

Article

Not peer-reviewed version

YOLOv8n-RMB: UAV Imagery Rubber Milk Bowl Detection Model for Autonomous Robots Natural Latex Harvest

[Yunfan Wang](#) , [Lin Yang](#) ^{*} , Pengze Zhong , [Xin Yang](#) , Chuanchuan Su , Yi Zhang , [Amir Hussain](#)

Posted Date: 25 August 2025

doi: 10.20944/preprints202508.1745.v1

Keywords: natural rubber milk; tapping robot; UAV imagery; YOLO; object detection



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

YOLOv8n-RMB: UAV Imagery Rubber Milk Bowl Detection Model for Autonomous Robots Natural Latex Harvest

Yunfan Wang ¹, Lin Yang ^{1,2,*}, Pengze Zhong ¹, Xin Yang ³, Chuanchuan Su ¹, Yi Zhang ¹ and Aamir Hussain ⁴

- ¹ College of Engineering, Huazhong Agricultural University, Wuhan 430070, China
- ² Key Laboratory of Agricultural Equipment in Mid-Lower Yangtze River, Ministry of Agriculture and Rural Affairs, Wuhan 430070, China
- ³ Leibniz Centre for Agricultural Landscape Research, Müncheberg 15374, Germany
- ⁴ Dept of Informatics, Modeling, Electronics, and Systems, University of Calabria, Rende 87036, Italy
- * Correspondence: lin.yang@hzau.edu.cn

Abstract

Natural latex harvest is pushing the boundary of unmanned agricultural production mode in rubber milk collection, by integrated robots in the land space of hilly and mountainous, like fixed and mobile tapping robots deployed in large-scale in the forest. As the bad working condition and complex natural environments among robber trees, the real-time and precision assessment of rubber milk yield status has emerged, as a key requirement for improving the efficiency and autonomy management of these kind of large-scale automatic tapping robots. However, the traditional manual rubber milk yield status detection methods are limited in their ability to operate effectively under conditions involving complex terrain, dense forest backgrounds, irregular surface geometries of rubber milk, and frequent occlusion of rubber milk bowls (RMB) by vegetation. To address this issue, this study presents an unmanned aerial vehicle (UAV) imagery rubber milk yield state detection method, termed YOLOv8n-RMB, in unstructured field environments instead of manual watching. The proposed method improved the original YOLOv8n by integrates structural enhancements across the backbone, neck, and head components of the network. First, a Receptive Field Attention Convolution (RFAConv) module is embedded within the backbone to improve the model's ability to extract target-relevant features in visually complex environments. Second, within the neck structure, a Bidirectional Feature Pyramid Network (BiFPN) is applied to strengthen the fusion of features across multiple spatial scales. Third, in the head, a content-aware dynamic upsampling module of DySample is adopted to enhance the reconstruction of spatial details and the preservation of object boundaries. Finally, the detection framework is integrated with the BoT-SORT tracking algorithm to achieve continuous multi-object association and dynamic state monitoring based on the filling status of RMB. Experimental evaluation shows that the proposed YOLOv8n-RMB model achieves an AP@0.5 of 94.9%, an AP@0.5:0.95 of 89.7%, a precision of 91.3%, and a recall of 91.9%. Moreover, the performances improve 2.7%, 2.9%, 3.9%, and 9.7%, respectively, comparing with the original YOLOv8n. Plus, the total number of parameters is kept within 3.0 million, and the computational cost is limited to 8.3 GFLOPs. And it meets the requirements of yield assessment tasks in computing resource-limited environments of rubber plantation fixed and mobile tapping robots.

Keywords: natural rubber milk; tapping robot; UAV imagery; YOLO; object detection

1. Introduction

Natural rubber milk is one of the four most important and scarce industrial materials, along with steel, oil and coal, and is the raw materials of more than 70,000 products, due to its excellent resilience,

electrical insulation, wear resistance, air tightness, and flexibility[1]. Moreover, the global demand of natural rubber milk is still speeding rising, and it is not satisfied by the natural latex production[2]. It is obtained from the latex tubes of rubber trees (*Hevea brasiliensis*) through regular tapping, and the rubber plantations are mostly located in hilly and mountainous areas with uneven terrain. In these regions, manual inspection and harvesting are limited by poor accessibility and high labor demands. [3] However, the traditional production processing remains heavily reliant on manual intensive for bark cutting and yield inspection, and it requires skilled tapping and collection workers under the bad working environment in night time in forest. Thus, the labor shortage with aging is a critical bottleneck in the natural rubber milk industry. Luckily, such challenges are being typically addressed by replacing manual labor with automated agricultural robotic systems, like both “one tree one machine”(OTOM) fixed-position and mobile rubber tapping robots [4].

Unmanned and automatic tapping robots generally are concentrated on the tapping procedure of tapping depth, tracing, start points via RGB and Lidar based vision.[5] The OTOM fixed robot is attached on one rubber tree, and automatically tapping the tree following the previous tapping tracing, lines, and depth [6] each three or four days. And the mobile robot is carried with a tapping machine, and travels among the trees, and tapes a tree assisted with tapping line image recognition and spatial control[7]. The tapping robots are triggered by fixed time-counter control, and starts the tapping operation guided by images [8] or laser scanning 3D points to achieve precise trajectory fitting, with strong adaptability to trees with irregular diameters and complex surface structures [9]. However, current OTOM and mobile tapping robots primarily focus on tasks such as tapping path planning and initial line positioning, while the perception of rubber milk collection status remains underexplored. Actually, in a real scenario of large-scale robots, the failure to recognize the remaining rubber milk in the rubber milk bowl (RMB) may easily result in rubber milk overflow or trees repeated tapping, leading to resource waste and reduced production efficiency. After tapping, rubber milk naturally flows into the rubber milk bowl through the incision, forming a visible liquid level. Based on the height of the rubber milk, the yield status of each rubber milk bowl can be classified into three categories: empty, partially filled, and full. So, the statue of RMB, like fill and unfilled, is one of key issue for the unmanned robber tapping robots’ management, especially related to robots’ decision of the suitable time of starting and natural latex harvest.

The status of RMB could be images from different scales of ground vehicle, Satellite, and unmanned aerial vehicle (UAV) [10]. Conventional approaches to rubber milk bowl image recognition typically involve mounting cameras directly onto tapping machines. However, such configurations face practical challenges for both OTOM and Mobility. For examples, each unit of OTOM with an independent vision system can significantly increase equipment costs and energy consumption in large-scale plantations. And the mobile tapping robots often encounter operational limitations in complex terrains, including difficulty in adjusting camera angles and frequent target occlusion. On the other hand, Satellite based imagery for the rubber yield prediction is explored [11] to estimate monthly rubber production at the plantation scale, with the overarching goal of stabilizing national rubber prices.[12] But such methods suffer from limited real-time applicability and relatively low accuracy, rendering them unsuitable for daily rubber milk collection monitoring at the individual farmer level. To overcome these issues, a UAV-based low-altitude visible-light imaging could be a suitable potential solution for efficient, cost-effective, and non-contact yield recognition of rubber milk bowls in the hilly and forest. The scenarios are shown in Figure 1(a), by integrating UAV imagery with computer vision algorithms to extract rubber milk level features, an artificial intelligent model can automatically determine the current yield status of each bowl. This enables precise localization and quantification of harvestable targets, providing robust perception support for intelligent rubber tapping systems.

UAV-imagery crop detection and counting by combining machine vision methods are successfully applied, like tree crown defoliation by extracting vegetation indices from multispectral imagery[13], and estimating equivalent water thickness in wheat under field conditions[14]. These early works of target detection and counting exhibit poor robustness to complex backgrounds and

varying lighting conditions, limited generalizability across crop varieties, and reduced computational efficiency in high-resolution imaging environments. And the based models of You Only Look Once (YOLO) are improved to address these challenges by changing and modifying the network structures. And the UAV images method is deployed in oil palm fruit detection[15], cotton crop [16], pine tree disease identification[17], pine tree diseases [18], in improving YOLOv3 for strawberries flowering and fruit counting[19], and in localization and accurate counting of peanut seedlings through frame-by-frame analysis via improved YOLOv8[20]. These UAV imagery recognition study provides technical models and basement for rapid dynamic monitoring and large-scale rubber bowl detection and counting.

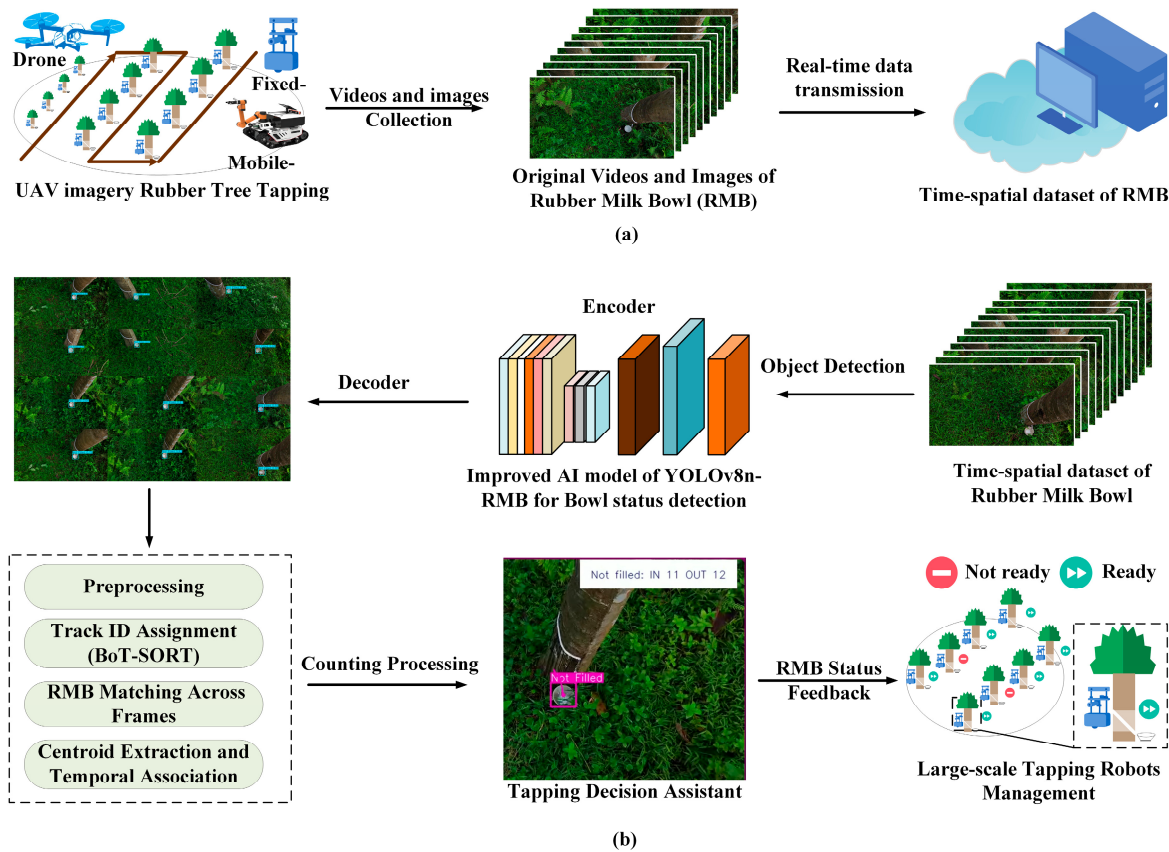


Figure 1. Scenarios and Concept of UAV assisted rubber milk tapping robots’ management. Figure(a) illustrates the data capturing and transmission via Drones, Figure (b) depicts the control logic of the rubber tapping machine via improved YOLOv8n-RMB.

The UAV imagery and target detection models has specific requirements from the features of procedures tapping robot control, RMB, and the environments as shown in Figure 1(b). In real scenarios, since the typically dark color of RMBs, visual confusion often arises between the bowls and background elements such as fallen leaves and tree trunks, which share similar chromatic features. This color similarity increases the likelihood of false detections. Moreover, the subtle visual differences between the “Not Filled” and “Filled” states frequently lead to mutual misclassification, thereby complicating the accurate assessment of rubber yield status. Thus, this work investigates rubber yield status counting under complex background conditions using an improved YOLOv8n-RMB model and UAV imagery. The main contributions are listed as below:

- (1) An originally and novel dataset of rubber milk bowl detection for large scale tapping robots’ control is built and reported. As the best as we know from literate review, this work is the first to capture and present UAV imagery natural rubber milk among the hilly and mountain forest.
- (2) An improved YOLOv8n-RMB is present to real-time natural rubber milk status recognition. The proposed model integrates the receptive field attention convolution (RFACONV) module,

replaces with a bidirectional feature pyramid network (BiFPN), and introduces the dynamic sampling module DySample into the upsampling stage.

(3)A pilot experiment of UAV imagery RMB target detection is implemented in real scenarios. The proposed YOLOv8n-RMB can be used in computing consumption devices and count and prediction the rubber milk in real time.

2. Materials and Methods

2.1. Data Collection and Annotation

The RMB images and videos of the dataset Opensource Dataset of RMB weblink https://drive.google.com/file/d/15-OnktSHcgs-WKghrHUoT0aG4VtBpCGi/view?usp=drive_link (open sourced in a google driver) is originally collected from main planting area rubber tree, Hainan province in the year 2025 and Yunnan province in the year 2024, China, as shown in Figure 2(a). And the training images of rubber liquid in the RMB come from the natural rubber forest of Danzhou city of Hainan Province and Ruili city of Yunnan Province, China. The images and videos were captured by using devices of both mobile phone and drone, and the resolution of there images and video are 1920x1080. Moreover, the images were collected at three times intervals of 1 hour, 12 hours, and 24 hours after tapping. To enable real-time and dynamic monitoring of rubber yield, a drone of the DJI FLIP is selected to conduct low-altitude flights among the rubber trees for imaging and video capturing, as shown in Figure 2(b). It is recorded and UAV-assisted for the recognition and counting of RMB states. The UAV flight scenario is illustrated in Figure 2(c), and a sample of the recognition output is shown in Figure 2(d).

As the dataset samples captured under complex background conditions, and that poses greater recognition challenges. This strategy not only enriches the variability of rubber milk volumes in the RMBs, but also ensures environmental diversity within the plantation. Representative sample images from the dataset are shown in Figure 3. The rubber milk surface in the RMBs was manually annotated using the labeling image annotation software. Three labels were assigned: None, not filled, and filled. These three states are forming a data set in txt format.

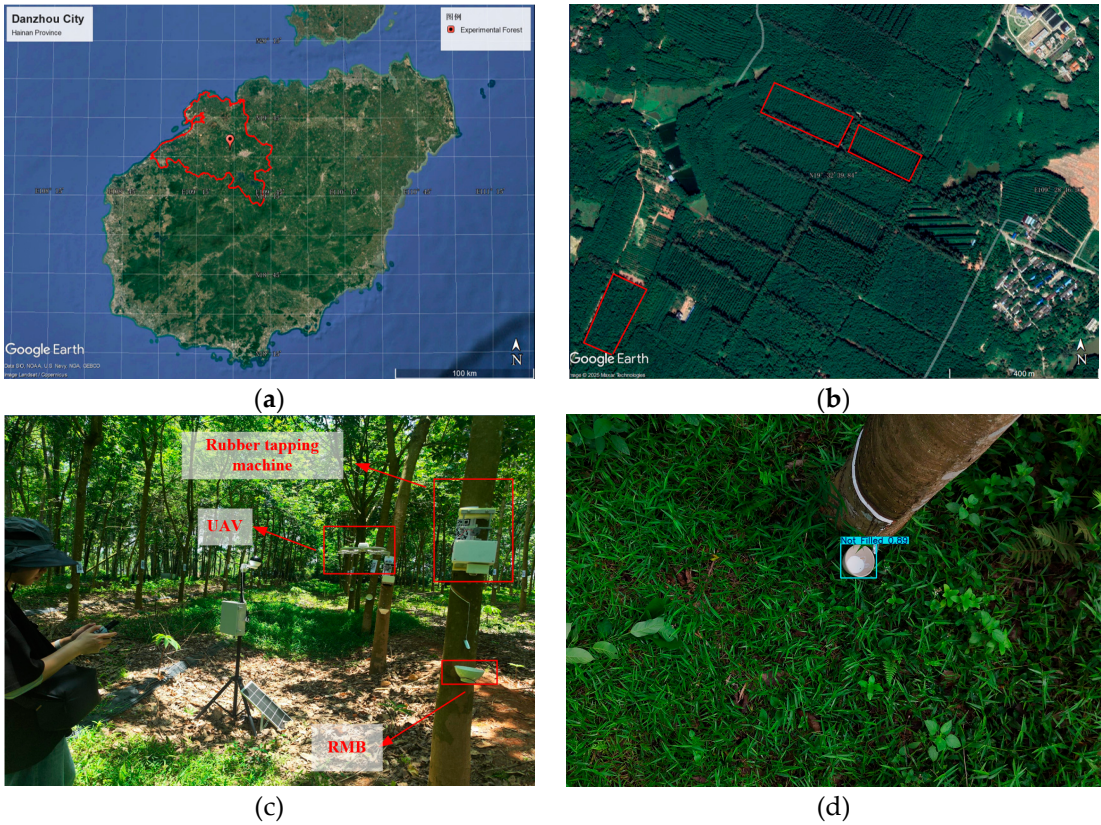


Figure 2. Figure(a) illustrates the location of Danzhou City on the map, Figure(b) presents the experimental rubber plantation, Figure(c) shows the experimental scene, and Figure d depicts the UAV-based recognition view.

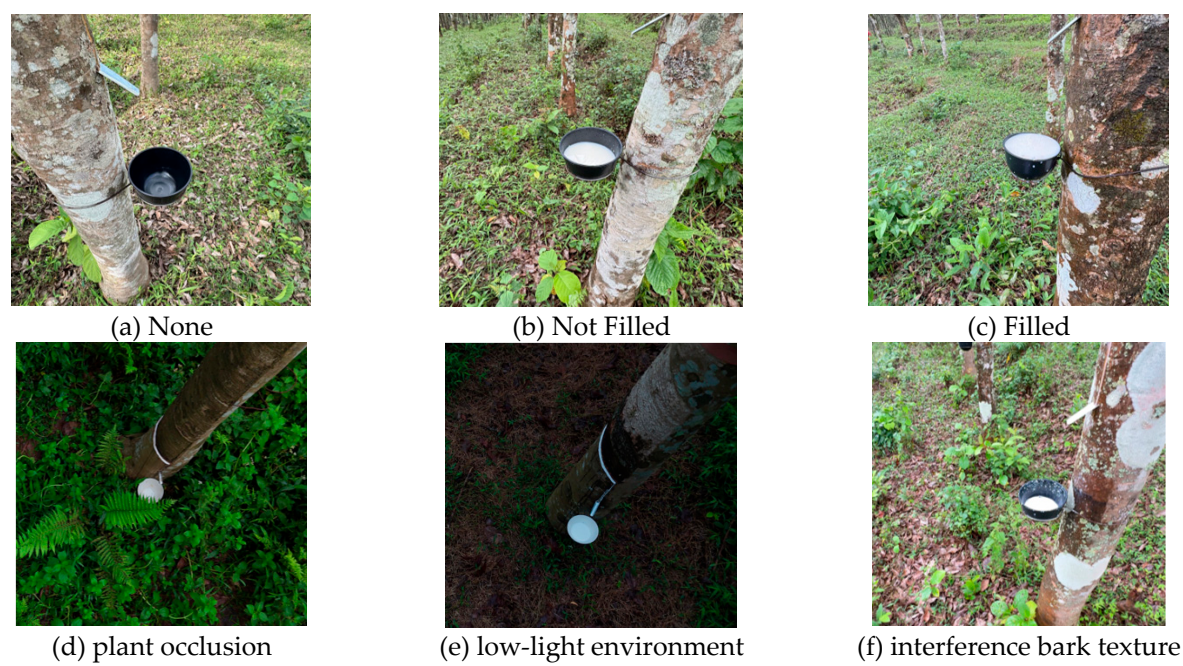


Figure 3. UVA imagery BMW status (a-c)and samples in complex background environment(d-f).

2.2. Data Enhancement

The performance of deep learning models depends on the quality of the dataset. Important factors include richness, diversity, and annotation accuracy. This dependency is particularly observed in the visual inspection of rubber tapping surfaces within complex rubber forest environments, environmental interference factors such as dynamic lighting changes, complex background noise and vegetation occlusion can easily lead to characteristic deviations in the collected sample data, which in turn causes the model to overfit and reduce its spatial generalization ability. In this study, random enhancement operations were applied to the original dataset images. These operations included image flipping, addition of Gaussian noise, and adjustment of brightness, color, and contrast, as shown in Figure 4. Finally, the number of enhanced data sets reached 2100 images, which were divided into training sets and validation sets in a ratio of 7 to 3.

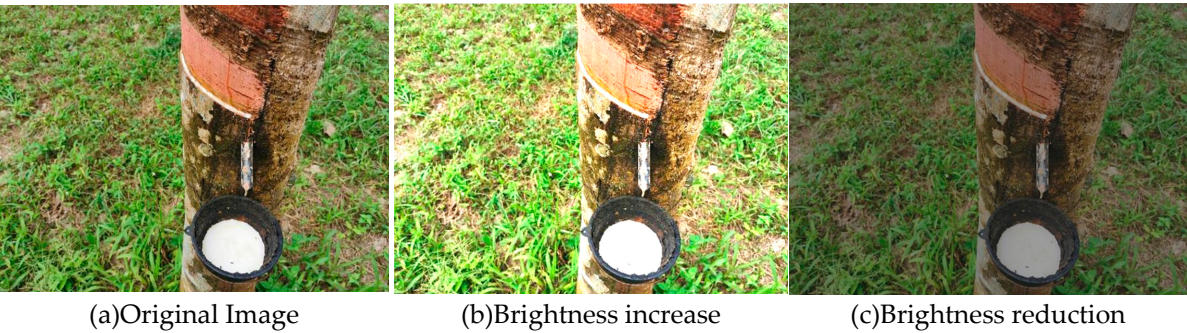




Figure 4. Data enhancement of image flipping, addition of Gaussian noise, and adjustment of brightness, color, and contrast.

2.3. Standard YOLOv8 Network Structure

YOLOv8 is a version of the YOLO object detection model that was released by Ultralytics in 2023. [21]. Maintaining the fundamental principle of “single-stage detection with high real-time performance,” YOLOv8 integrates a series of architectural refinements and algorithmic innovations, markedly enhancing detection accuracy, robustness, and computational efficiency[22]. Due to these improvements, YOLOv8 has been widely used in areas such as precision agriculture, industrial inspection, and autonomous systems. The model provides five scalable variants—YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large)—to meet different performance and hardware needs. Given the real-time inference requirements and limited computational resources in rubber tapping scenarios, this study adopts the lightweight YOLOv8n model as the base detection architecture to balance accuracy and efficiency. In this study, the lightweight object detection model YOLOv8n was used as the baseline architecture. YOLOv8n is the smallest model in the YOLOv8 series. It provides fast inference and low computational cost, which supports deployment on resource-limited platforms and in real-time detection tasks. To improve detection accuracy and classification consistency of RMBs in complex forest environments.

The overall architecture of YOLOv8 is organized into three functional parts: the backbone, the neck, and the head. The input image is first processed by the backbone network, which applies modules such as CBS (Convolution-BatchNorm-SiLU), SPPF (Spatial Pyramid Pooling - Fast) and C2F (Cross-Stage Partial Fusion) to extract multi-scale semantic and spatial features. These features are then passed through the neck, which uses a Path Aggregation Feature Pyramid Network (PAFPN) to enhance feature integration across scales while preserving contextual information. The head network outputs object categories, confidence scores, and bounding box coordinates, enabling accurate detection under complex environmental conditions.

3. Improvement of YOLOv8n-RMB Network

3.1. The Improved Network Architecture of YOLOv8n-RMB

The YOLOv8n-RMB (<https://github.com/coiuy/YOLOv8n-RMB.git>) integrates three core modules: RFACONV, BiFPN, and DySample. This design maintains model compactness while improving detection accuracy. The model performs well in identifying the three states of RMBs—empty, partially filled, and fully filled—under various complex field conditions. The overall network structure of YOLOv8n-RMB is shown in Figure 5.

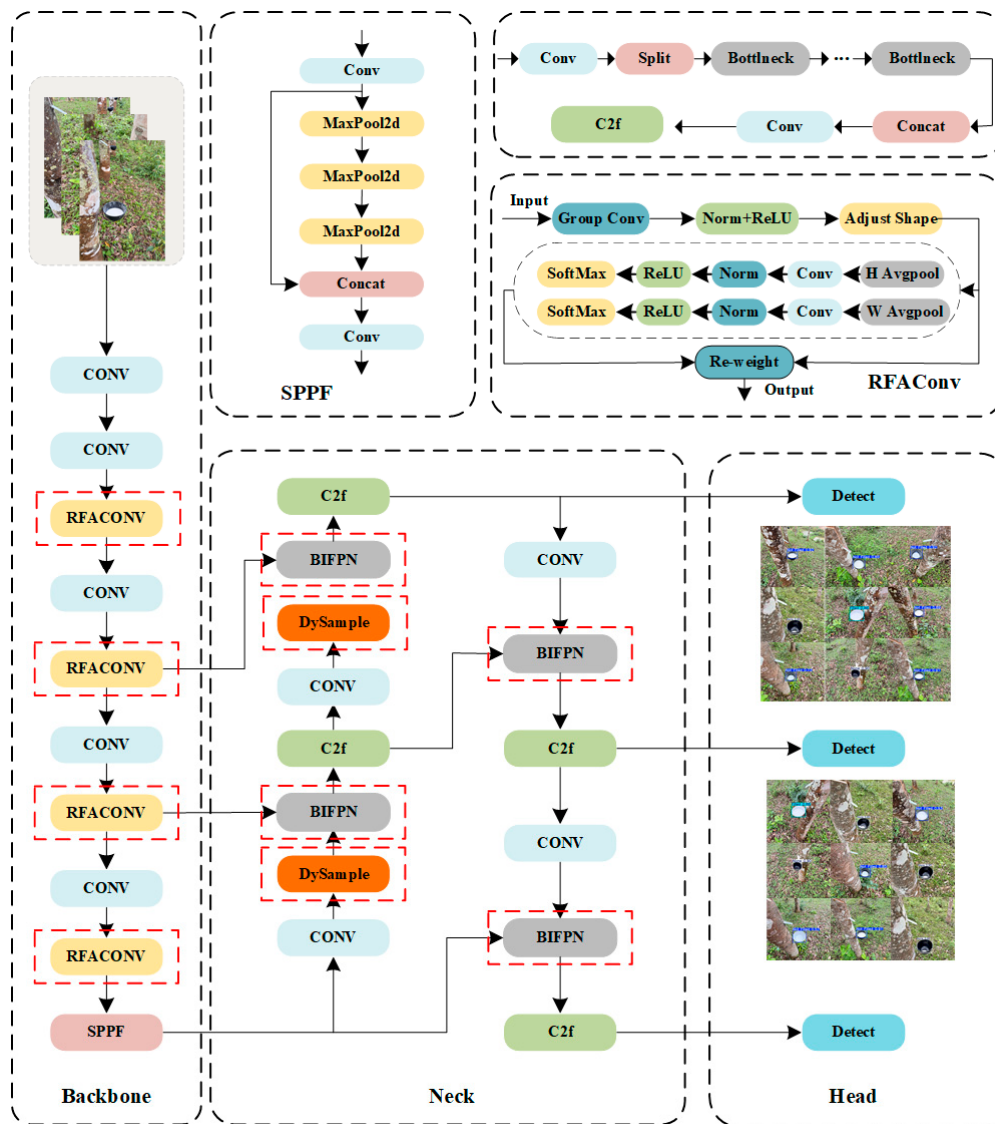


Figure 5. Structure of YOLOv8n-RMB. The red dashed boxes in the figure indicate the modules that replace components of the original YOLOv8n. In RFACONV, AvgPool is used to capture global information within each receptive field and SoftMax emphasizes important features. In C2f, MaxPool2D denotes the max pooling operation.

3.2. Improvement of Backbone Network

In YOLOv8, standard convolution operations use the same parameters for all receptive fields. Feature information is extracted through convolutional kernels without considering positional variation. This uniform approach often causes redundant information to appear in the extracted data. The introduction of spatial attention mechanisms allows models to focus on salient features [23–25]. This improves the network’s ability to capture fine-grained feature representations.

At present, spatial attention mechanisms such as the Convolutional Block Attention Module (CBAM) and Channel Attention (CA) are widely used to improve the performance of convolutional neural networks. The structures of CA and CBAM are shown in Figure 6. However, these spatial attention mechanisms mainly focus on spatial features and do not effectively handle the issue of parameter sharing in convolutional kernels.

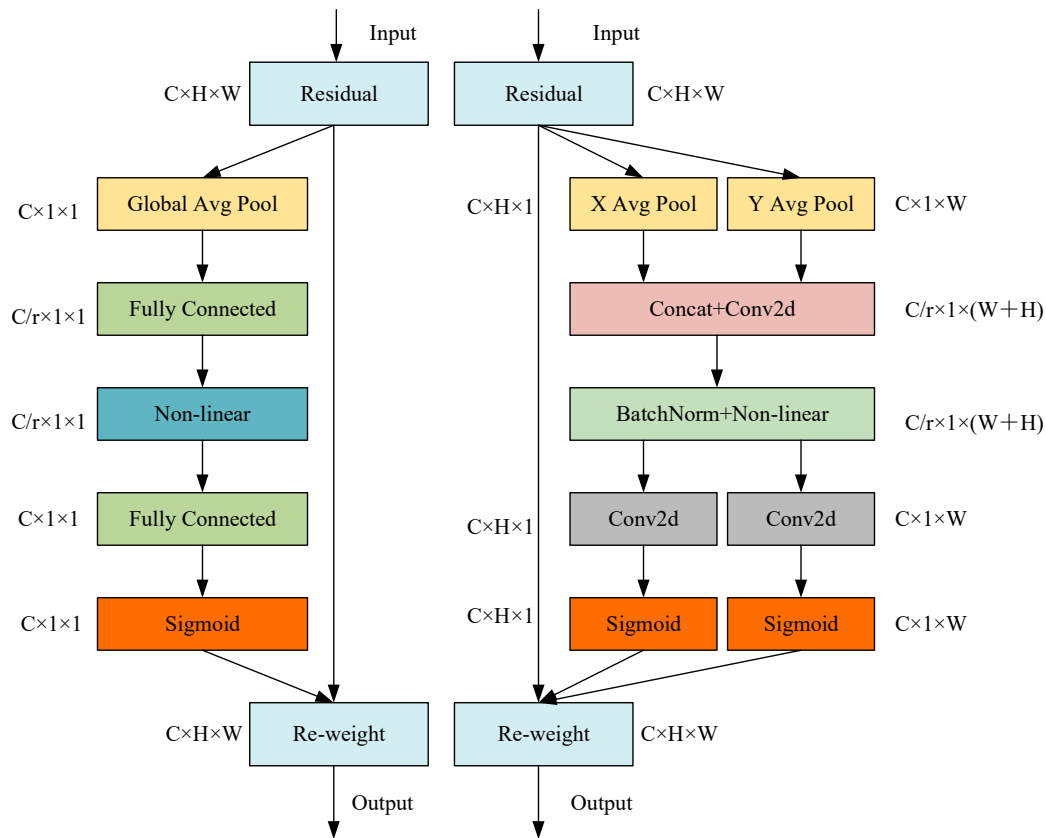


Figure 6. The Structure of CA and CBAM.

To overcome the limitations of existing spatial attention approaches, this study introduces a novel Receptive Field Attention Convolution[26] (RFAConv). RFAConv not only emphasizes the importance of distinct features within the receptive field window but also prioritizes the spatial features of receptive fields, effectively addressing the problem of parameter sharing in convolution operations[27]. As shown in Figure 7, the RFAConv structure uses 3×3 convolutional kernels. In this structure, C , H , and W represent the number of channels, input height, and input width, respectively. “ 3×3 Group Conv” refers to a 3×3 group convolution operation, and “AvgPool” represents an average pooling operation.

First, the convolution layer applies an initial convolution to the input features to extract low-level information and produce a preliminary feature map, and the model adjusts the contribution of each feature based on this map to better capture the most important information in the input. The feature weights are then updated through a reweighting mechanism. Next, average pooling is used to aggregate the global information of each receptive field to reduce the computational overhead caused by the interaction between features and a 1×1 group convolution is applied to enable information interaction, and Softmax is used afterward to emphasize the importance of features within each receptive field.

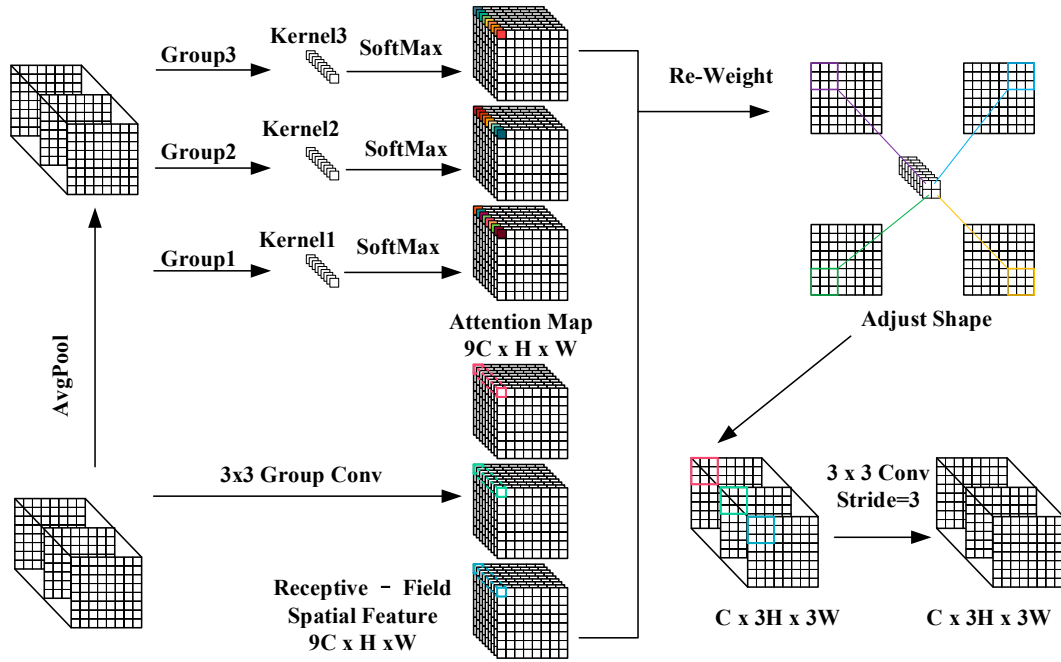


Figure 7. The Structure of RFACnv.

In general, the computation of receptive field attention (RFA) can be expressed as follows:

$$F = \text{Softmax}(g^{i \times i}(\text{AvgPool}(X))) \times \text{ReLU}(\text{Norm}(g^{k \times k}(X))) = A_{\{rf\}} \times F_{\{rf\}} \quad (1)$$

Where $g^{i \times i}$ represents a grouped convolution of size $i \times i$, k represents the convolution kernel size, Norm represents normalization, X represents the input feature map, and F is obtained by multiplying the attention map $A_{\{rf\}}$ with the transformed receptive field space feature $F_{\{rf\}}$. Compared with traditional convolutional attention modules, RFA generates attention maps for individual receptive field features.

3.3. Improvement of Neck Network

Feature fusion is important in target detection tasks and helps combine information from different scales to improve detection accuracy. The traditional FPN combines features across levels (P3 to P7) using a top-down information flow, as shown in Figure 8(a). However, due to the characteristics of unidirectional transmission, FPN has certain limitations in processing positioning information, which easily leads to the loss of spatial details. To this end, the path aggregation network (PANet) introduces a bottom-up path based on FPN to form a bidirectional information flow, as shown in Figure 8(b), which effectively compensates for the lack of positioning features and enriches semantic features. In YOLOv8, the feature extraction network adopts a combination of FPN and PANet, and improves the representation of the feature pyramid through feature fusion of P4-N4 and P5-N5. However, the PAN-FPN structure still suffers from several limitations in the context of RMB detection under complex background conditions, including feature redundancy, insufficient utilization of shallow features, and low feature reuse efficiency. These shortcomings reduce the model's ability to capture key cues such as bowl edge contours and subtle differences in yield states.

To solve the above limitations, this paper introduces a bidirectional feature pyramid network[28] (BiFPN), whose structure is shown in Figure 8(c). BiFPN was proposed by Google in EfficientDet to improve the transfer and combination of features across different scales. It introduces a bidirectional structure that allows information to move both upward and downward between layers, and applies learnable weights to adjust the contribution of each input during feature merging. Compared with PANet, BiFPN uses a simpler structure and eliminates nodes that have only one input, which helps reduce unnecessary computations. At the same time, new paths are introduced that link input and output features within the same resolution layer, which further improves the diversity and quality

of the fused features. In addition, BiFPN introduces a weighted fusion mechanism, which avoids the information loss problem caused by simple feature superposition by assigning learnable weights to different input features, and adopts a fast normalization method to improve training speed and data consistency.

The integration of BiFPN not only preserves the advantages of both FPN and PAN structures, but also enhances feature interaction through the addition of two lateral pathways. The introduction of the P2 layer provides higher-resolution feature maps, which are beneficial for capturing fine-grained details such as RMB edges and yield states. Furthermore, the repeated two-way feature pathways in BiFPN support full multi-scale integration, which improves the model's response to state changes and increases classification accuracy in complex background conditions. In summary, BiFPN offers a structurally simple yet computationally efficient feature fusion strategy, making it an effective architectural enhancement for addressing the challenges posed by environmental variability and inter-class similarity in rubber plantations. This provides essential support for intelligent agricultural tasks, including yield estimation and automated RMB counting.

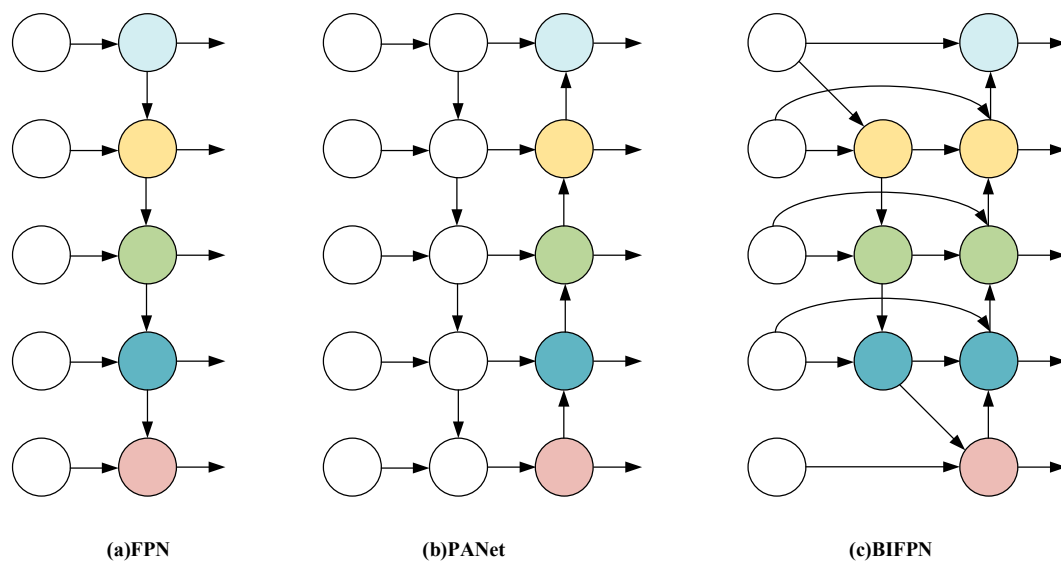


Figure 8. Comparison of neck network structure.

3.4. Improvement of Upsampling Module

To effectively balance the semantic information of deep features with the spatial information of shallow features during the feature fusion stage, neural networks typically employ upsampling operations to restore spatial resolution. In the original YOLOv8 architecture, bilinear interpolation is commonly used for upsampling, where sampling locations are fixed and independent of the input image content. This static interpolation strategy presents limitations in complex scenes or for irregular objects, as it fails to adapt sampling based on the characteristics of the feature map, thereby compromising the precision and robustness of feature fusion.

To solve this problem, the DySample mechanism is introduced to adjust the upsampling process in YOLOv8 based on input feature characteristics[29]. DySample dynamically generates sampling locations based on feature content. Its core concept lies in learning spatial offsets for each pixel in the input feature map to adaptively adjust sampling positions. Specifically, DySample first transforms the input features via a linear layer to generate displacement offsets, which are then multiplied by a scaling factor to compute relative coordinate shifts. These are embedded into the feature map using a pixel shuffle operation and added to the base sampling coordinates to obtain dynamic sampling positions. Feature resampling is subsequently performed using grid sampling based on these coordinates.

As illustrated in Figure 9, the DySample module consists of key components such as the input feature map, linear transformation layer, offset computation, pixel shuffle, and grid sampling. Unlike

fixed interpolation methods, DySample adjusts the upsampling strategy based on image content. This approach is effective in complex backgrounds, unstructured environments, and cases involving uneven target distribution. In rubber plantations and other natural environments, DySample improves the model’s ability to adapt to uneven terrain and detect edge objects. This leads to more accurate feature restoration and fusion, resulting in balanced and detailed feature representations.

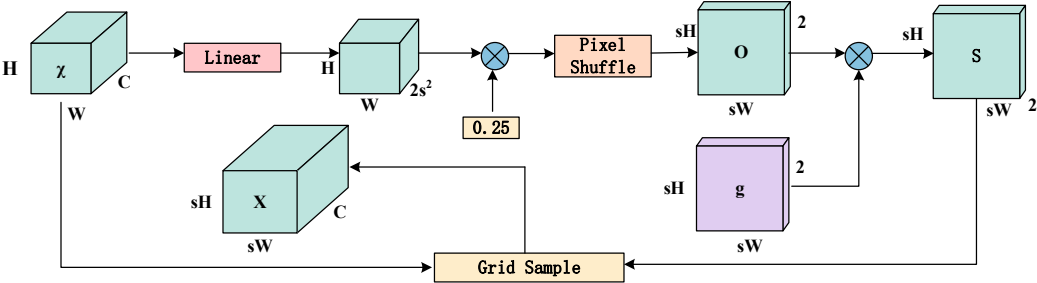


Figure 9. The Structure of Dysample.

3.5. Real-Time Video Counting Algorithm

This study developed a video analysis system that integrates the YOLOv8n-RMB model with the BoT-SORT algorithm to achieve automated detection, state recognition, and accurate counting of RMBs in UAV aerial videos. The system first establishes a stable video streaming platform by configuring an NGINX server with the RTMP module, which streams real-time plantation footage captured by the UAV to the server using the RTMP protocol, ensuring low-latency and high-frame-rate video input. Subsequently, the detection script utilizes multimedia processing libraries such as OpenCV and FFmpeg, and employs cv2.VideoCapture to continuously acquire frames from the RTMP stream, thereby providing real-time input for the target analysis module. For each frame, the optimized YOLOv8n-RMB model is applied to detect RMB targets and classify their yield status into three categories: empty, partially filled, and fully filled. The BoT-SORT (ByteTrack with Transformer-enhanced appearance features) multi-object tracking algorithm is then used to assign a unique identifier (ID) to each target. By leveraging motion trajectories and appearance features, the algorithm achieves robust cross-frame association and tracking, ensuring coherent and stable target trajectories under complex plantation backgrounds. Based on the identified status of each bowl, the model draws corresponding bounding boxes and labels while maintaining a record of its historical positions and recognition states.

To ensure accurate counting, the system predefines reference lines or zones in the video frames. When a detected target crosses these boundaries or is consistently classified as “fully filled” over consecutive frames, it is added to the count of RMBs ready for collection, effectively minimizing duplicate counting. All processed frames are displayed in real time on the front-end interface with detection and counting results, and optionally stored as video files for traceability and accuracy verification. By integrating detection, recognition, tracking, and counting functionalities, this system provides a comprehensive and efficient video-based intelligent processing solution for unmanned rubber milk collection operations.

3.6. Experimental Environment and Model Evaluation Indicators

3.6.1. Experimental Environment

The experiments in this study were implemented using the open-source machine learning framework PyTorch 1.11. The hardware environment consisted of an Intel® Xeon® Platinum 8352V CPU @ 2.10 GHz, 64 GB of RAM, and an NVIDIA RTX 4090 GPU with 24 GB of video memory, operating under CUDA 11.3. The software platform was based on Ubuntu 18.04 with Python 3.8.

Model training was conducted over 100 epochs with a batch size of 32. The initial learning rate was set to 0.01, while the momentum parameter was configured at 0.937 to accelerate convergence and suppress oscillation. Furthermore, a weight decay coefficient of 0.0005 was applied to prevent overfitting by penalizing large weights during optimization.

3.6.2. Model Evaluation Indicators

The evaluation of model performance in this study is based on both detection accuracy and inference efficiency. Specifically, Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP) are adopted as the primary metrics to assess detection accuracy. In addition, to evaluate the inference speed with respect to hardware constraints, model parameters, floating point operations (GFLOPs), frames per second (FPS), and latency are used as the main speed-related indicators. The calculation formulas for Precision, Recall, AP, and mAP are defined as follows:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP = \frac{1}{N} \sum_{c=1}^N AP_c \quad (5)$$

In the equations, TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Let $P(R)$ be the precision at each recall level. N is the total number of object classes, and AP_c is the average precision for class c . P is the ratio of true positives to all predicted positives. It reflects how well the model avoids false positives. R is the ratio of true positives to all actual targets. It shows the model's ability to reduce missed detections. AP is the area under the precision-recall curve. It is calculated by summing precision over different recall thresholds. mAP is the average of AP over all classes. It is often used to measure object detection performance. A higher mAP means better accuracy and stability.

4. Results and Discussion

4.1. Comparison of Model Before and After Improvement

To further evaluate the performance of the proposed YOLOv8n-RMB model, this study compares the training loss curves of the original YOLOv8n and the improved YOLOv8n-RMB. Loss curves show how the loss function changes across training epochs. They help assess the effect of each loss component and offer information about the model's generalization and training stability on the validation set. The comparison results are shown in Figure 10. The comparison results are shown in Figure 10. Box_loss is the bounding box regression loss. It measures the error between predicted boxes and ground-truth boxes and this value reflects the accuracy of object localization. Dfl_loss is the distribution focal loss. It captures uncertainty in classification and localization. cls_loss is the classification loss, measuring the model's ability to correctly predict object categories. In general, lower and more stable loss values are indicative of better model performance.

As shown in Figure 10, the YOLOv8n-RMB model gives lower box_loss and dfl_loss in most training epochs. This result indicates better bounding box regression and higher localization accuracy compared to the original YOLOv8n. Furthermore, the improved model exhibits smaller fluctuations, indicating enhanced training stability. Although the cls_loss curve of YOLOv8n-RMB shows minor oscillations during the initial training phase, it converges to a lower value in the later epochs, reflecting improved classification performance in distinguishing between the three RMB states—"None," "Not Filled," and "Filled"—especially under complex background conditions. The comparative loss curves substantiate the performance advantage of the YOLOv8n-RMB model in the RMB state recognition task within visually challenging environments.

Figure 11 provides a visual comparison of the PR curves before and after model enhancements. AP measures the area under the PR curve, where a larger area indicates higher detection accuracy for a given class. Subfigure (a) illustrates the performance of the original YOLOv8n model across the three RMB states, while subfigure (b) presents the PR curves of the improved YOLOv8n-RMB model incorporating RFA, BiFPN, and DySample modules. It is evident that YOLOv8n-RMB exhibits higher precision and recall across all classes, with significantly larger enclosed areas under the curves, particularly for the “Not Filled” and “Filled” categories, which are more prone to misclassification.

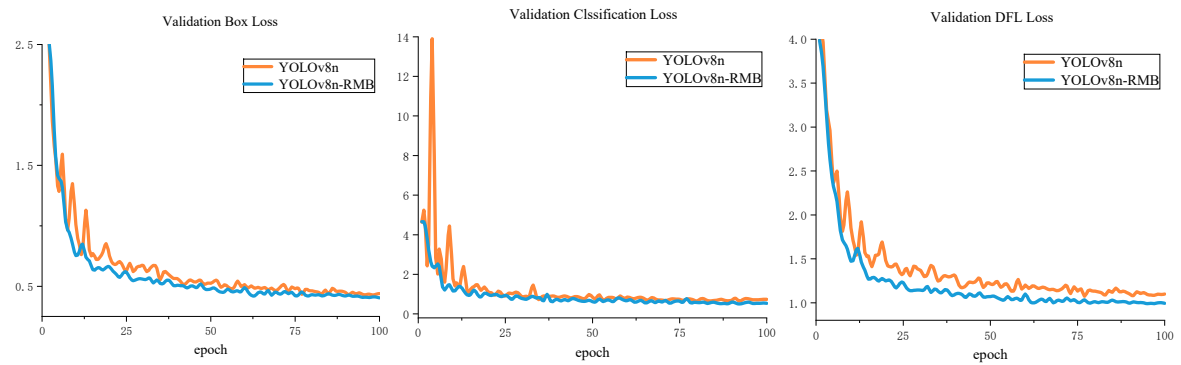


Figure 10. Comparison of loss curves before and after model improvement.

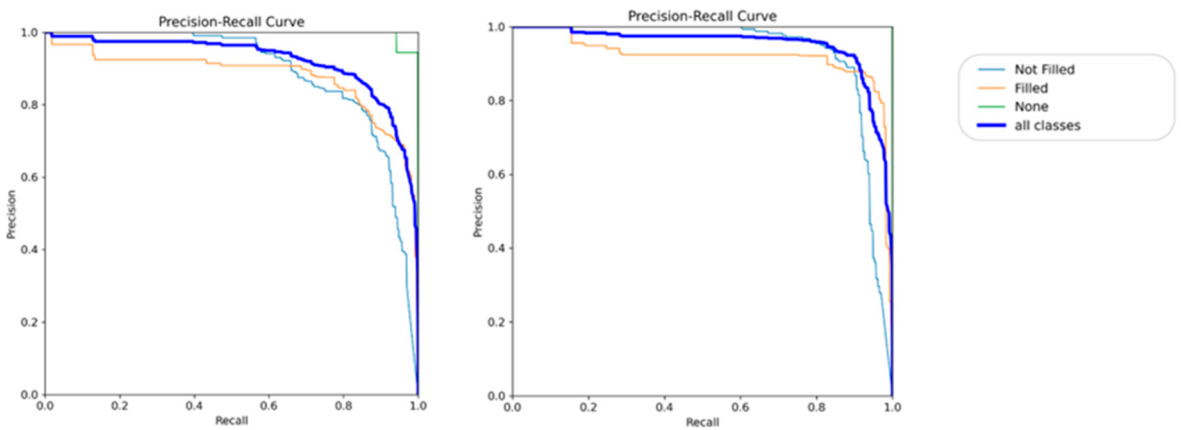


Figure 11. Comparison of P-R curves. The left shows the PR curve of YOLOv8n, and the right shows the PR curve of YOLOv8n-RMB.

As shown in Table 1, the proposed YOLOv8n-RMB model outperforms the baseline in terms of mAP across all three RMB states. Specifically, the mAP for the “None” class improved from 99.2% to 99.5%, marking a 0.3 percentage point gain. For the “Not Filled” category, mAP increased from 89.0% to 93.1%, and for the “Filled” class, it rose from 88.0% to 91.9%, representing improvements of 4.1 and 3.9 percentage points, respectively. These results underscore the enhanced discriminative capability of the improved model in fine-grained classification of RMB states.

Figure 12 presents a comparative analysis of detection results before and after model enhancement. As shown in Figures 12(b), the original model was prone to false positives in complex backgrounds and low light quality, often misled by visual interference from leaves and tree trunks. Additionally, it occasionally misclassified “Not Filled” bowls as “Filled,” indicating limitations in fine-grained state discrimination. In contrast, the improved YOLOv8n-RMB model substantially reduced the false detection rate and exhibited enhanced capability in distinguishing between different bowl states. As illustrated in Figures 12c), the refined model produced more accurate and reliable detection results, demonstrating its robustness in challenging field conditions.

Table 1. mAP comparison before and after improvement.

Models	Category	mAP
YOLOv8n	None	99.2%
	NotFilled	89.0%
	Filled	88.0%
YOLOv8n-RMB	None	99.5%
	NotFilled	93.1%
	Filled	91.9%

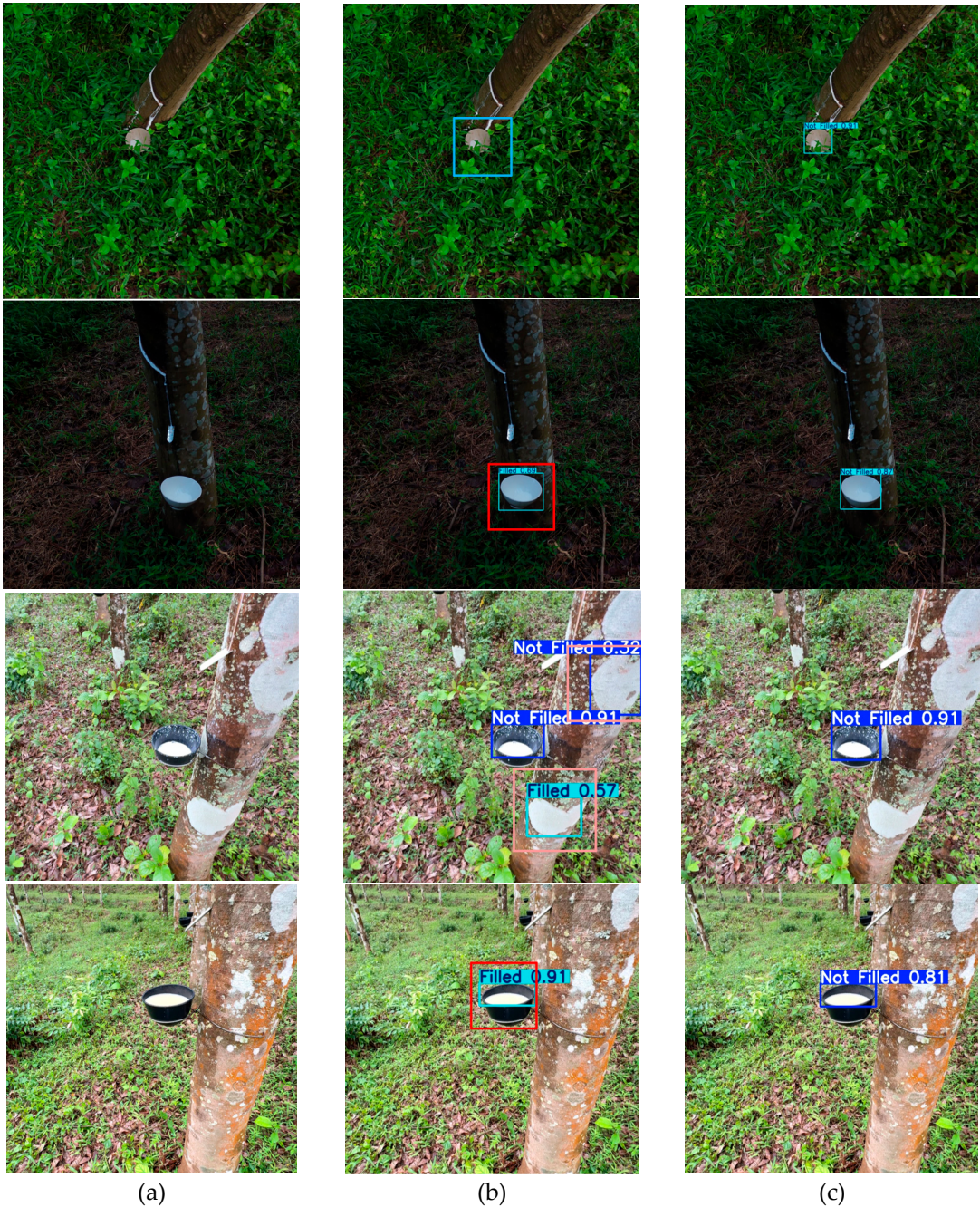


Figure 12. Comparison of model effects before and after improvement, Among them, (a) is the image before processing, (b) is the image detected by YOLOv8n, and (c) is the image detected by YOLOv8n-RMB. The blue boxes indicate misclassified objects, the red boxes represent missed detections, and the pink boxes denote false positives caused by background misclassification.

To intuitively analyze the differences in attention to RMB target regions before and after model improvement, we employed heatmap visualization to compare the feature response regions of the YOLOv8n and YOLOv8n-RMB models, with the results presented in Figure 13. In the heatmaps, brighter colors represent regions with higher model attention, while darker colors indicate lower attention. A comparative analysis reveals that the original YOLOv8n model exhibits relatively limited attention regions, with some heat-activated areas failing to fully cover the RMB body and even misfocusing on background regions. In contrast, the improved YOLOv8n-RMB model demonstrates a broader and more concentrated attention range, with significantly enhanced focus on the RMB itself. The YOLOv8n-RMB model shows higher response intensity and more precise spatial localization of the target region. This result supports the effectiveness of the structural changes in improving detection accuracy and robustness under complex background conditions.

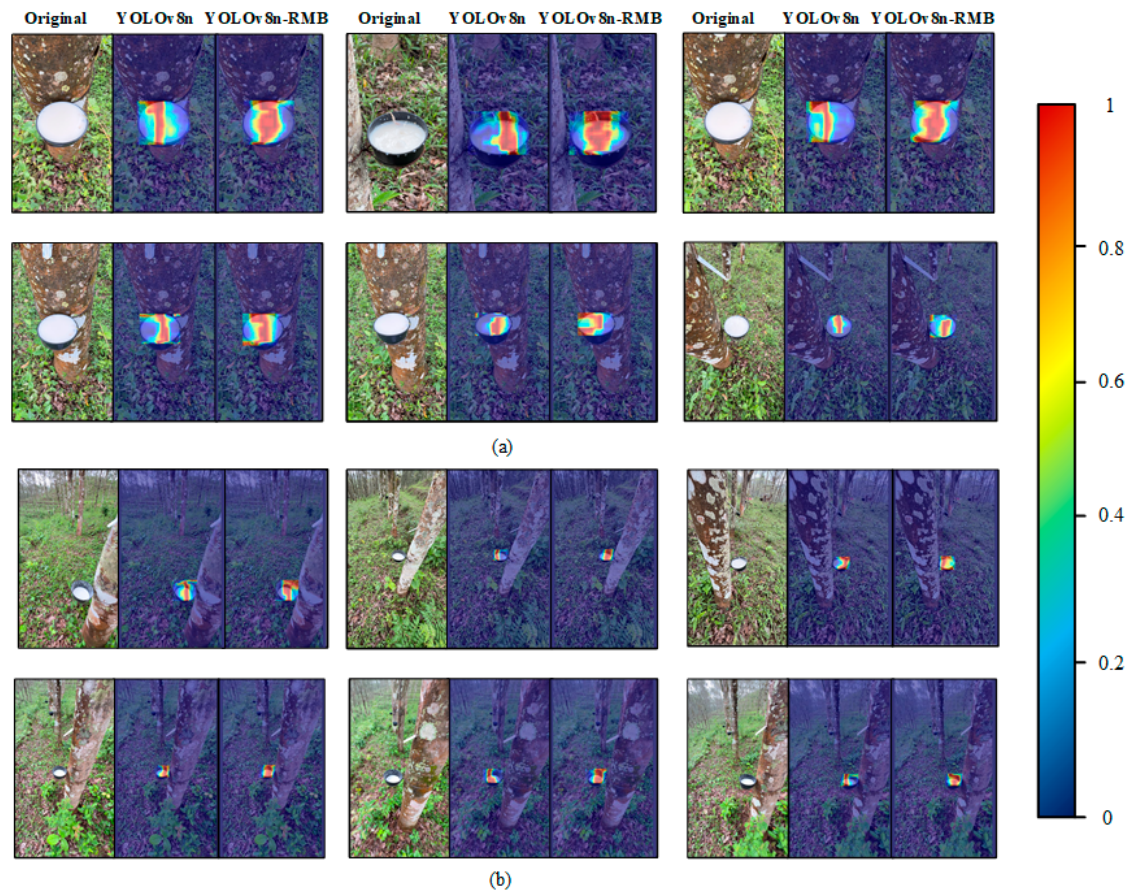


Figure 13. Comparison of heatmaps before and after model improvement. Figure (a) presents the heatmap-based recognition results for close-up views of RMBs, while Figure (b) shows the results for distant views. In each group of images, the sequence from left to right represents the original image, the recognition result from the baseline model, and the result from the improved model, respectively. In the heatmaps, brighter colors represent regions with higher model attention, while darker colors indicate lower attention.

Overall, the proposed YOLOv8n-RMB model achieved optimal performance in the task of RMB and counting. Compared with the original YOLOv8n model, It showed increases of 2.8% in $mAP@0.5$, 2.9% in $mAP@0.5:0.95$, 3.9% in Precision, and 9.7% in Recall. These results reflect improvements in detection accuracy and robustness across multiple metrics. Notably, while maintaining high detection precision, the YOLOv8n-RMB model retained its lightweight architecture. This was achieved by embedding the RFAConv module to strengthen local feature extraction, introducing the BiFPN structure to adjust feature fusion across scales, and incorporating the Dysample module for high-quality upsampling. These structural enhancements collectively improved the model’s feature representation capacity while maintaining computational efficiency.

4.2. Ablation Experiment

Compared with the original YOLOv8n detection model, the RMB yield detection and RMB counting model proposed in this study has been improved in three aspects. Part I : The convolution RFAConv with receptive field attention mechanism replaces the conventional convolution module in the backbone network; Part II : The neck network adopts an improved BIFPN structure to enhance the multi-scale feature fusion capability; Part III : In the head detection network, the DynamicHead detection head is introduced to improve detection accuracy and robustness. In order to verify the contribution of various improvements to the overall model performance, this study conducted ablation experiments on the RMB yield detection and RMB counting models under complex backgrounds. The results are shown in Table 2.

Table 2. Ablation Experiment Results. A represents RFAConv, B represents BIFPN, and C represents DySample. The bold font represents the best data in this indicator.

Models	mAP50/%	mAP95/%	Precision/%	Recall/%	Gflows/G
Yolov8n	92.1	86.9	87.4	82.2	6.9
Yolov8n+RFA	92.7	87.2	86.5	83.0	8.2
Yolov8n+BIFPN	93.4	88.2	89.2	83.4	6.9
Yolov8n+DySample	93.3	88.4	90.4	85.4	6.8
Yolov8n+RFA+BIFPN	92.5	87.5	87.4	85.6	8.3
Yolov8n+RFA+DySample	94.1	88.8	90.4	84.9	7.0
Yolov8n+BIFPN+DySample	93.2	87.5	88.3	90.0	7.0
Yolov8n-RMB	94.9	89.7	91.3	91.9	8.3

After incorporating the RFA convolution module into the backbone network, the model exhibited a notable improvement in feature extraction under complex backgrounds. Compared to the unmodified baseline, the model enhanced with RFA convolution achieved an increase in recall from 82.2% to 83.0%, and an improvement in mAP@0.5 by 0.6%. Although the precision decreased slightly from 87.4% to 86.5%, the overall gain in recall indicates that RFA convolution effectively enlarges the receptive field and adaptively focuses on objects of varying scales and shapes. This enhancement is effective in scenes with strong background interference. It increases the model’s capacity to recognize detailed characteristics of RMBs. However, the increased feature sensitivity may have also led to a marginal rise in false positives, slightly lowering the precision. Introducing the BiFPN structure into the neck network optimized multi-scale feature fusion and transmission. In experiments using only BiFPN, precision improved to 89.2%, mAP@0.5 reached 93.4%, and recall increased to 83.4%, all without increasing the model’s parameter GFlows. These results shows that BiFPN, with its learnable fusion weights, effectively enhances the expressive power of features while maintaining computational efficiency. By adaptively weighting feature importance, BiFPN enables robust detection across varying scales. Further improvements were observed when the DySample module was integrated into the upsampling process. Both precision and recall increased significantly, reaching 90.4% and 85.4%, respectively. Moreover, improvements were also observed in mAP@0.5 and mAP@0.5:0.95. These gains affirm that DySample enhances feature resolution recovery and spatial alignment, especially under complex background conditions. The content-aware dynamic sampling mechanism significantly improves feature reconstruction and overall detection accuracy in challenging environments.

When submodules were combined, synergistic effects became more apparent. The integration of RFA and BiFPN further increased recall but caused a slight drop in precision, suggesting that while these modules jointly enhance fine-grained feature modeling and multi-scale fusion, further refinement is needed to balance false positive suppression. The combination of RFA and DySample yielded precision of 90.4% and mAP@0.5 of 94.1%, demonstrating that the interplay between high-level feature extraction and adaptive upsampling substantially enhances detection performance in complex scenes. The BiFPN-DySample configuration significantly improved recall to 90.0%,

underscoring the complementary strengths of feature fusion and dynamic sampling in recall-oriented tasks.

The fully integrated model incorporating RFA, BiFPN, and DySample achieved the best overall performance: mAP@0.5 reached 94.9%, mAP@0.5:0.95 achieved 89.7%, and precision and recall improved to 91.3% and 91.9%, respectively. These results confirm that the proposed YOLOv8n-RMB framework effectively optimizes feature extraction, feature fusion, and detection decision-making across the backbone, neck, and head stages. The combined improvements significantly enhanced the model's accuracy, robustness, and practical applicability in RMB detection and yield estimation under complex backgrounds, all while maintaining low parameter of GFlows count and computational cost.

Figure 14 provides an intuitive visualization of performance trends across different model configurations during training. In the first 30 epochs, all metrics—including precision, recall, and mAP—rose rapidly, indicating strong convergence behavior. The improvements introduced did not cause training instability despite the increased architectural complexity. As shown in Figure 14(a), the YOLOv8n-RMB consistently maintained the highest precision throughout training, with superior stability emerging after 50 epochs. Figures 14(b) shows that YOLOv8n-RMB also achieved the highest recall, with earlier and more substantial gains, verifying the effectiveness of BiFPN and DySample in detecting small objects. Figures 14(c) and 5(d) present the trends of mAP@0.5 and mAP@0.5:0.95, respectively. The full model incorporating RFA, BiFPN, and DySample not only attained rapid improvements under the looser mAP@0.5 criterion but also maintained its lead under the more stringent mAP@0.5:0.95 standard. This confirms its superiority in both localization accuracy and classification robustness. Furthermore, the volatility in early training stages was lower for all improved models compared to the baseline, indicating that the proposed architectural modifications improved convergence stability without introducing negative side effects during training.

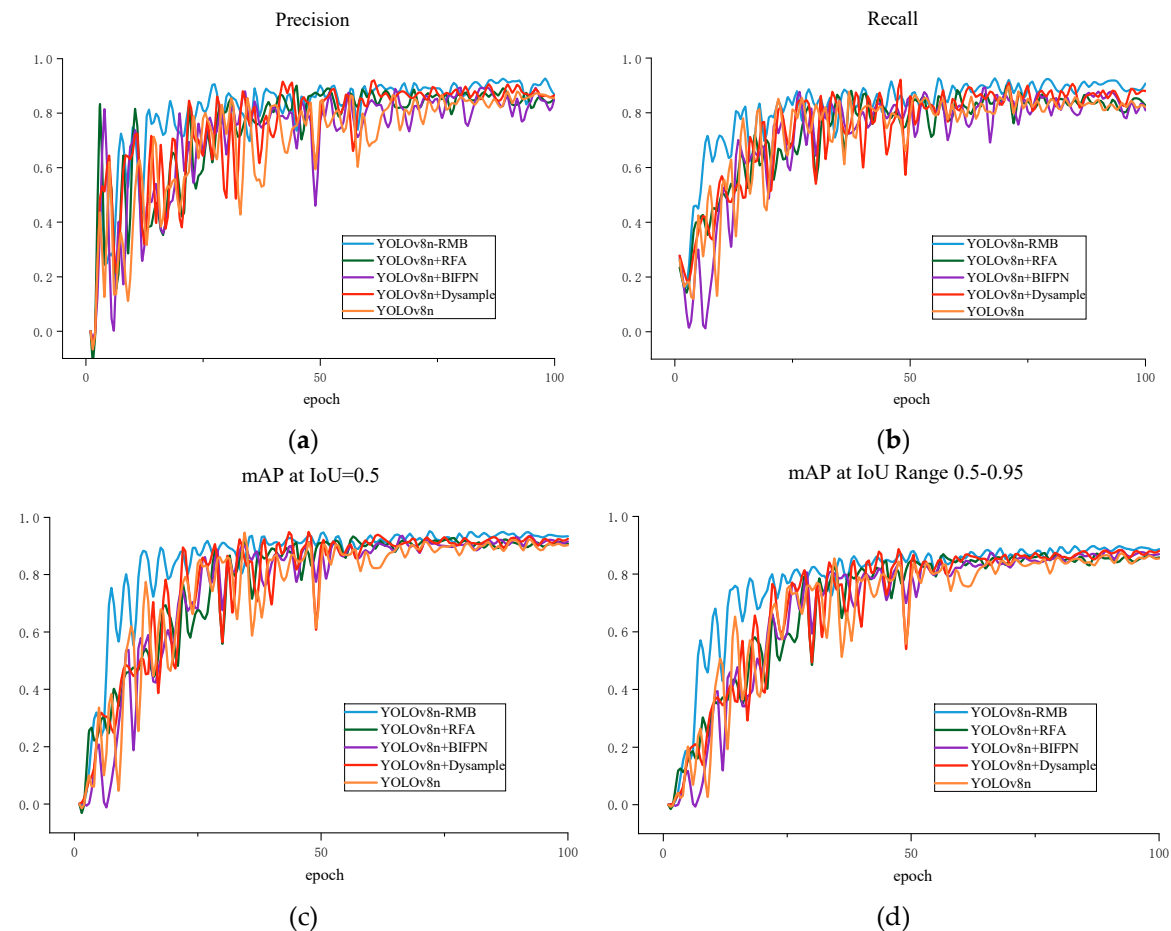


Figure 14. Results of each experiment. (a) The change curve of precision, (b) The change curve of recall, (c) The change curve of mAP (0.5), (d) The change curve of mAP (0.5:0.95).

4.3. Comparison Between Different Models

To comprehensively evaluate the effectiveness of the proposed YOLOv8n-RMB model in the RMB detection task, a comparative analysis was conducted against several state-of-the-art lightweight object detection models, including RT-DETR, YOLOv3-Tiny, YOLOv5n, YOLOv10n, and YOLOv8n. As shown in Table 3, the YOLOv8n-RMB model outperformed all counterparts across all evaluation metrics, achieving a Precision of 91.3%, Recall of 91.9%, mAP@0.5 of 94.9%, and mAP@0.5:0.95 of 89.7%. Compared to the baseline YOLOv8n, the YOLOv8n-RMB-enhanced variant achieved an improvement of 2.8% in mAP@0.5 and 2.8% in mAP@0.5:0.95, highlighting the significant gains in detection accuracy brought by the proposed architectural enhancements.

Table 3. Performance comparison of different models with proposed YOLOv8n-RMB.

Models	mAP50/%	mAP95/%	Precision/%	Recall/%	Gflows/G
RT-DETR	84.3	78.3	87.4	81.0	103.4
Yolov3-Tiny	90.6	81.4	84.5	85.1	14.3
Yolov5n	89.7	79.6	89.7	77.7	5.8
Yolov10n	91.3	85.7	86.0	83.2	8.2
Yolov8n	92.1	86.9	87.4	82.2	6.9
YOLOv8n-RMB	94.9	89.7	91.3	91.9	8.3

Despite these improvements, the YOLOv8n-RMB model maintained a lightweight design, with a computational cost of only 8.3 GFLOPs, marginally higher than YOLOv8n (6.9 GFLOPs), but substantially lower than RT-DETR (103.4 GFLOPs), demonstrating its superior suitability for deployment in resource-constrained environments. While YOLOv5n and YOLOv3-Tiny showed moderate performance, both Precision and Recall were consistently lower than those of YOLOv8n-RMB, and a marked performance drop was observed under the stricter mAP@0.5:0.95 metric (79.6% and 81.4%, respectively), indicating reduced robustness in complex scenes and near-boundary object detection. YOLOv10n provided a balance between accuracy and complexity but showed lower mAP and Recall compared to YOLOv8n-RMB. Notably, RT-DETR exhibited the highest computational cost among all models and failed to demonstrate any significant performance advantage in this task.

In summary, the YOLOv8n-RMB model provides high detection accuracy and robustness while keeping computational cost low. This makes it suitable for real-time, high-precision deployment in UAV-based RMB recognition tasks. Its strong performance under complex environmental conditions validates its effectiveness for intelligent rubber yield assessment and precision plantation management.

4.4. Rubber Bowl Production Status Counting Experiment

To evaluate the practical counting performance of the proposed RMB status detection and counting model, three field experiments were conducted in the same natural rubber plantation involving a total of 590 bowls. In each experiment, the number of RBMs in the three states, Empty, Not Filled, and Filled, was first recorded by manual counting and the YOLOv8n-RMB-based detection and counting model was then used to automatically identify and count the same RBMs in real time using UAV images. The model’s output was compared with the manually recorded ground truth counts. The results of the experiment are listed in Table 4.

Table 4. Counting test results, proposed UAV imagery method vs manual.

Counting Method	None	NotFilled	Filled	Rate(%)	Time(min)
Test1	99.2%	69	479	100	43

Test2	Artificial	89.0%	53	491	100	47
Test3		88.0%	119	463	100	38
Test1	YOLOv8n-RMB	99.5%	69	465	97.6	18
Test2		93.1%	53	482	98.4	24
Test3		91.9%	133	454	97.6	20

The experimental results demonstrate that the proposed YOLOv8n-RMB model exhibits strong practicality and generalization capability in complex rubber plantation environments. Compared with manual counting, the model reached accuracy rates of 97.6%, 98.4%, and 97.6% in three field trials for identifying and counting RMB states. In terms of efficiency, the YOLOv8n-RMB model needed about 21 minutes on average to complete one full counting task per scene, while manual counting took around 43 minutes. This result shows more than twice the efficiency.

In addition, using real-time video streams from UAVs, the YOLOv8n-RMB model shows better adaptability and easier deployment compared to traditional methods based on static images. It makes the model suitable for large-scale unmanned rubber-tapping operations. The integration of RFAConv, BiFPN, and DySample modules significantly improves both accuracy and inference speed, confirming the model’s potential for real-world engineering applications. The deployment of this model offers technical support for rubber yield monitoring and intelligent tapping systems. It also supports the development of forestry informatization and the modernization of the rubber industry.

5. Conclusions and Future Work

This study proposes a novel RMB yield recognition and counting model, YOLOv8n-RMB, designed for UAV-based video detection tasks in rubber plantations with complex forest backgrounds. First, a RFAConv is added to the backbone to address the limited ability of standard shared-parameter convolution kernels in capturing long-range features. This significantly enhances the model’s ability to perceive edge details and subtle textures of RMB under complex conditions. Second, BiFPN is introduced to improve fusion between shallow and deep semantic features. It helps reduce feature degradation and remove redundant information. Finally, a content-aware dynamic resampling module, DySample, is incorporated during the upsampling stage to achieve precise recovery of fine-grained spatial details and clearer classification boundaries for RMB states. In real-world scenarios characterized by diverse backgrounds, variable object scales, and high visual similarity between RMBs and environmental elements, the proposed model successfully identifies and counts three distinct RMB states—empty, partially filled, and fully filled. This provides robust visual support for target selection and yield estimation in automated rubber tapping operations.

(1) Ablation Study: Compared with the baseline YOLOv8n model, YOLOv8n-RMB achieves substantial improvements in all detection metrics, with mAP@0.5, mAP@0.5:0.95, Precision, and Recall increasing by 2.8%, 2.9%, 3.9%, and 9.7%, respectively. These gains are attributed to BiFPN’s enhanced feature expression via frequent bidirectional fusion and DySample’s dynamic sampling mechanism for optimized spatial reconstruction. Although the model’s parameters increased from 2.7M to 3.0M and FLOPs rose from 6.9G to 8.3G, the inference speed stayed high at 91 FPS. This meets the real-time detection needs in UAV-based rubber plantation monitoring. These results show that the model maintains a balance between detection accuracy and speed, indicating its applicability in real field conditions.

(2) Comparative Evaluation: In comparison with several mainstream lightweight detectors, the YOLOv8n-RMB model achieves the highest mAP@0.5 and mAP@0.5:0.95 scores, at 94.9% and 89.7%, respectively. Visualized results demonstrate its ability to stably and accurately detect the three RMB states with clear target boundaries and a low false detection rate, outperforming baseline models such as YOLOv8n and YOLOv5n. While RT-DETR shows low detection accuracy (84.3% mAP@0.5, 78.3% mAP@0.95) and its large computational cost (103.4 GFLOPs) makes it unsuitable for real-time UAV deployment on resource-constrained edge devices. These findings further validate YOLOv8n-

RMB's superiority in balancing detection precision and computational efficiency, making it highly suitable for practical deployment in RMB yield monitoring tasks.

(3) Field Counting Experiments: In three field tests, the YOLOv8n-RMB model achieved recognition and counting accuracies of 97.6%, 98.4%, and 97.6%, respectively, demonstrating strong generalization and adaptability across real plantation conditions. Furthermore, the model significantly outperforms manual counting in terms of time efficiency, completing each task in an average of 21 minutes compared to 43 minutes manually—more than doubling the operational efficiency. By processing UAV video streams in real time and using the BoT-SORT multi-object tracking algorithm, the model performs automated detection and counting of RMBs in three states. These results show that the method ensures high counting accuracy and reduces labor demand.

Despite the promising performance, the proposed counting method still exhibits instances of false negatives and false positives. Moreover, the current experiments were conducted exclusively in rubber plantations located in Hainan and Yunnan Province, China. Given the substantial variability in planting density, trunk morphology, lighting conditions, and background interference across different regions, the generalizability of the model to other rubber plantations remains to be validated. Future work will focus on expanding the dataset to encompass more diverse environmental conditions, thereby enhancing the model's feature learning capacity. Additionally, further architectural refinements will be explored to reduce both false detection and missed detection rates.

Author Contributions: Conceptualization, Y.W. and L.Y.; methodology, Y.W., L.Y., and P.Z.; software, Y.W., P.Z.; validation, C.S., Y.Z., and A.H.; formal analysis, X.Y. and Y.W.; investigation, L.Y.; resources, L.Y.; data curation, Y.W., C.S., and L.Y.; writing—original draft preparation, Y.W. and L.Y.; writing—review and editing, L.Y., Y.W., and X.Y.; visualization, Y.W., L.Y.; supervision, L.Y.; project administration, L.Y.; funding acquisition, L.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Technology Project, grant number NK2022160603; and the Fundamental Research Funds for the Central Universities, grant number 2662025GXPY004.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, Y.; Yang, H.; Liu, J.; Zhang, Z.; Zhang, X. Tapping Line Detection and Rubber Tapping Pose Estimation for Natural Rubber Trees Based on Improved YOLOv8 and RGB-D Information Fusion. *Sci. Rep.* **2024**, *14*, 28717, doi:10.1038/s41598-024-79132-5.
2. Yong, J.; Lu, G.; An, Y.; Lang, S.; Zhang, H.; Chen, R. Characterization of Cis-Polyisoprene Produced in *Periploca Sepium*, a Novel Promising Alternative Source of Natural Rubber. *Commun. Biol.* **2025**, *8*, 1–11, doi:10.1038/s42003-025-07734-4.
3. Wang, L.; Huang, C.; Li, T.; Cao, J.; Zheng, Y.; Huang, J. An Optimization Study on a Novel Mechanical Rubber Tree Tapping Mechanism and Technology. *Forests* **2023**, *14*, 2421, doi:10.3390/f14122421.
4. Zhang, X.; Ma, W.; Liu, J.; Xu, R.; Chen, X.; Liu, Y.; Zhang, Z. An Improved YOLOv8n-IRP Model for Natural Rubber Tree Tapping Surface Detection and Tapping Key Point Positioning. *Front. Plant Sci.* **2024**, *15*, 1468188, doi:10.3389/fpls.2024.1468188.
5. Xirui, Z.; Chao, C.A.O.; Li'na, Z.; Jiejie, X.; Junxiao, L.I.U.; Xuehu, D.O. Design and Test of Profiling Progressive Natural Rubber Automatic Tapping Machine. *Nongye Jixie Xuebao/trans. Chin. Soc. Agric. Mach.* **2022**, *53*.

6. Ning, T.; Liang, D.; Zhang, Y.; Fu, W.; Ru, S. Design and Experimental Research of Fixed Compound Motion Track Rubber Tapping Machine. *J. Southwest Univ. Nat. Sci. Ed.* **2022**, *44*, 100–109.
7. Xiongwei, W.; Guisheng, G.; Fucheng, L.I.; Xiaofeng, Z.; Zhichao, G. Design of Fixed Automatic Intelligent Control Rubber Tapping Machine. *Agric. Eng.* **2020**, *10*, 79–84.
8. Zhou, H.; Zhang, S.; Zhai, Y.; Wang, S.; Zhang, C.; Zhang, J.; Li, W. Vision Servo Control Method and Tapping Experiment of Natural Rubber Tapping Robot. **2021**.
9. Zeng, S.; Wu, Y.; Zeng, F.; Yu, K.; Ma, L.; Yang, W. Design and Test of Automatic Rubber Tapping Machine. *J. Southwest Univ. Nat. Sci. Ed.* **2024**, *46*, 92–102.
10. Wang, C.; Li, H.; Deng, X.; Liu, Y.; Wu, T.; Liu, W.; Xiao, R.; Wang, Z.; Wang, B. Improved You Only Look Once v.8 Model Based on Deep Learning: Precision Detection and Recognition of Fresh Leaves from Yunnan Large-Leaf Tea Tree. *Agriculture* **2024**, *14*, 2324, doi:10.3390/agriculture14122324.
11. Bhumiphan, N.; Nontapon, J.; Kaewplang, S.; Srihanu, N.; Koedsin, W.; Huete, A. Estimation of Rubber Yield Using Sentinel-2 Satellite Data. *Sustainability* **2023**, *15*, 7223, doi:10.3390/su15097223.
12. Chen, G.; Liu, Z.; Wen, Q.; Tan, R.; Wang, Y.; Zhao, J.; Feng, J. Identification of Rubber Plantations in Southwestern China Based on Multi-Source Remote Sensing Data and Phenology Windows. *Remote Sens.* **2023**, *15*, 1228, doi:10.3390/rs15051228.
13. Otsu, K.; Pla, M.; Duane, A.; Cardil, A.; Brotons, L. Estimating the Threshold of Detection on Tree Crown Defoliation Using Vegetation Indices from UAS Multispectral Imagery. *Drones* **2019**, *3*, 80, doi:10.3390/drones3040080.
14. Traore, A.; Ata-Ul-Karim, S.T.; Duan, A.; Sothar, M.K.; Traore, S.; Zhao, B. Predicting Equivalent Water Thickness in Wheat Using UAV Mounted Multispectral Sensor through Deep Learning Techniques. *Remote Sens.* **2021**, *13*, 4476, doi:10.3390/rs13214476.
15. Junos, M.H.; Mohd Khairuddin, A.S.; Thannirmalai, S.; Dahari, M. Automatic Detection of Oil Palm Fruits from UAV Images Using an Improved YOLO Model. *Visual Comput.* **2022**, *38*, 2341–2355, doi:10.1007/s00371-021-02116-3.
16. Evaluation of Cotton Emergence Using UAV-Based Imagery and Deep Learning. *Comput. Electron. Agric.* **2020**, *177*, 105711, doi:10.1016/j.compag.2020.105711.
17. Ong, P.; Teo, K.S.; Sia, C.K. UAV-Based Weed Detection in Chinese Cabbage Using Deep Learning. *Smart Agric. Technol.* **2023**, *4*, 100181, doi:10.1016/j.atech.2023.100181.
18. Hu, G.; Yin, C.; Wan, M.; Zhang, Y.; Fang, Y. Recognition of Diseased Pinus Trees in UAV Images Using Deep Learning and AdaBoost Classifier. *Biosyst. Eng.* **2020**, *194*, 138–151, doi:10.1016/j.biosysteng.2020.03.021.
19. Zhou, X.; Lee, W.S.; Ampatzidis, Y.; Chen, Y.; Peres, N.; Fraisse, C. Strawberry Maturity Classification from UAV and Near-Ground Imaging Using Deep Learning. *Smart Agric. Technol.* **2021**, *1*, 100001, doi:10.1016/j.atech.2021.100001.
20. Zhang, F.; Zhao, L.; Wang, D.; Wang, J.; Smirnov, I.; Li, J. MS-YOLOv8: Multi-Scale Adaptive Recognition and Counting Model for Peanut Seedlings under Salt-Alkali Stress from Remote Sensing. *Front. Plant Sci.* **2024**, *15*, 1434968, doi:10.3389/fpls.2024.1434968.
21. Karakuş, S.; Kaya, M.; Tuncer, S.A. Real-Time Detection and Identification of Suspects in Forensic Imagery Using Advanced YOLOv8 Object Recognition Models. *Trait. Du Signal* **2023**, *40*, 2029–2039, doi:10.18280/ts.400521.

22. Khan, A.T.; Jensen, S.M.; Khan, A.R. Advancing Precision Agriculture: A Comparative Analysis of YOLOv8 for Multi-Class Weed Detection in Cotton Cultivation. *Artif. Intell. Agric.* **2025**, *15*, 182–191, doi:10.1016/j.aiia.2025.01.013.
23. Cai, S.; Zhang, X.; Mo, Y. A Lightweight Underwater Detector Enhanced by Attention Mechanism, GSConv and WIoU on YOLOv8. *Sci. Rep.* **2024**, *14*, 25797, doi:10.1038/s41598-024-75809-z.
24. Liu, H.; Zhang, Y.; Chen, Y. A Symmetric Efficient Spatial and Channel Attention (ESCA) Module Based on Convolutional Neural Networks. *Symmetry* **2024**, *16*, 952, doi:10.3390/sym16080952.
25. Lv, B.; Zhang, S.; Gong, H.; Zhang, H.; Dong, B.; Wang, J.; Du, C.; Wu, J. Pavement Disease Visual Detection by Structure Perception and Feature Attention Network. *Appl. Sci.* **2025**, *15*, 551, doi:10.3390/app15020551.
26. Zhang, X.; Liu, C.; Yang, D.; Song, T.; Ye, Y.; Li, K.; Song, Y. RFACnv: Innovating Spatial Attention and Standard Convolutional Operation Available online: <https://arxiv.org/abs/2304.03198v6> (accessed on 24 April 2025).
27. Sui, J.; Liu, L.; Wang, Z.; Yang, L. RE-YOLO: An Apple Picking Detection Algorithm Fusing Receptive-Field Attention Convolution and Efficient Multi-Scale Attention. *PLOS One* **2025**, *20*, e0319041, doi:10.1371/journal.pone.0319041.
28. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection 2020.
29. Liu, W.; Lu, H.; Fu, H.; Cao, Z. Learning to Upsample by Learning to Sample 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.