

ORIGINAL RESEARCH

Open access

Deep Learning-Based Online Cabbage Recognition under Light Occlusion in Open-Field Environments

Mateo Alvarez^{1*}, Sofia Herrera², Diego Cruz¹, Andres Castro³

Abstract

To address poor accuracy in identifying open-field vegetables when light is blocked, this investigation used cabbage as the subject. It established an online target identification model based on deep learning. Using YOLOv8n as the backbone network, an approach was formulated to reduce the impact of light occlusion on the precision of online cabbage identification. A sequence of cabbage image filters was engineered to remove the consequences of light occlusion. A unit for adaptive parameter acquisition tailored to the cabbage image filters was built. The image filter sequence and the adaptive acquisition unit were woven into the YOLOv8n object detection architecture. This fusion enabled precise, real-time identification of cabbage in light-occluded environments. Experimental outcomes indicated identification precisions of 97.5% with the normal lighting dataset, 93.1% with the light occlusion dataset, and 95.0% with the mixed dataset. When the degree of light occlusion exceeded 0.4, the identification precision increased by 9.9% and 13.7% compared with the YOLOv5n and YOLOv8n architectures. The model achieved identification precisions of 99.3% with the Chinese cabbage dataset and 98.3% with the broccoli dataset. The model was deployed on an Nvidia Jetson Orin NX edge device, where it processed images at 26.32 frames per second. Field assessments showed identification precisions of 96.0% in ordinary light and 91.2% when light was blocked. The presented online cabbage identification model facilitates instant recognition and spatial location of cabbage in complex open-field conditions, providing technical support for targeted spraying.

Keywords Object detection, Cabbage, Light occlusion, Online recognition

*Correspondence:

Mateo Alvarez

mateo.alvarez@unal.edu.co

¹ Department of Open-Field Crop Recognition, Faculty of Agriculture, National University of Colombia, Bogota, Colombia

² Department of Occlusion-Tolerant Deep Learning, Faculty of Agriculture, University of Chile, Santiago, Chile

³ Department of Cabbage Detection for Precision Harvesting, Faculty of Agriculture, University of Antioquia, Medellin, Colombia

Introduction

A swelling global population and heightened demand for food and vegetables create serious hurdles for crop production on scarce arable land. Cabbage is among the vegetables grown on the largest scale worldwide. Figures published by the Food and Agriculture Organization (FAO) show that, in 2021, the worldwide area devoted to cabbage harvesting was 2.4506 million hectares, and the quantity produced reached 71.7072 million tons. Diseases and pests pose principal threats to cabbage quality and yield, and the use of chemical pesticides remains one of the most effective control tactics. At present, blanket continuous spraying is widely used in cabbage cultivation to suppress pests and diseases. Throughout the early and middle development periods, cabbage crops exhibit large gaps between rows and between individual plants, leading to poor pesticide uptake when conventional application procedures are used. Research indicates that only about 30% of open-field-sprayed pesticides adhere to vegetation.

In comparison, the remaining 60%–70% settles into the soil or drifts into the air [1], resulting in chemical waste and ecological contamination. Target-focused spraying technology constitutes a potent remedy for these difficulties [2]. By employing immediate target sensing and precise regulation, this technique ensures that chemicals land precisely on the intended targets.

Precise identification of targets stands as a prerequisite for target-focused spraying in open-field cabbage. The chief technologies for detecting open-field crop targets include ultrasonic, LiDAR, and machine vision sensing [3–5]. Ultrasonic sensors and LiDAR systems detect targets by interpreting reflected ultrasonic or laser waves. Yet they fall short of capturing rich semantic data, complicating the separation of crops from weeds. Machine vision technology extracts crop images packed with semantic clues. It deploys image manipulation routines rooted in machine learning or deep learning to identify crop traits and recognize and sort targets in the images. This tactic yields marked benefits for identifying open-field cabbage.

Image manipulation methods grounded in machine learning require the handcrafted extraction of features such as texture, form, and edges. These characteristics are subsequently plugged into machine learning categorizers to perform sorting and identification [6]

6. [7] pulled three texture characteristics from the gray-level co-occurrence matrix (GLCM) of lettuce pictures. They harnessed a support vector machine (SVM) categorizer to group these characteristic sets, achieving a mean classification accuracy of 94.73% for single plants observed against a simple background. Machine learning procedures require manual feature mining for crop sorting. In exceptionally tangled field situations, the exhaustive extraction of useful target traits is difficult, \ procedures rooted in machine learning are generally restricted to sorting for only one target category per picture, hindering the complete, end-to-end spatial localization of crop targets [8].

Object detection frameworks in deep learning extract image features using feature extraction architectures and then apply regression models to estimate target locations. This strategy operates as a potent means for both target recognition and spatial location in imagery [9, 10]. A large number of investigators have undertaken wide-ranging inquiries into target recognition schemes built upon object detection frameworks. [11] developed a sensing model for hydroponically grown lettuce seedlings by refining the feature extraction architecture and loss function, achieving 86.2% precision on a greenhouse-collected dataset. [12] used a YOLOv5 model integrated with attention mechanisms to pinpoint *Solanum rostratum* Dunal seedlings under steady illumination, achieving a precision of 94.65%. [7] developed a rapeseed target-sensing model based on YOLOv5s, achieving 99.6% recognition accuracy under specific indoor illumination conditions. The investigations noted above all involved offline image manipulation under static, regulated illumination. How these models behave has not yet been verified in unregulated field settings, especially those characterized by patchy illumination.

Shifts in the sun's angle of elevation, together with shadows thrown by operational machinery in unregulated field settings, repeatedly bring about light occlusion inside the camera's viewing area. This reduces the target's useful traits, thereby lowering recognition precision. Putting together setups and procedures that maintain high recognition precision is fundamental to reliable target spotting when light is blocked. Several scholars and robotics enterprises have assembled image-capturing setups using shrouds and powered light banks to provide steady illumination for cameras. placed cameras within a spacious LED-lit enclosure to build a cabbage-and-weed spotting setup, achieving a precision of 93.5%. [13, 14], and Blue River Technology have rolled out spotting setups that depend on powered light banks [15]. Depending on powered light banks to offset light occlusion and patchy brightness calls for kitting out machinery with bulky enclosures, culminating in hefty apparatus with reduced nimbleness, higher energy draw, and shorter battery endurance. Several scholars have proposed recognition schemes to improve model performance under low-light conditions. [16] switched color spaces and developed color metrics for images fed to deep learning models, thereby bolstering performance under varying lighting conditions. Several scholars have applied adaptive histogram equalization schemes to enhance images, thereby reducing the impact of natural outdoor illumination on recognition accuracy. [17–19]. [20, 21] introduced methods grounded in Retinex theory to reduce patchy brightness, thereby improving the accuracy of target

detection in deep learning models. The picture manipulation schemes cited earlier, including adaptive histogram equalization and Retinex-rooted methods, require manual setting of hyperparameters. A solitary preset value cannot adapt to shifting outdoor illumination, leading to weak stability and long manipulation intervals, which consequently hinder real-time spotting.

To address the challenges outlined, this work developed a real-time cabbage spotting model based on YOLOv8n. It knitted together a sequence of cabbage picture filters and an adaptive parameter-learning unit, working to reduce the footprint of light occlusion on spotting accuracy. The model funnels incoming cabbage images through the adaptive parameter-learning unit to determine the required hyperparameters for the filter set. The adaptive parameter learning unit is woven directly into the YOLOv8n object detection model, with the same loss formula applied for joint tuning during training. Cabbage pictures, once filtered, are fed into the YOLOv8n object detection model to tease out target details associated with cabbage in the pictures. The online cabbage-spotting model proposed in this paper meaningfully reduces the footprint of light occlusion and provides robust technical support for subsequent target-focused cabbage spraying.

Materials and Methods

Development of the cabbage dataset

From April 2021 through September 2024, imagery of cabbage was acquired at the National Precision Agriculture Research and Demonstration Base in Xiaotangshan Town, Beijing (40°10'36" N, 116°26'44"E). Sampling focused on specimens from the fifth day after transplanting through the early rosette phase. Throughout this window, individual plants stayed spatially separated, yielding a corpus of 5694 RGB captures. Two camera models served as collection instruments: an Intel RealSense D455, which output frames at 1280 × 800 pixels, and a Logitech C930, which recorded at 1920 × 1080 pixels. The imaging sensor remained coplanar with the soil surface throughout all acquisitions. Elevating the camera to 110 cm ensured that four cabbage rows along a single bed fell within the camera's field of view. Both cabbage specimens and weeds were present across the gathered pictures; the composition of the resulting dataset is detailed in [Table 1](#). A Transformer-driven recognition architecture established in prior research was employed for preprocessing the raw captures. Inference outputs from that model yielded cabbage bounding boxes, which were persisted as .txt annotations conforming to the Yolo label specification. Labeling was then used to correct any erroneous tags and manually fill in missing ones. A subset of 530 captures characterized by light occlusion was selected as the light occlusion test collection. OpenCV routines were used to manipulate occlusion severity, which was quantified using a luminance ratio that linked shadowed patches to brightly lit patches. This procedure generated multiple test groupings, each embodying a distinct occlusion intensity. Another 517 uniformly illuminated captures constituted the even-lighting test collection. Merging these two produced the mixed test collection. The remaining 4647 captures were partitioned at a 9:1 ratio into training and validation sets. Network weights were fitted on the training segment, post-training generalization performance was gauged on the validation segment, and the hyperparameters guiding training were refined on the test segment to elevate model efficacy. The pipeline spanning image gathering and dataset assembly is portrayed in [Figure 1](#).

Table 1. Dataset composition.

Data type	Different growth periods/day											Different lighting	
	5	12	17	19	22	25	27	29	30	33	42	Light occlusion	Normal lighting
Number	477	855	779	355	330	658	118	820	904	183	215	1595	4099

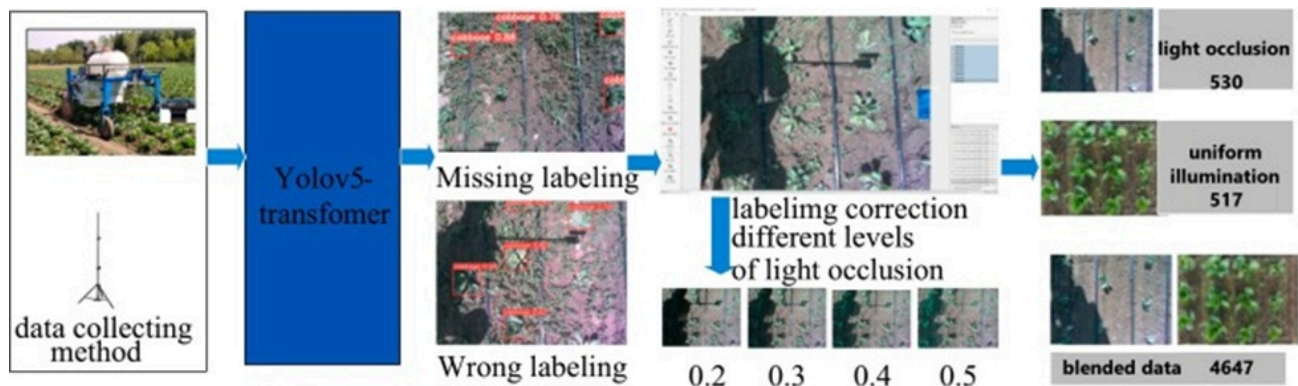


Figure 1. Data set production process.

Construction of a cabbage online recognition model

Design of a cabbage image filter combination to mitigate the effects of light occlusion

Variations in the sun's altitude during field acquisition often cast shadows across portions of individual frames, undercutting recognition fidelity. To counteract this, a bespoke chain of cabbage image filters was engineered to suppress the detrimental effect of light blockage on detection scores. The chain depicted in **Figure 2** combines three elements: a linear adjustment filter, a gamma correction filter, and a sharpening filter. The linear adjustment filter primarily controls the overall exposure of the imagery. The gamma correction filter addresses regions that are washed out or overly dark. The sharpening filter amplifies edge definition, strengthening the prominence of contours, textural patterns, and other salient cues in both the cabbage foreground and the surrounding backdrop.

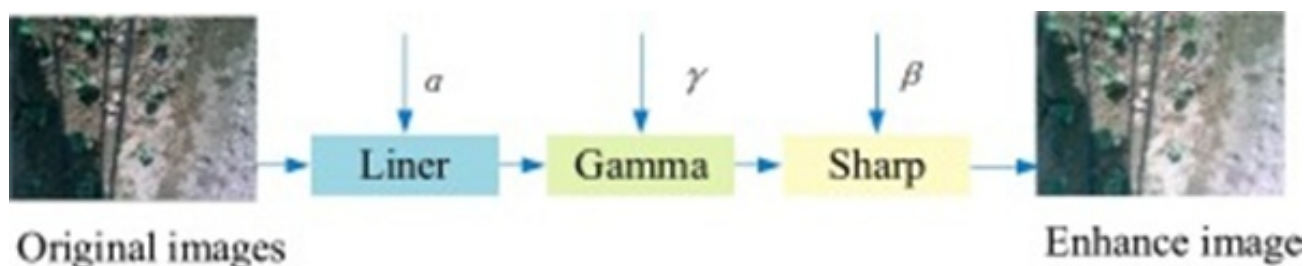


Figure 2. Cabbage image filter combination.

The filter chain gets embedded directly into the detection architecture, with gradient descent driving the optimization during the training cycle. Differentiability represents a necessary design requirement for these filters. Following **Eq. 1**, the linear adjustment filter applies a linear mapping to rescale global luminance. Details within the frame become more conspicuous, increasing the likelihood that the model exploits informative signals. The coefficient α holds considerable weight; a poor choice can push the frame into overexposure or severe underexposure. In the present work, α is not held fixed but instead adapts to image-level characteristics, broadening the filter's operational envelope.

$$G(x, y) = H(x, y) * e^{\alpha} \quad (1)$$

Within this expression, $G(x, y)$ corresponds to the resultant frame, $H(x, y)$ corresponds to the source frame, and α designates the coefficient governing the linear adjustment filter.

The gamma correction filter operates according to **Eq. 2**. Through an exponential nonlinearity, it selectively remaps pixel intensities in zones affected by over- or underexposure—the steepness of this remapping is determined by the parameter

γ . Values of γ exceeding 1 amplify cabbage detail populating brightly illuminated image sectors. Values falling short of 1 do the same for details within shadowed image sectors. By judiciously setting γ , the gamma filter harmonizes illumination disparities across different image regions to a meaningful degree. The result is a more uniform light distribution across the globe. Minute features inside bleached-out or heavily shaded patches are reinforced, while the integrity of features inside normally lit patches is preserved. Determination of γ proceeds through adaptive learning of the image's intrinsic properties, carried out by the parameter adaptive learning module.

$$G(x, y) = H(x, y)^\gamma \quad (2)$$

Within this expression, $G(x, y)$ corresponds to the resultant frame, $H(x, y)$ corresponds to the source frame, and γ designates the parameter of the gamma correction filter.

The sharpening filter's formulation appears in [Eq. 3](#). First, Gaussian smoothing is imposed on the cabbage frame to strip away high-frequency disturbances tied to irregular lighting and specular leaf reflections. This operation isolates a low-frequency base layer. Subtracting that base layer from the original frame yields a high-frequency residual map that encodes fine-scale plant details. That residual carries edge contour signatures and venation textural patterns, both of which rank among the primitive features that shallow stages of convolutional architectures latch onto. The residual map is then reintegrated with the original frame through a weighted summation governed by a fusion ratio. The outcome is an augmented frame that accentuates intricate plant characteristics. This enrichment expands the neural network's pool of actionable signals and improves recognition fidelity. The fusion-scale parameter β is dynamically tuned to the frame's feature makeup by the parameter-adaptive learning module.

$$G(x, y) = H(x, y) + \beta(H(x, y) - \text{Gau}(H(x, y))) \quad (3)$$

Within this expression, $G(x, y)$ corresponds to the resultant frame, $H(x, y)$ corresponds to the source frame, Gau signals the Gaussian smoothing operator, and β denotes the parameter tied to the sharpening filter.

Filter parameter adaptive learning module

Legacy image processing pipelines hinge on hand-tuned filter settings to handle cabbage frames marred by light obstruction. However, real-time manual intervention is infeasible once models are fielded operationally. To bridge this gap, the current investigation introduces an adaptive filter parameter-learning module based on convolutional neural networks (CNNs). This module is composed of a stack of convolutional layers. The filters themselves act upon low-level image attributes—luminosity, chromaticity, texture, and edge structure—within the cabbage frames [\[22\]](#). The necessary parameter values are synthesized by first pulling these same low-level attributes from the imagery. A CNN built with as few as 5 convolutional tiers is already capable of reliably extracting color, texture, and contour cues [\[23\]](#). Given the cabbage recognition pipeline's latency constraints, minimizing the tier count was a priority. A side-by-side appraisal was conducted to assess the parameter-generation fidelity of stacks with 5, 6, and 7 convolutional tiers. A count of 6 was ultimately settled upon; the precise layout is provided in [Figure 3a](#). To reduce the processing overhead introduced by the module, incoming color cabbage frames are first resized to 256×256 pixels. The frame is then funneled through a succession of 6 depthwise-separable convolutions that progressively reduce its spatial dimensions, with the feature volume shapes evolving as $128 \times 128 \times 16$, $64 \times 64 \times 32$, $32 \times 32 \times 64$, $16 \times 16 \times 128$, $8 \times 8 \times 128$. For the ultimate layer, an 8×8 kernel performs the computation, emitting the three required filter parameters as a $1 \times 1 \times 3$ tensor. The depthwise-separable convolution scheme first applies a 2D kernel independently to each channel of the color frame to isolate single-channel cabbage features, then applies a $1 \times 1 \times M$ kernel to fuse information across the separate channel-specific feature maps, as illustrated in [Figure 3b](#). The module is placed before the YOLOv8n backbone feature extractor, enabling the entire assembly to be trained end-to-end in a single pass.

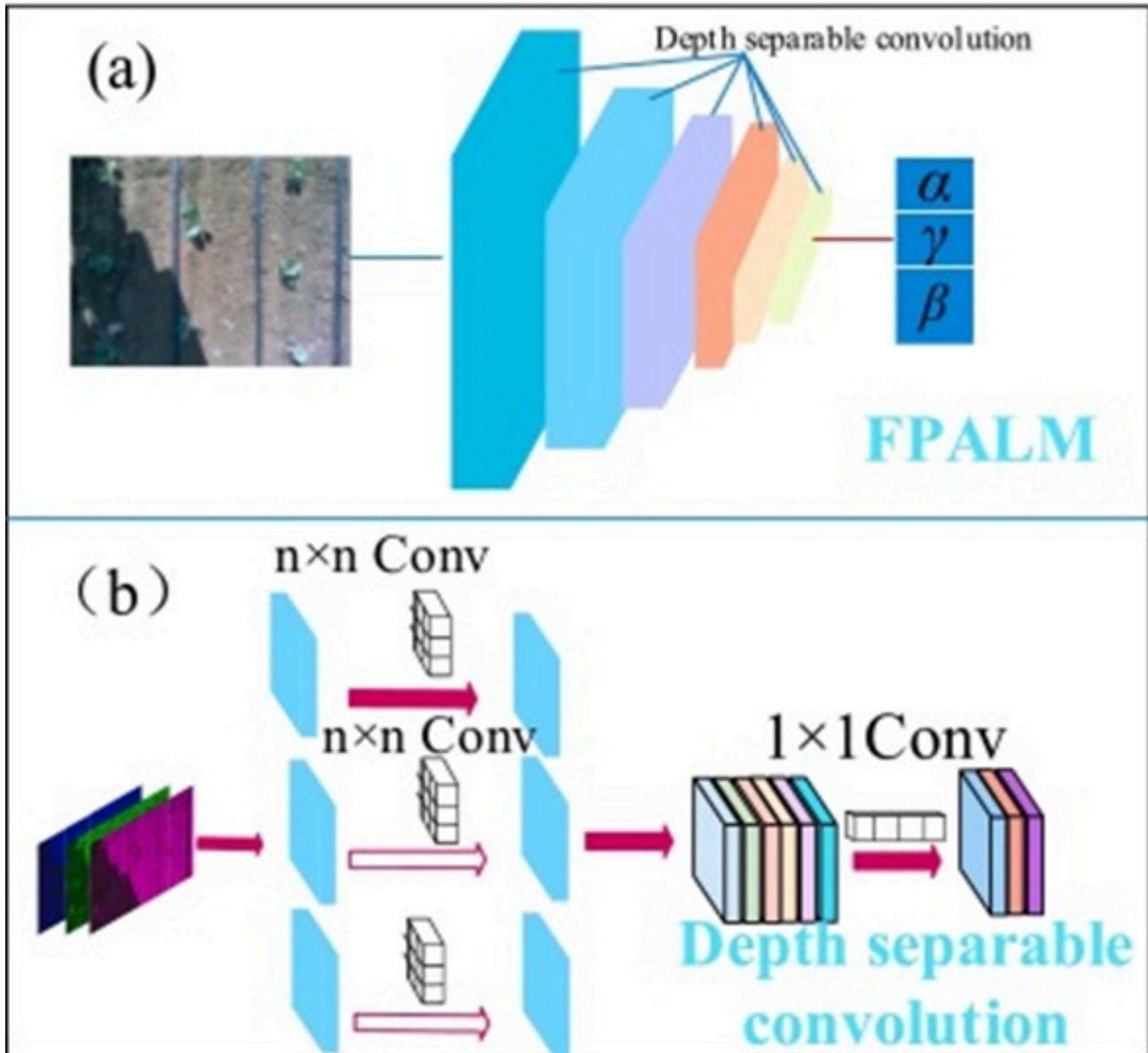


Figure 3. Filter parameter adaptive learning module.

Overall structure of the online cabbage recognition model

Depicted in **Figure 4** is the end-to-end architecture of the online cabbage recognition system devised in this investigation. Three interconnected subsystems form the pipeline: an adaptive module that learns filter parameters, a cascade of image filters, and the YOLOv8n-based cabbage detection module. Live imagery arriving from the camera is routed in parallel to both the parameter-learning module and the filter cascade. Upon receiving a frame, the parameter-learning module first shrinks its spatial dimensions to 256×256 pixels, then passes it through a hierarchy of convolutional layers. A concluding fully connected layer then projects the learned representation into the three scalar filter coefficients needed downstream. In the parallel branch, the untouched frame passes through the filter cascade, where linear exposure correction, gamma-based tonal remapping, and edge sharpening are applied in sequence. The resulting visually enhanced frame, now richer in cabbage-defining features, is handed off to the YOLOv8n module, which performs the final detection. The entire assembly is trained by feeding imagery through the parameter-learning module, onward through the filter cascade, and finally into the YOLOv8 detector. Throughout optimization, the parameter-learning module is tightly coupled with YOLOv8, effectively becoming a native subcomponent of YOLOv8n. A single unified loss function and a shared optimizer

govern simultaneous weight updates across all parts, enabling genuine end-to-end training and ensuring that the filter parameters evolve in lockstep with the detection backbone.

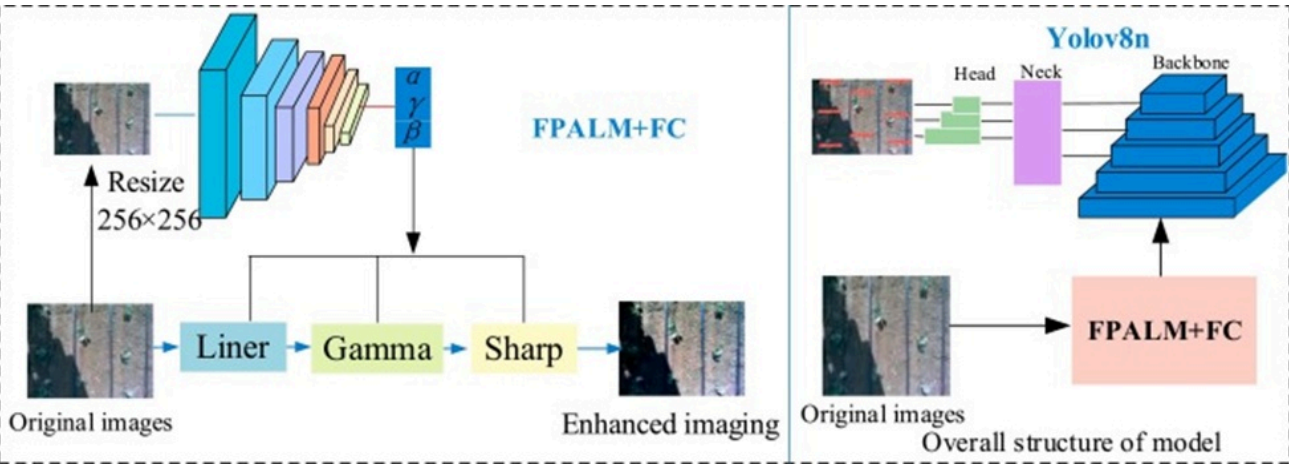


Figure 4. Overall structure of the cabbage recognition model.

The YOLOv8n-powered detection module itself comprises three functional blocks: a backbone for feature extraction, a Neck block, and a detection head. The backbone's job is to distill hierarchical cabbage feature representations from the incoming enhanced frames. At its terminal stage, a Spatial Pyramid Pooling Fast (SPPF) unit applies max pooling across three kernel scales and then fuses the pooled maps with the original, unpooled volume. This multi-scale aggregation expands the representational richness of the features, yielding improved detection robustness across cabbages that differ markedly in apparent size due to varying growth stages. The Neck block weaves together Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN). Through its top-down pathway, the FPN harvests cabbage-related signals spanning from early, shallow layers to late, deep layers at multiple resolutions, thereby sharpening the model's sensitivity to small cabbage instances. Complementing this, the PAN's bottom-up pathway merges coarse semantic information with fine-grained spatial detail from the cabbage frames, reinforcing the positional cues that anchor bounding boxes. This dual-path fusion acts as a safeguard against the progressive dilution of spatial detail that repeated convolution and downsampling can cause. From the Neck's outputs, the detection head produces predictions at three distinct grid resolutions relative to the input dimensions. Every head yields a set of candidate bounding boxes, each tagged with a confidence score. A Non-Maximum Suppression (NMS) procedure subsequently filters these candidates, discarding redundant boxes and retaining the single highest-confidence prediction per cabbage.

Model training and performance validation

Training platform specifications

Implementation, training, and inference of the cabbage recognition model were carried out within the deep learning ecosystem comprising PyTorch 1.7 and Torchvision 0.8. A complete enumeration of the desktop workstation's hardware and software configuration utilized during training is supplied in [Table 2](#).

Training method

Table 2. Workstation specifications.

Name	Parameter
CPU	Intel(R)Xeon(R) Gold 5218R
RAM	128 GB

ROM	4 T
GPU	NVIDIA RTX 3090
System	WIN 10

To meet the speed requirements for real-time cabbage recognition, the architecture introduced here was trained and benchmarked head-to-head against YOLOv5n and YOLOv8n, all trained on the mixed dataset. The spatial resolution of training samples was fixed at 480 × 384 pixels, and gradient updates were computed over mini-batches of 128 samples. Training spanned 500 epochs for every model variant. To drive faster convergence of the objective, the Stochastic Gradient Descent (SGD) update rule was adopted. The optimizer’s knob settings were: a base learning rate of 0.025, a momentum coefficient of 0.6, and a weight decay penalty of 0.001.

Model performance verification methods

Versions of the filter-parameter adaptive learning module, built with differing convolutional depths, were each inserted into the cabbage recognition pipeline and trained; their respective detection accuracies on the light occlusion test set were then compared. This comparison pinpointed the convolutional depth that yields the best parameter-generation performance. The proposed models, Yolov5n and Yolov8n, were then evaluated across three distinct test beds—the mixed dataset, the normal light dataset, and the light occlusion dataset—so that performance metrics could be contrasted both across architectures and across illumination regimes. To probe the internal mechanisms underlying performance differences, GRAD-CAM-derived visual explanations were generated, and changes in image entropy from raw to enhanced frames were quantified. For on-device benchmarking, the model was ported to an Nvidia Jetson Orin NX edge computing unit, and the frame processing throughput of each architecture was measured on this constrained hardware. By feeding the model sets of cabbage imagery spanning a gradient of illumination evenness, the model’s recognition fidelity under increasing light occlusion severity was systematically verified. The model’s capacity to generalize to other vegetable morphologies was tested against broccoli and Chinese cabbage collections that the research team had accumulated in earlier efforts. Finally, to gauge real-world field readiness, video streams of cabbage captured during field operations were fed into the recognition model, and a manual census of missed plants and spurious detections was conducted under conditions mimicking live deployment.

Model evaluation metrics

Metric selection is pivotal for both guiding hyperparameter choices during training and delivering a conclusive judgment on the cabbage recognition model’s quality. This study relies on mean Average Precision (mAP50) and per-frame processing latency to form a two-pronged evaluation. The computation underlying mAP50 is formalized in [Eq. 4](#). Conceptually, its magnitude equals the area under the precision-recall curve, bounded by [0, 1]. Values approaching unity denote stronger model capability. The intermediate quantities of precision (P) and recall (R) are defined via [Eqs. 5 and 6](#).

$$mAP_{50} = \int_0^1 P(R) dR \tag{4}$$

$$P = \frac{T_P}{(T_P + F_P)} \tag{5}$$

$$R = \frac{T_P}{(T_P + F_N)} \tag{6}$$

In the above notation, mAP50—mean Average Precision; P—Precision; R—Recall; TP—True Positives (cabbage plants correctly localized); FP—False Positives (background or weed patches wrongly labeled as cabbage); FN—False Negatives (cabbage plants that the model failed to flag).

To obtain the processing speed metric, frames drawn from the validation set are streamed through the cabbage recognition model for inference. The wall-clock time consumed by the full set is divided by the total number of frames, yielding a per-image latency. Taking the reciprocal of this latency yields the throughput figure, quoted in frames per second (fps).

Results and Discussion

Recognition accuracy of the adaptive filter parameter learning module with different numbers of convolutional layers

The detection performance of the adaptive filter parameter learning module, when configured with varying convolutional depths under light occlusion, is catalogued in [Table 3](#). Among the tested configurations, the 6-layer stack stands out, achieving a recognition accuracy of 93.1 %—a mark that eclipses that of every other variant examined. There is a direct relationship between convolutional depth and model footprint: adding layers increases both the parameter count and the size of the stored weight file.

Table 3. Recognition accuracy of the adaptive filter parameter learning module with different numbers of convolutional layers.

Convolutional layers	<i>mAP</i> ₅₀ /%	Params/M	Size/MB
5	81.3	3.22	6.38
6	93.1	3.29	6.51
7	89.8	3.43	6.77

Experimental results of different filter combination orders

The influence of filter chain ordering on outcomes was tested, and the results are presented in [Table 4](#). The data leaves little ambiguity: arranging the sequence as Liner-Gamma-Sharp delivers optimal performance on the light occlusion collection. The logic behind this ordering unfolds as follows. Leading with the linear filter normalizes the pixel intensity distribution, ironing out gross illumination disparities. Global contrast is elevated, which pulls subtle image details into sharper relief. A richer set of these fine details persists through the pipeline because the initial contrast boost prevents the downstream gamma and sharpening stages from discarding information that would otherwise be lost to insufficient tonal separation [24]. Operating on an already contrast-enhanced frame, the gamma correction step can then modulate luminance with greater precision—brightening dim zones without veering into overexposure of already-bright zones. The sharpening filter, being a localized operator, zeroes in on zones of steep intensity gradients, namely edge regions, rendering the structural skeleton of the image more distinct. Reversing the order creates problems: running sharpening ahead of adequate brightness and contrast conditioning amplifies native noise signatures and irregular artifacts, pushing spurious content into prominence. When the sequence deviates from the optimal ordering, the network, during its optimization phase, learns to suppress the filters' corrective influence, and recognition scores on the light-occluded dataset suffer as a direct consequence.

Table 4. Experimental results of different filter combination orders.

Order of filter combinations			<i>mAP₅₀</i> I%		
			Light occlusion dataset	Mixed dataset	Normal lighting dataset
Liner	Gamma	Sharp	93.1	95.0	97.5
Liner	Sharp	Gamma	84.0	90.6	98.1
Gamma	Liner	Sharp	82.2	90.5	97.8
Gamma	Sharp	Liner	89.6	93.4	97.9
Sharp	Liner	Gamma	85.6	91.3	98.0
Sharp	Gamma	Liner	82.5	89.8	97.9

Analysis of the ablation experiment

Table 5 documents the ablation findings, with the symbol √ indicating that a given module was kept operational and × denoting its removal. That performance consistently erodes upon the excision of any single filter speaks to the non-redundant contribution each makes to improving aggregate accuracy. Quantifying these contributions: pulling the linear filter out of the pipeline shaved 4.6% off accuracy, while dropping the sharpening filter cost 4.0%. The heaviest penalty was incurred by removing the Gamma filter—accuracy on the light occlusion set plunged to 80.9%, a precipitous 12.2% drop, by far the largest swing registered in the ablation series. Running the Gamma filter in isolation reduced light occlusion accuracy by just 1.7%, a finding that cements the Gamma filter's status as the single most consequential element for feature rehabilitation. What accounts for its outsized impact is its mode of operation: a nonlinear, power-law remapping of pixel values that jointly manipulates brightness and contrast, thereby breathing new life into features entombed within the deeply shadowed pockets of cabbage imagery caused by occluded lighting. Turning to the normally illuminated dataset, the story changes: activating or deactivating any solitary filter, or indeed any combination thereof, perturbed recognition accuracy by no more than 0.3%. This near-negligible swing reveals that the filter suite offers little additional benefit when the raw imagery already enjoys uniform lighting.

Table 5. Ablation experiment results.

Yolov8n	Liner	Gamma	Sharp	<i>mAP₅₀</i> I%		
				Light occlusion dataset	Mixed dataset	Normal lighting dataset
√	×	×	×	83.3	87.5	97.5
√	√	×	×	84.7	90.8	97.6
√	×	√	×	91.4	94.5	97.5
√	×	×	√	84.1	90.4	97.5
√	√	√	×	89.1	93.5	97.5
√	√	×	√	80.9	88.7	97.6
√	×	√	√	88.5	92.8	97.3
√	√	√	√	93.1	95.0	97.5

Recognition accuracy of different models on different test datasets

Figure 5 offers a side-by-side visual comparison of how the different architectures behave on sample cabbage imagery. Under benign, uniformly lit conditions, cabbage textural motifs and boundary contours stand out crisply, and the proposed

architecture returns strong detection results alongside Yolov5n and YOLOv8n. The picture changes dramatically once light occlusion enters the scene. Frames then become patchworks of specular hot spots and deep shadow pools. The YOLOv5 model frequently hallucinated detections on cabbage specimens sitting in shadowed pockets, while simultaneously overlooking small cabbage targets in brightly lit zones. YOLOv8n likewise registered a large number of false-positive detections confined to the shadowed terrain. In stark contrast, the architecture described in this paper maintained its detection fidelity regardless of whether the cabbage was in shadow or glare.

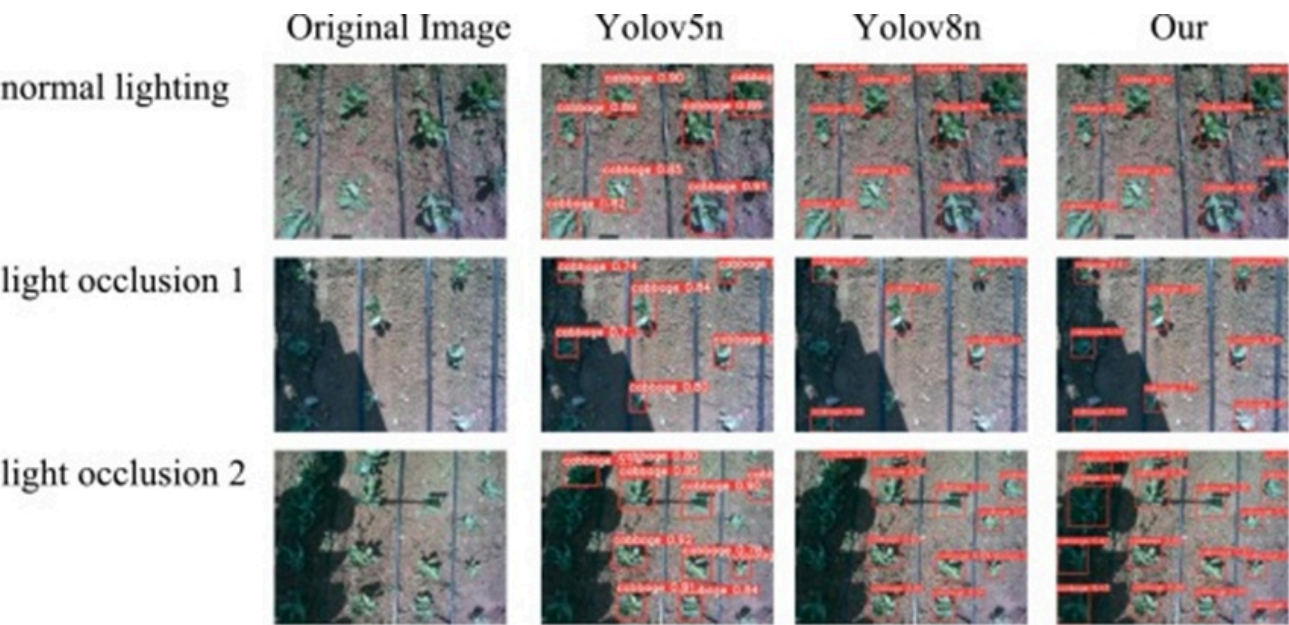


Figure 5. Recognition effect of different models.

Plotted in **Figure 6** are the accuracy metrics recorded by each model variant across the three evaluation suites: the mixed collection, the light-occluded collection, and the normally lit collection. On the mixed dataset, the model introduced in this work achieved 95.0% accuracy, translating into advantages of 7.5% over YOLOv8n and 8.2% over YOLOv5n. On the light occlusion dataset specifically, the proposed model achieved a peak accuracy of 93.1%, giving it a lead of 9.8% over YOLOv8n and 13.6% over YOLOv5n. Even on the normal lighting dataset—where all models performed respectably—the proposed model’s 97.5% edged out YOLOv5n’s 95.2%. The pattern is consistent: across testing regimes, the model proposed in this investigation demonstrated a clear accuracy advantage over both YOLOv8n and YOLOv5n.

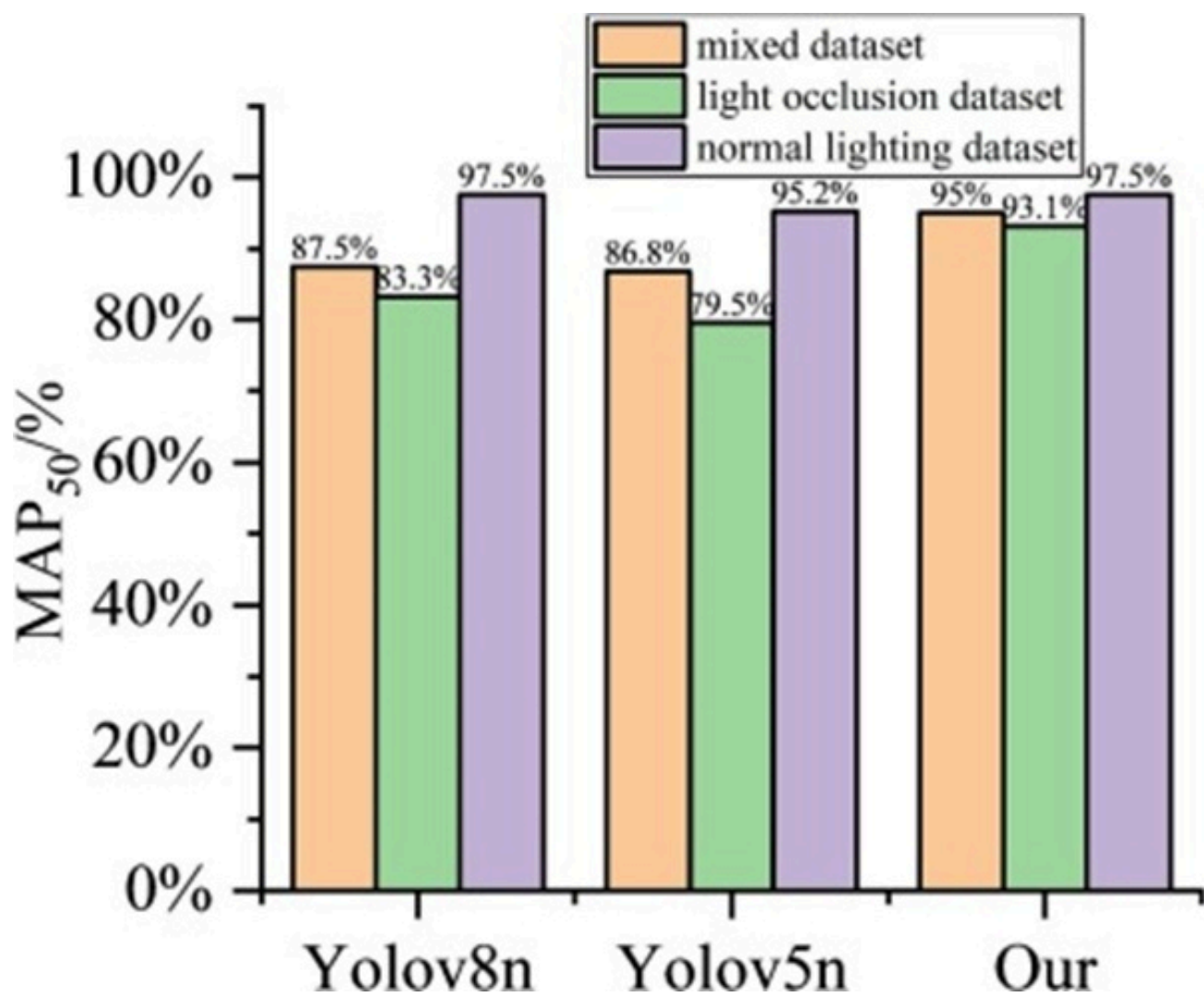


Figure 6. Recognition accuracy of different models on various datasets.

To shed light on the mechanisms driving these improvements, Grad-CAM attribution maps were generated and displayed as heatmap overlays in [Figure 7](#). Regions glowing with warmer hues denote areas exerting stronger influence on the recognition decision. What emerges from these visualizations is that YOLOv5n and YOLOv8n produce substantially weaker activations than the proposed model in shadowed zones cast by light occlusion on cabbage. This observation corroborates the notion that the anti-occlusion methodology engineered in this study meaningfully amplifies the content of cabbage-defining features in the imagery, thereby powering the observed accuracy gains. A complementary line of evidence comes from information-theoretic measurement: raw, unprocessed frames carried an average entropy of 13.5. After traversing the filter cascade, the average climbed to 18.3, up 4.8. An upward shift in entropy signals an expansion of informational richness within the frames, and this enrichment underpins the gains in recognition fidelity.

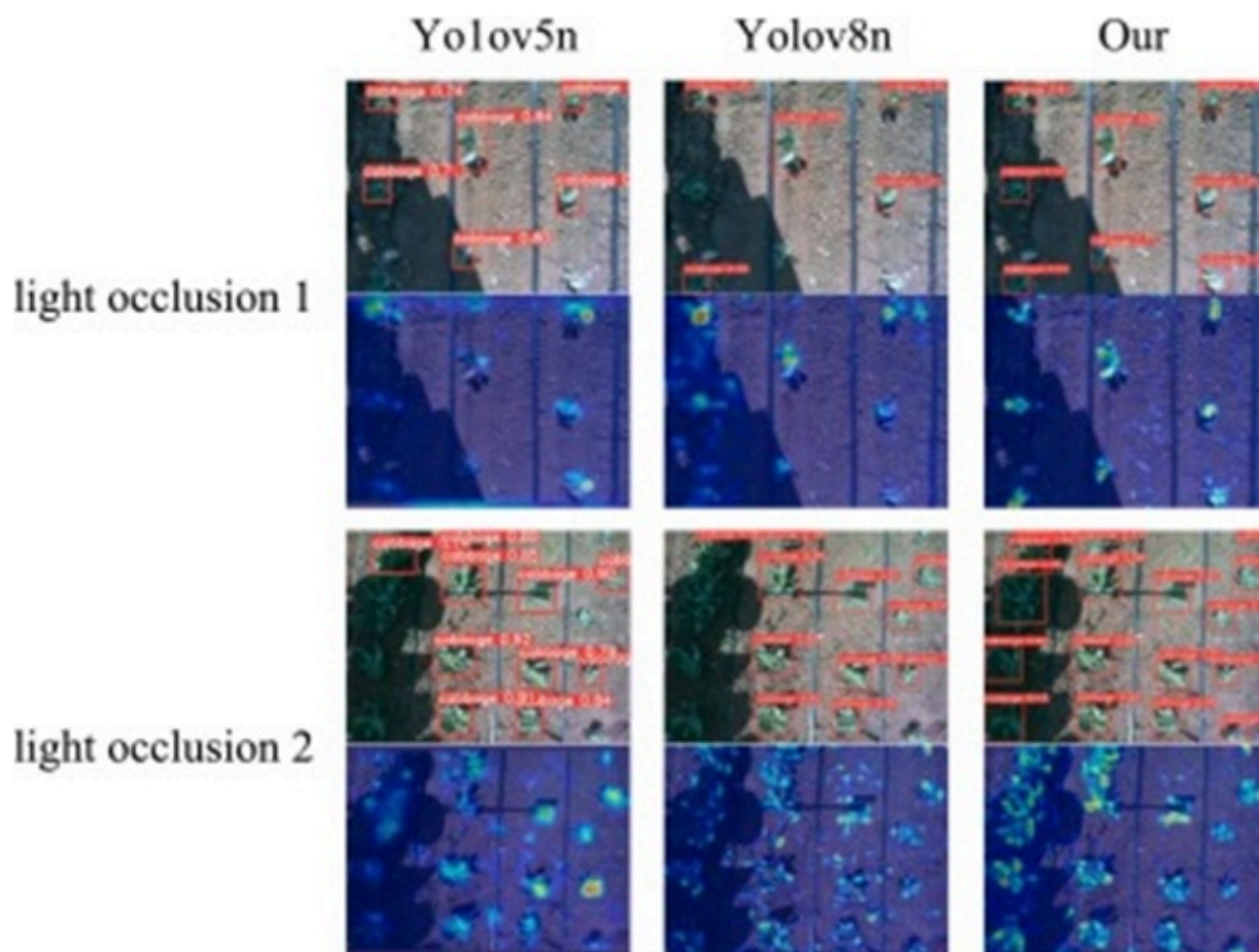


Figure 7. Model activation heatmap.

Recognition results at different levels of light occlusion

Figure 8 maps the detection behavior as the severity of light occlusion is progressively increased. A hard failure boundary emerges when the luminance ratio—comparing shadowed patches spawned by occlusion with their normally lit counterparts—falls below 0.2. In this extreme regime, cabbage within occluded zones is entirely undetected, regardless of whether the proposed models, YOLOv5n or YOLOv8n, are used. Within the intermediate band, where the luminance ratio lies between 0.2 and 0.4, all three architectures recover a partial picture, detecting some fraction of the occluded cabbage. The decisive separation occurs once the luminance ratio exceeds 0.4: here, the model proposed in this study outperforms the two alternatives in recognition accuracy.

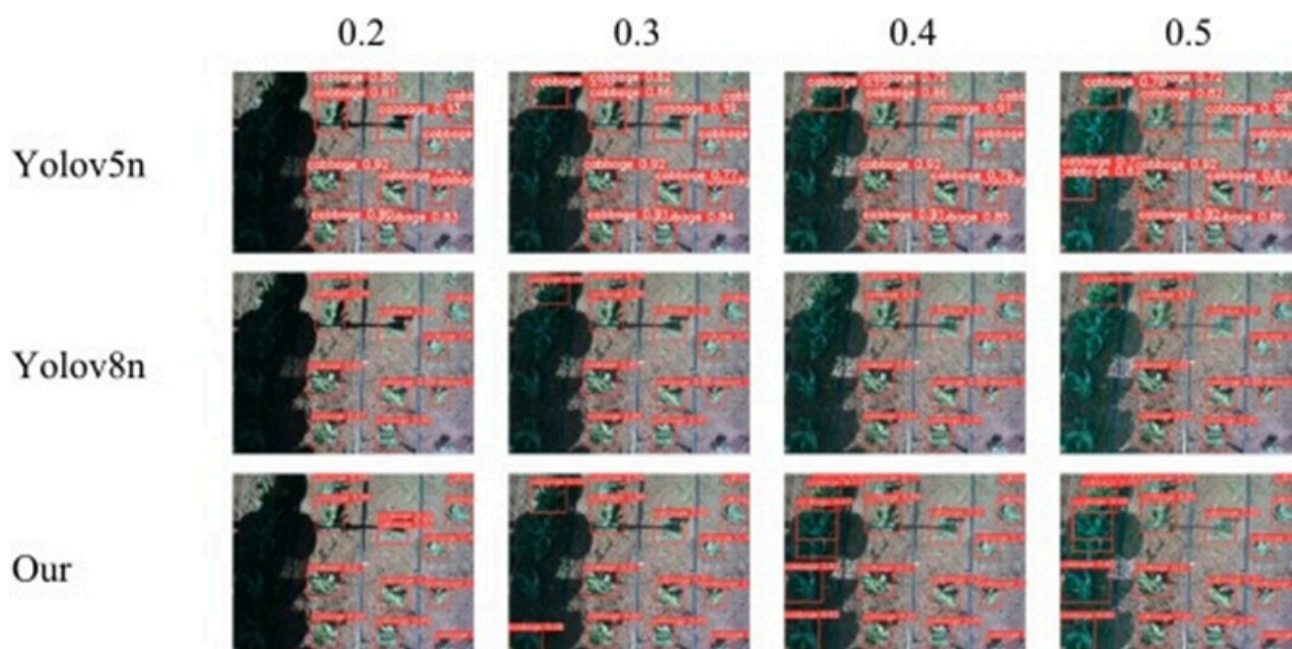


Figure 8. Recognition results at different levels of light occlusion.

Figure 9 traces the accuracy trajectories of the three models as the degree of light occlusion relaxes. A monotonic increase in accuracy across all architectures accompanies the reduction in occlusion severity. At the tightest occlusion level evaluated—0.2—the proposed model fell short of Yolov8n by 0.8% but outperformed Yolov5n by 8.2%. Beyond the 0.4 threshold, the proposed cabbage recognition model stabilized at 93.1% accuracy, opening a 9.9% margin over YOLOv8n and a 13.7% margin over YOLOv5n. This leveling-off in performance indicates that the proposed model consistently and substantially outperforms both YOLOv8n and YOLOv5n when the degree of light occlusion exceeds 0.4.

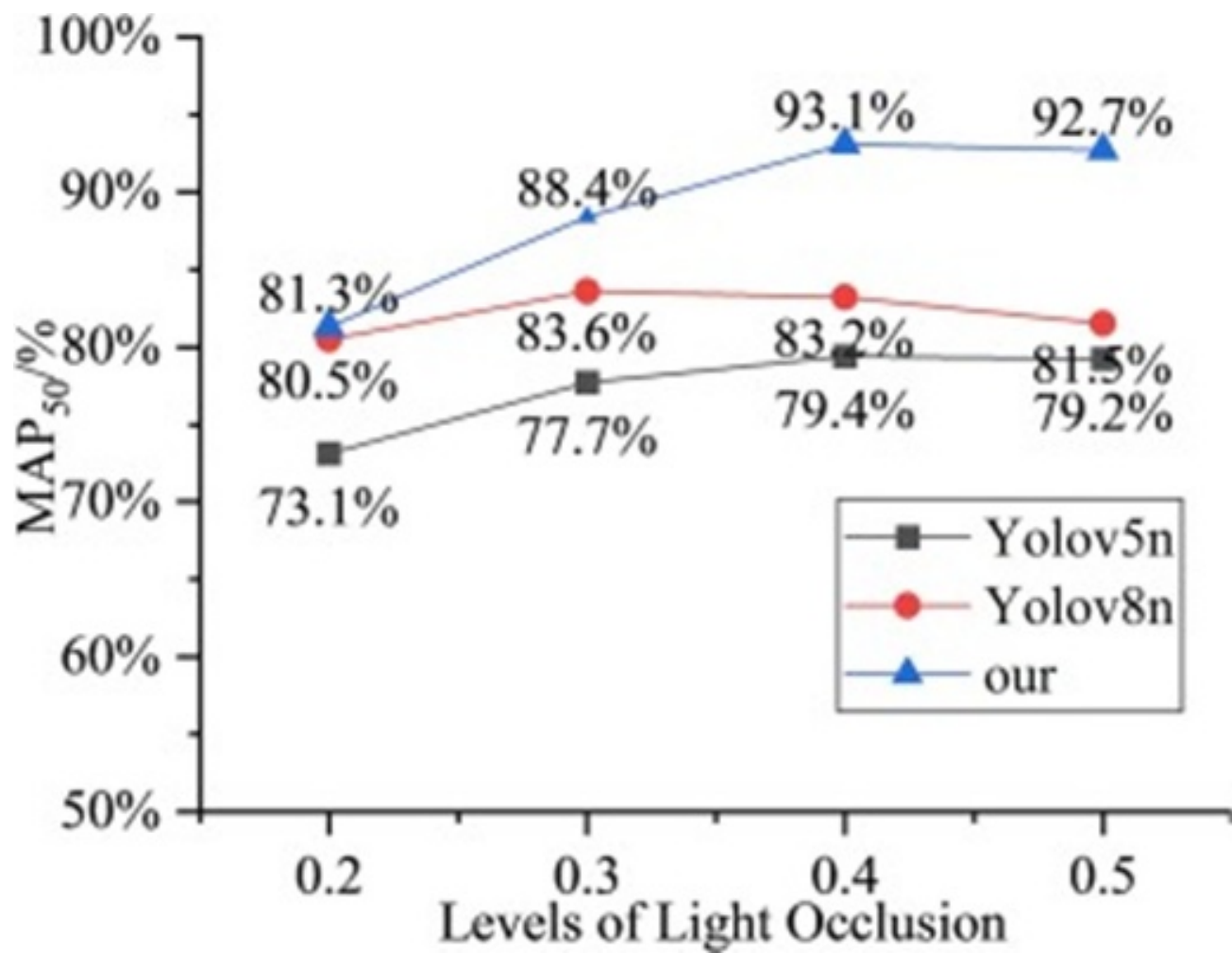


Figure 9. Recognition accuracy at different levels of light occlusion.

Performance comparison of the model deployed on edge computing devices

The model was ported to an Nvidia Jetson Orin NX edge computing unit operating at 10 W, and the resulting performance characteristics are presented in Table 6. As the table documents, incorporating the adaptive learning filter module increased parameter volume and the size of the weight files. Relative to YOLOv5n and YOLOv8n, the model's parameter count grew by 0.28 M and 2.05 M, while its weight file expanded by 0.57 MB and 2.94 MB, respectively. The throughput achieved by the proposed model reached 26.32 f/s, falling short of the comparison models by 8.04 and 7.35 fps. Timing measurements captured via the time module revealed that the image filter and the filter parameter adaptive module together consumed 11.47 ms of runtime, accounting for why the proposed model's frame processing rate trails that of the unmodified YOLOv8n.

Table 6. Performance comparison of models deployed on edge computing devices.

Model	Params/M	FPS/fps	Size/MB
Yolov5n	1.77	34.36	3.57
Yolov8n	3.01	33.67	5.94
Our	3.29	26.32	6.51

Recognition accuracy of cabbage at different growth cycles

Table 7 compiles the recognition outcomes for cabbage sampled across multiple growth intervals. A clear pattern emerges: detection accuracy tends to climb as the growing period lengthens, though environmental context exerts a simultaneous influence. For cabbage cultivated over a mere 12 days, accuracy settled at 88.5%, a consequence of their diminutive stature and the complicating presence of shadows. Because the collected cabbage specimens originated from separate planting cohorts, field management practices varied. The batch grown for 33 days suffered from poor maintenance and heavy weed infestation, as depicted in **Figure 10**, which dragged down recognition accuracy. In contrast, cabbage grown for 30 days under standard agronomic care and in weed-free plots was recognized with 99.4 accuracy.

Table 7. Results of recognition at different Growth periods.

Growth period/day	5	12	17	19	22	25	27	29	30	33	42
<i>mAP50</i> /%	95.3	88.5	95.3	98.2	97.9	95.5	99.4	94.7	99.1	93.7	91.8



(a)5

(b)12

(c) 30

(d) 33

Figure 10. Recognition of different growth periods.

Results of validation experiments on broccoli and Chinese cabbage datasets

The vegetable recognition model proposed in this work was trained from scratch on collections of broccoli and Chinese cabbage imagery. The detection outputs for these two vegetables are shown in **Figures 6 and 11a**. Inspection confirms that the model reliably identifies Chinese cabbage and broccoli targets, whether they are within light-occluded zones or in normally illuminated regions. On their respective test sets, the model achieved recognition accuracies of 99.3% for Chinese cabbage and 98.3% for broccoli. These figures attest that the cabbage recognition architecture presented here generalizes to other vegetable types with strong performance.

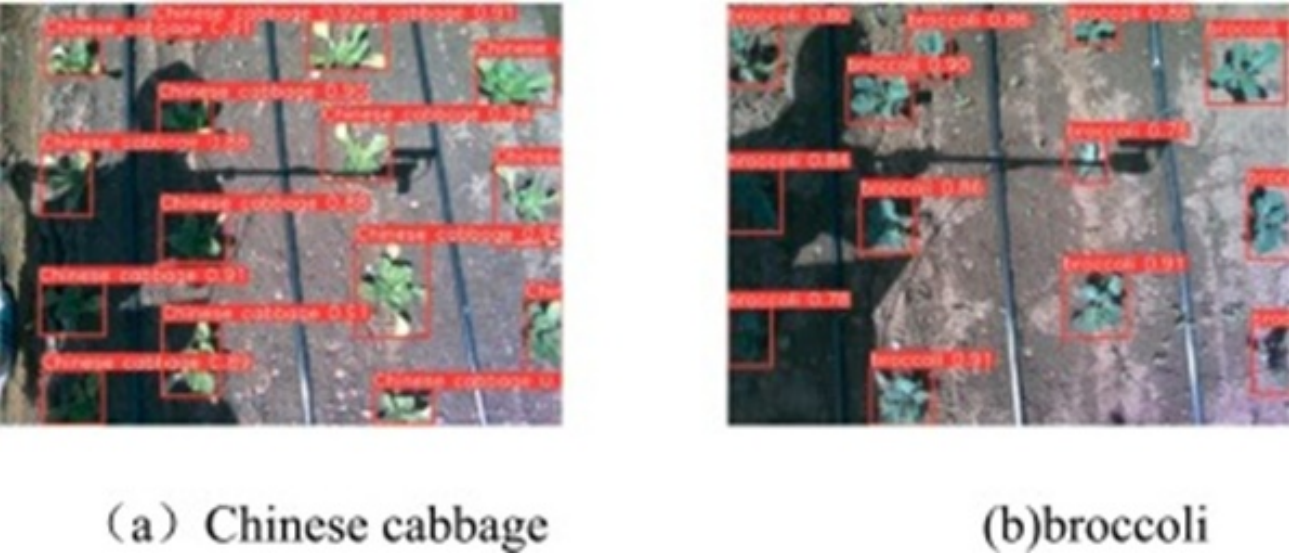


Figure 11. Recognition results of Chinese cabbage and broccoli.

Field experiment results

Video footage collected under both normal and light-occluded illumination was fed into the proposed cabbage recognition pipeline, and the resulting tallies are shown in [Table 8](#). Across the two conditions, the total number of cabbage plants was 846 and 817, respectively. Operating on normally lit footage, the model correctly identified 812 plants, yielding an accuracy of 96.0%; under light occlusion, the accuracy reached 91.2%. These field results substantiate that the proposed methodology performs well across diverse illumination conditions, making it suitable for online cabbage detection in open-field production settings.

Table 8. Field experiment results.

Empty Cell	Actual number/plant	TP/plant	FN/plant	FP/plant	P/%
normal lighting	846	812	34	47	96.0 %
light occlusion	817	745	72	26	91.2 %

Centered on cabbage grown in open fields, this work achieved real-time detection under partial light-blocking conditions and conducted a systematic examination of recognition accuracy across multiple datasets. The experimental evidence demonstrates that, when presented with light occlusion, the constructed cabbage recognition model outperformed YOLOv5n by 13.6% and YOLOv8n by 9.8%. This outcome underscores the potency of the parameter-adaptive filter as an image processing tool. Entropy values were computed for both raw and enhanced image data. Entropy, understood as the mean information content of an image, quantifies its informational richness in information theory; higher entropy values reflect greater information density. The analysis revealed that the average entropy of frames increased by 4.8 after filtering, indicating that the technique effectively enhances the amount of usable information in imagery degraded by light occlusion, and this enrichment directly fuels the observed gains in cabbage recognition accuracy.

In this investigation, an adaptive filter parameter module operating in tandem with a filter combination was employed to process cabbage imagery. The introduction of these components incurred processing speed penalties of 8.04 and 7.35 frames per second relative to YOLOv8n and YOLOv5n, respectively. Instrumentation via a time module showed that the combined execution time of the image filter and the filter-parameter-adaptive module was 11.47 ms, which explains the throughput deficit of the proposed cabbage recognition model relative to YOLOv8n and YOLOv5n.

The cabbage recognition model equipped with the image filter combination achieved 93.1% accuracy on frames where the light uniformity metric exceeded 0.4, outpacing YOLOv8n by 9.9% and YOLOv5n by 13.7%. Beyond this 0.4 threshold, the filter combination's responsiveness to further changes in light occlusion diminishes, causing the rate of performance gain to decelerate. Within the 0.2 to 0.4 occlusion band, both the absolute accuracy and the upward trajectory of the proposed model surpass those of YOLOv8n and YOLOv5n, indicating that the parameter-adaptive enhancement module and filter combination are particularly attuned to cabbage image features within this occlusion range, where their feature-boosting effect is most pronounced. At occlusion levels below 0.2, the proposed method achieves accuracy on par with YOLOv8n. Synthesizing these observations, the filter combination delivers meaningful feature amplification for cabbage imagery exhibiting light occlusion degrees above 0.2. For images where occlusion drops below 0.2, a stark luminance divide separates occluded and lit zones. The mean grayscale value within shadowed regions measures 27.4, while illuminated regions register 137; normalizing these yields values of 0.107 and 0.537, respectively. When subjected to the Gamma filter, pixel intensities undergo the transformations mapped in [Figure 12](#). Shadowed-area grayscale values experience pronounced stretching. However, raising a shadowed region to a normalized brightness of 0.5 simultaneously pushes the originally normal brightness region to 0.83, corresponding to a pixel value of 212, thereby inducing overexposure in the properly lit zone and undermining overall recognition accuracy. Throughout the adaptive parameter generation process, the model tunes filter coefficients to maximize global recognition accuracy. For imagery exhibiting mild shadowing, the enhancement of useful features is held in check by collateral damage to overall accuracy, making fully correct recognition difficult to attain. Continued inquiry is warranted to devise feature enhancement strategies tailored to single-image captures exhibiting extreme brightness disparities, with the goal of further fortifying model adaptability.

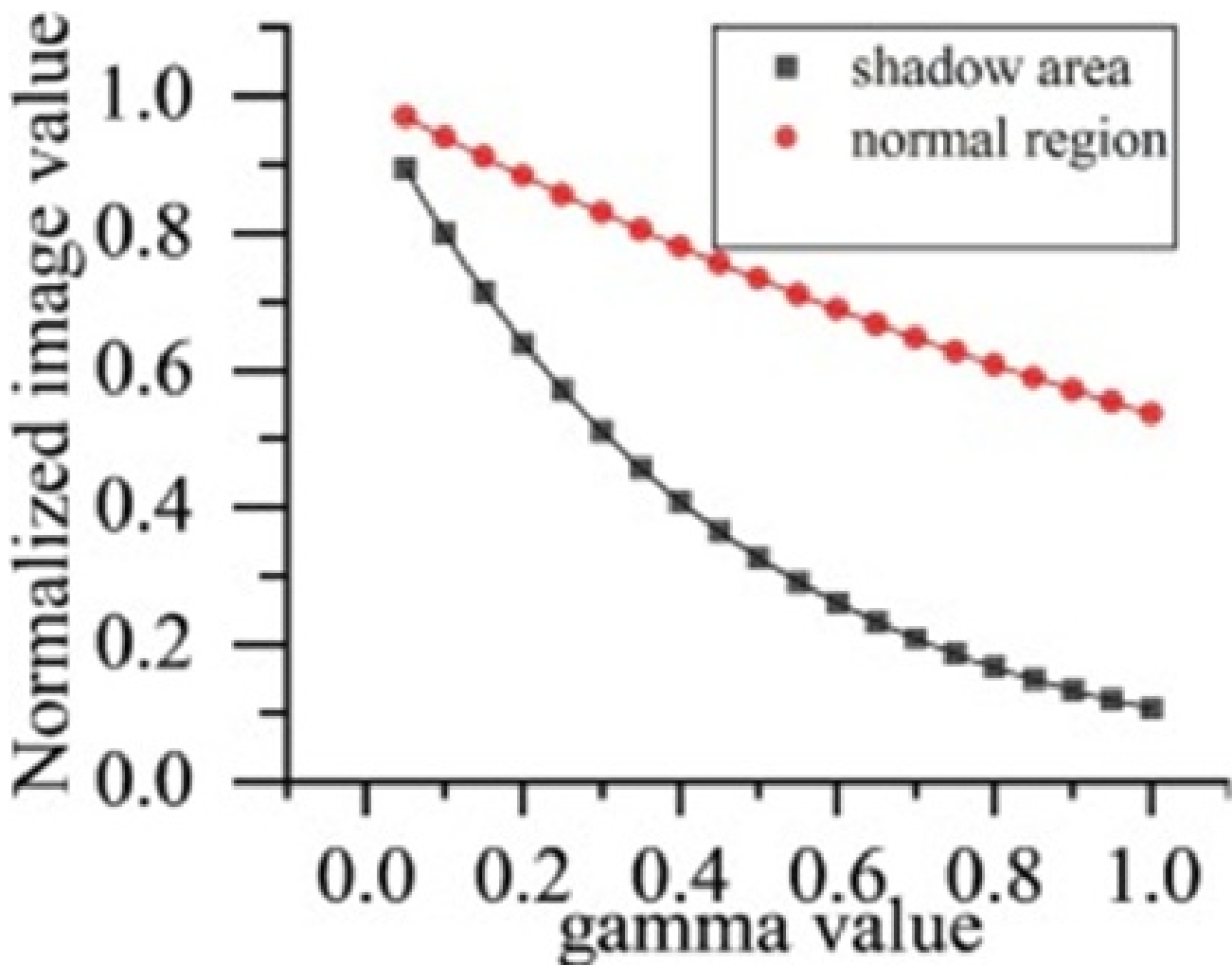


Figure 12. Gamma filter transformed image.

Turning to the field evaluation, recognition accuracy under both normal and light-occluded conditions fell below the figures recorded on the curated test datasets, with a higher incidence of missed targets. As shown in **Figure 13a**, a portion of the cabbage plants experienced disease stress, resulting in chlorotic leaves and reduced plant stature. These conditions strip away key features, leading to missed detections. To remedy this, future investigations can delve into precision recognition methodologies for cabbage subjected to the stresses of pathogens, pests, and inclement weather, aiming to broaden the model's operational envelope. In **Figure 13b**, certain individual cabbages exhibit both small size and a clustered leaf architecture, which also led to detection failures. The usable information captured by the camera shrinks, with the cabbage bounding box occupying only 28×25 pixels—0.0022 of the total frame area at a resolution of 640×480 pixels. This fraction falls below the 0.03 benchmark, categorizing the problem as one of small-object detection. The backbone feature-extraction network in Yolo v8n downsamples the cabbage frames by a factor of 32. For targets whose dimensions are smaller than 32×32 pixels, feature representations are completely obliterated, leading to non-detection. To improve recognition performance for extremely diminutive individual cabbages, forthcoming research must engineer a backbone feature-extraction network that employs a less aggressive downsampling factor while preserving a rich feature-extraction capacity.



Figure 13. Unrecognized cabbage.

Conclusion

This investigation proposed a lightweight, online method for recognizing cabbage under challenging lighting conditions with occlusions. The approach marries an adaptive convolutional neural network-based module for image filter parameters with a suite of complementary image filters. Filter coefficients are inferred from the feature content of cabbage imagery, and the filter suite is subsequently applied to amplify image features, thereby elevating cabbage recognition accuracy in the face of light occlusion. Both the filter-parameter adaptive learning module and the image filter combination are directly woven into the object detection network. An end-to-end training paradigm is adopted, allowing the adaptive learning module and the detection network to be jointly optimized. Experimental outcomes place recognition accuracy at 93.1% on the light occlusion dataset, 97.5% on the normal lighting dataset, and 95.0% on the mixed dataset, figures that signal markedly superior detection fidelity. When benchmarked against other cabbage recognition models described in the literature, the architecture proposed in this study holds advantages in both accuracy and throughput. Deployed on the Nvidia Jetson Orin NX edge computing device, it processes frames at 26.32 fps, satisfying the requirements for real-time recognition. This research provides a reference point for detecting field-grown cabbage and other crops under lighting interference, provides technical support for targeted spraying, and advances the intelligent evolution of plant protection operations. Looking forward, deeper exploration can focus on accurate recognition of cabbage under stress from disease, pests, and severe weather, as well as on detection techniques for very small individual plants, thereby expanding the model's scope and applicability.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Youth Science Fund Project) (32201647), the Special Fund for Innovation Capacity Building of Beijing Academy of Agriculture and Forestry Sciences (KJ CX20230409), and the Agricultural “Spark Technology” Cultivation—Targeted Spraying Robot for Open-field Cabbage (13230366).

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 14 Feb 2023

Revised: 29 Mar 2023

Accepted: 18 May 2023

Published online: 15 September 2023

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Özlüoymak ÖB. Development of a servo-controlled nozzle on the mobile robot for target-oriented micro-dose spraying in precision weed control. *Semin Cienc Agrar*. 2021;42(2):635-56.
<https://doi.org/10.5433/1679-0359.2021v42n2p635>.
- Lameski P, Zdravetski E, Kulakov A. Review of automated weed control approaches: an environmental impact perspective. In: Kalajdziski S, Ackovska N, editors. *ICT Innovations 2018: Engineering and Life Sciences*. Cham: Springer International Publishing; 2018. p. 132-47. (Communications in Computer and Information Science).
https://doi.org/10.1007/978-3-030-00825-3_12.
- Andujar D, Weis M, Gerhards R. An ultrasonic system for weed detection in cereal crops. *Sensors (Basel)*. 2012;12(12):17343-57.
<https://doi.org/10.3390/s121217343>.
- Zou K, Chen X, Wang Y, Zhang C, Zhang F. A modified U-Net with a specific data argumentation method for semantic segmentation of weed images in the field. *Comput Electron Agric*. 2021;187:106242.
<https://doi.org/10.1016/j.compag.2021.106242>.
- Cruz Ulloa C, Krus A, Barrientos A, del Cerro J, Valero C. Robotic fertilization in strip cropping using a CNN vegetables detection-characterization method. *Comput Electron Agric*. 2022;193:106684.

<https://doi.org/10.1016/j.compag.2021.106684>.

Xu Y, Shu L, Xie Q, Song M, Zhu Y, Cao W, et al. Precision weed detection in wheat fields for agriculture 4.0: a survey of enabling technologies, methods, and research challenges. *Comput Electron Agric.* 2023;212:108106.

<https://doi.org/10.1016/j.compag.2023.108106>.

Zhang L, Zhang Z, Wu C, Sun L. Segmentation algorithm for overlap recognition of seedling lettuce and weeds based on SVM and image blocking. *Comput Electron Agric.* 2022;201:107284.

<https://doi.org/10.1016/j.compag.2022.107284>.

Xu K, Li H, Cao W, Zhu Y, Chen R, Ni J. Recognition of weeds in wheat fields based on the fusion of RGB images and depth images. *IEEE Access.* 2020;8:110362-70.

<https://doi.org/10.1109/ACCESS.2020.3001999>.

Ariza-Sentís M, Vélez S, Martínez-Peña R, Baja H, Valente J. Object detection and tracking in precision farming: a systematic review. *Comput Electron Agric.* 2024;219:108757.

<https://doi.org/10.1016/j.compag.2024.108757>.

Zhai C, Fu H, Zheng K, Zheng S, Wu H, Zhao X. Establishment and experiment of an online field cabbage recognition model based on deep learning. *Trans Chin Soc Agric Mach.* 2022;53:293-303.

Li Z, Li Y, Yang Y, Guo R, Yang J, Yue J, et al. A high-precision detection method of hydroponic lettuce seedlings status based on improved faster RCNN. *Comput Electron Agric.* 2021;182:106054.

<https://doi.org/10.1016/j.compag.2021.106054>.

Wang Q, Cheng M, Huang S, Cai Z, Zhang J, Yuan H. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* Dunal seedlings. *Comput Electron Agric.* 2022;199:107194.

<https://doi.org/10.1016/j.compag.2022.107194>.

FarmWise [Internet]. 2020 [cited 2026 Aug 2]. Available from: <https://farmwise.io/>

Carbon Robotics [Internet]. 2021 [cited 2026 Aug 2]. Available from: <https://carbonrobotics.com/>

EcoRobotix [Internet]. 2020 [cited 2026 Aug 2]. Available from: <https://www.ecorobotix.com/fr/avo-robot-desherbant-autonome/>

Wang A, Xu Y, Wei X, Cui B. Semantic segmentation of crop and weed using an encoder-decoder network and image enhancement method under uncontrolled outdoor illumination. *IEEE Access.* 2020;8:81724-34.

<https://doi.org/10.1109/ACCESS.2020.2991354>.

Choi D, Lee WS, Ehsani R, Schueller J, Roka FM. Detection of dropped citrus fruit on the ground and evaluation of decay stages in varying illumination conditions. *Comput Electron Agric.* 2016;127:109-19.

<https://doi.org/10.1016/j.compag.2016.05.020>.

Bao W, Cheng T, Zhou XG, Guo W, Wang Y, Zhang X, et al. An improved DenseNet model to classify the damage caused by cotton aphid. *Comput Electron Agric.* 2022;203:107485.

<https://doi.org/10.1016/j.compag.2022.107485>.

Ni W, Wei D, Peng Z, Ma Z, Zhu S, Tang R, et al. An appetite assessment method for fish in outdoor ponds with anti-shadow disturbance. *Comput Electron Agric.* 2024;221:108940.

<https://doi.org/10.1016/j.coempag.2024.108940>.

Bao W, Yang X, Liang D, Hu G, Yang X. Lightweight convolutional neural network model for field wheat ear disease recognition. *Comput Electron Agric.* 2021;189:106367.

<https://doi.org/10.1016/j.compag.2021.106367>.

Bao W, Zhu Z, Hu G, Zhou X, Zhang D, Yang X. UAV remote sensing detection of tea leaf blight based on DDMA-YOLO. *Comput Electron Agric.* 2023;205:107637.

<https://doi.org/10.1016/j.compag.2023.107637>.

Kalwar S, Patel D, Aanegola A, Konda KR, Garg S, Krishna KM. GDIP: gated differentiable image processing for object detection in adverse conditions. In: 2023 IEEE International Conference on Robotics and Automation (ICRA); 2023 May 29-Jun 2; London, UK. IEEE; 2023. p. 7083-9. (Crossref added)

<https://doi.org/10.1109/ICRA48891.2023.10160787>.

Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer Vision – ECCV 2014*. Cham: Springer; 2014. p. 818-33. (Lecture Notes in Computer Science; vol. 8689).

https://doi.org/10.1007/978-3-319-10590-1_53.

Yin T, Subaramaniam K, Shibghatullah A, Mansor N. Enhancement of low-light image using homomorphic filtering, unsharp masking, and gamma correction. *Int J Adv Comput Sci Appl.* 2022;13(8).

<https://doi.org/10.14569/IJACSA.2022.0130864>.