Vietnam Journal of Artificial Intelligence

Journal homepage: https://vjai.org.vn/en/vjai/

Published by East Asia University of Technology, Vietnam, ISSN: 3126-2872

RL-RiceBox: A Reinforcement Learning Framework for Bounding Box Refinement in Rice Leaf Disease Localization

Phi Dinh Huynh[1], Tran Quang Tiep[2], Kieu Ba Thinh[2], Truong Cong Doan[2], Nguyen Thi Quynh Anh[2], Nguyen Viet Hung[3,*]

Review

Open Access

Affiliation:

[1]Faculty of Information Technology, East Asia University of Technology, Phan Tay Nhac, Hanoi, Vietnam.

[2]International School, Vietnam National University, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam.

[3]International Training and Cooperation Institute, East Asia University of Technology, Phan Tay Nhac, Hanoi, Vietnam

Email:

*Corresponding author: hungnv@eaut.edu.vn;

Contributing authors: 20222072@eaut.edu.vn; 23070340@vnu.edu.vn; 23070247@vnu.edu.vn; tcdoan@vnu.edu; quynhanhnguyen@vnu.edu.vn; hungnv@eaut.edu.vn

Vietnam J. Art. Intell. **2026**, 1(1), 1-18
DOI: https://doi.org/10.65153/qf9jp371

Received: 23 March 2026
Accepted: 11 May 2026
Published: 24 July 2026

Abstract

Precise localization of diseased regions on rice leaves is critical for reliable diagnosis, yet bounding-box annotations and detector outputs are often coarse and inconsistent. We propose RL-RiceBox, a deep reinforcement learning framework that automatically refines an initial (inaccurate) bounding box through a short sequence of discrete geometric actions (translation, scaling, and termination). The refinement process is formulated as a Markov Decision Process, and a Deep Q-Network learns an action policy from a fixed cropped patch together with the current box coordinates, using an IoU-driven reward to encourage fast convergence to tighter lesion localization. To emulate realistic initialization noise, training and evaluation start from synthetically shifted boxes generated from the ground truth. Experiments on a rice leaf disease dataset show that RLRiceBox consistently improves localization quality, increasing mean IoU from 0.53 to 0.89 and achieving 90.4% and 88.6% success rates at IoU thresholds of0.50 and 0.75, respectively, with an average of 7.1 refinement steps. The proposed approach provides an effective post-processing module for improving box quality and strengthening agricultural vision pipelines.

Keywords: Computer Vision, Rice Disease Detection, Bounding box refinement, Deep Learning, Deep Q-Network (DQN).

1. Introduction

Rice is a staple food crop for a large share of the world's population and plays a critical role in global food security [1–3]. However, its productivity is highly vulnerable to foliar diseases such as bacterial blight, brown spot, and blast, which can cause substantial yield and quality losses if not detected and treated in time [4]. In many rice-producing regions, disease monitoring still relies heavily on manual scouting by human experts, which is labor-intensive, subjective, and difficult to scale to large fields or smallholder settings.

In the last decade, advances in deep learning for computer vision have enabled image-based plant disease diagnosis systems that can automatically recognize visual symptoms on leaf surfaces. Convolutional neural networks (CNNs) and their variants have been successfully applied to classify a wide range of crop diseases from RGB images, often achieving accuracy comparable to or exceeding human experts [5–8]. Comprehensive surveys further highlight that deep learning models have become the de facto standard for plant disease detection, covering classification, detection, and segmentation tasks across many crops and datasets [9]. These developments have motivated a new generation of decision-support tools that can assist farmers and agronomists in early disease detection. Rice leaf diseases have received particular attention in this context. A growing body of work has explored CNN-based approaches for rice leaf disease classification, leveraging increasingly larger and more diverse image collections captured in laboratory and field conditions [4]. At the same time, public datasets of rice leaf images with expert annotations have stimulated the development and benchmarking of new models. Among them, the Rice Disease dataset provides bounding-box annotations for three major diseases (bacterial blight, brown spot, and rice blast), making it well suited for object detection tasks that localize lesions on the leaf surface rather than only assign an image-level label [10].

While pure image-level classification can be sufficient for some agronomic applications, explicit localization of diseased regions offers several advantages. Bounding-box or region-level predictions can (i) support quantitative estimates of disease severity, (ii) provide more interpretable outputs for agronomists, and (iii) facilitate downstream tasks such as lesion segmentation or region-of-interest tracking over time. Consequently, several studies have adopted one-stage detectors such as YOLO variants for rice leaf disease detection, reporting promising performance in localizing multiple disease types in a single leaf image [11–13]. Nevertheless, these methods typically treat localization as a one-shot regression problem: the detector predicts bounding boxes in a single forward pass, and any inaccuracies in the box position or scale directly degrade the Intersection-over-Union (IoU) and the perceived quality of the annotations.

Producing high-quality bounding boxes is also expensive from an annotation perspective. Manually drawing precise boxes for thousands of lesions is time-consuming and prone to human error, and small inconsistencies in the labeling process can propagate to training and evaluation. To alleviate this, reinforcement learning (RL) has been investigated as a way to iteratively refine bounding boxes through a sequence of actions that translate, expand, or shrink the box [14]. In particular, BAR (Bounding box Automated Refinement) introduced an RL agent that learns to correct inaccurate bounding boxes generated by a detector or by imperfect human annotations, significantly improving IoU while reducing the amount of manual correction required [14]. Such approaches shift the focus from predicting boxes in a single step to learning a policy that gradually moves an initial box toward the ground truth.

Despite this progress, directly applying general-purpose bounding-box refinement methods, such as BAR [14], to agricultural imagery presents unique and non-trivial challenges. Unlike generic visual objects (e.g., vehicles or humans) which typically possess distinct structural boundaries, rice leaf lesions exhibit extreme morphological variations. They range from tiny, low-contrast anomalies (e.g., Brown Spot) to elongated, highly irregular streaks that blend seamlessly into the leaf edges (e.g., Leaf Blight). These domain-specific characteristics often cause standard one-stage object detectors [11–13] to produce bounding boxes that are highly noisy, skewed, or overextended. Consequently, a refinement policy must be specifically adapted to navigate these subtle visual boundaries and high intra-class variability. This gap motivates the development of a tailored RL-based refinement framework, shifting the paradigm from generic object refinement to specialized precision agriculture localization.

Recent work has explored hybrid deep learning and machine learning approaches for plant disease analysis. For example, Hung et al. [15] proposed a tomato leaf disease classification framework that integrates CNN-based feature extraction with traditional classifiers, demonstrating that hybrid pipelines can enhance robustness and interpretability in real agricultural settings. Motivated by this shift toward more structured and region-aware analysis, we extend the problem from image-level classification to region-level localization and refinement for rice leaf diseases. In addition, Reinforcement Learning (RL) has shown strong capabilities in spatial decision-making tasks; notably, Hung et al. [16] applied RL in the HEVERL model to optimize viewport prediction in 360-degree video streaming, highlighting RL's effectiveness in dynamic spatial adjustment. Although not directly related to plant pathology, this evidence suggests that RL is a promising direction for bounding box refinement, informing our design of an RL-based framework for more precise rice leaf disease localization.

In this paper, we propose a hybrid object-detection and deep reinforcement learning framework for automated localization of rice leaf diseases with refined bounding boxes. Starting from initial proposals generated by a modern one-stage detector trained on the Rice Disease dataset [10], we train an RL agent inspired by BAR [14] to iteratively adjust the box through a discrete set of actions that translate and rescale it. The agent is rewarded according to the improvement in IoU with respect to the ground-truth box, encouraging policies that quickly converge to tight and accurate localizations. We experimentally demonstrate that the proposed refinement agent improves the quality of lesion bounding boxes over the baseline detector, particularly for challenging cases with small or elongated lesions.

As illustrated in Figure 1, the conventional object detection pipeline treats the detector output as the final prediction: the model receives rice leaf images and directly returns bounding boxes around diseased regions. In practice, these one-shot predictions can still suffer from inaccurate localization, particularly in cluttered scenes or for small lesions. In our framework, the detector is followed by a reinforcement learning (RL) agent that operates on the initial bounding boxes.

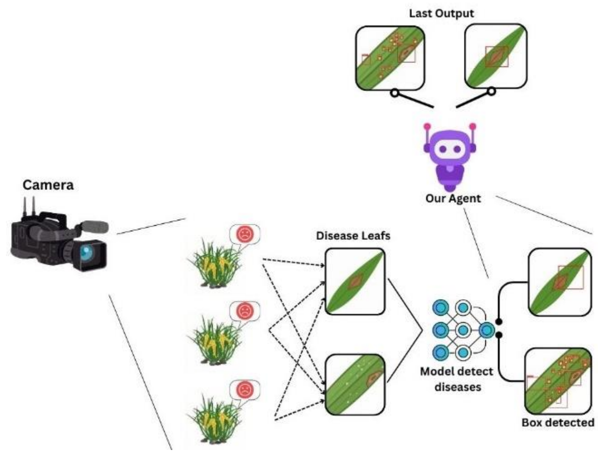

Fig. 1: Overall concept of the proposed framework.

The agent observes local image patches together with the current bounding box coordinates, which jointly constitute the state of the environment, and applies a sequence of learned actions to iteratively adjust the box. Unlike direct regression-based refinement, this sequential decision-making process allows the model to adaptively correct coarse detections. As a result, the initial predictions from the baseline detector are refined into tighter and more reliable lesion bounding boxes, leading to more accurate localization in our experiments.

The main contributions of this work are summarized as follows:

- We identify a critical gap in current disease detection pipelines where standard detectors struggle with extreme aspect ratios and visually subtle lesions. Rather than a simple plug-and-play application of existing RL methods, we formulate, to the best of our knowledge, the first systematic integration of deep reinforcement learning specifically engineered for bounding-box refinement in plant pathology. We rigorously evaluate how an RL agent navigates the complex morphological variations of rice leaf diseases through discrete geometric actions.

- We propose a specialized refinement agent tailored to the challenging characteristics of agricultural imagery. By structuring the state representation and IoU-driven reward mechanism to accommodate multi-class lesion patterns with ambiguous boundaries, our approach stabilizes Deep Q-Network training and effectively corrects the structured localization noise typically generated by agricultural object detectors.
- We conduct an extensive evaluation across IoU thresholds and controlled annotation-noise settings. The results demonstrate that our RL-based refinement consistently improves localization accuracy over a strong baseline detector while adding minimal computational overhead, highligh-ting its practicality for real-world agricultural inspection systems.

The structure of this manuscript is organized to guide the reader from foundational context to methodological and empirical contributions. Section 2 synthesizes prior studies on plant disease recognition, publicly available rice leaf disease corpora, and reinforcement learning strategies for bounding-box refinement. Section 4 then delineates the proposed hybrid framework together with its RL formulation. Subsequently, Section 5 details the experimental protocol and reports the resulting performance analyses. Finally, Section 6 summarizes the key findings and discusses prospective avenues for advancing this line of research.

2. Related Work

Hung et al. [17] demonstrated the strong capability of deep learning models in image based recognition tasks. Although their work centered on human motion analysis, the underlying preprocessing strategies and feature learning techniques are broadly transferable and can be adapted to agricultural applications such as rice plant and leaf disease recognition. Complementing this perspective, Mohanty et al. [18] showed that convolutional neural networks (CNNs) trained on large-scale leaf image datasets can achieve high accuracy across diverse crop–disease combinations, underscoring the viability of data-driven approaches for plant health monitoring. More recent surveys [5, 9] further consolidate advances in CNN- and transformer-based plant disease detection, highlighting the critical roles of dataset diversity, strong augmentation strategies, and rigorous evaluation protocols in developing reliable models.

Beyond pure end-to-end CNN classifiers, several works have explored hybrid architectures that combine deep feature extractors with traditional machine learning models in order to improve robustness and interpretability. For tomato leaf diseases, Hung et al. [15] proposed a framework that couples CNN-based feature extraction with shallow classifiers, reporting consistent gains over purely deep baselines on a real-world data set. These studies collectively suggest that hybrid pipelines can be beneficial in agricultural scenarios where data are heterogeneous, class distributions are imbalanced, and deployment constraints (e.g., edge devices) must be considered.

Rice leaf diseases have attracted increasing attention as a critical application domain for agricultural computer vision. Simhadri et al. [4] surveyed deep learning approaches for rice disease detection, covering both classification and detection pipelines, and highlighted the need for models that generalize across cultivars, growth stages, and imaging conditions. Public datasets with expert annotations, such as the Rice Disease dataset with bounding boxes for bacterial blight, brown spot, and blast lesions [10], have enabled the community to move from image-level classification toward explicit lesion localization.

A substantial body of work has adopted YOLO-style detectors for rice leaf disease detection. Haque et al. [19] developed a YOLOv5-based model for joint classification and detection of rice leaf diseases, reporting competitive precision and mean Average Precision (mAP) on a custom annotated dataset. More recent efforts focus on improving detection of small lesions and enhancing efficiency for deployment. For instance, Pan et al. [20] introduced SSD-YOLO, a lightweight detector that integrates a dynamic up sampling module and a Shape-aware IoU loss to better handle subtle lesion patterns on complex backgrounds. Fang et al. [21] proposed RLDD-YOLOv11n, which incorporates residual attention modules into the neck of YOLOv11 to boost performance on small and overlapping rice lesions, while maintaining a relatively compact model.

Other works aim at enhancing deployment flexibility on mobile and embedded platforms. Li et al. [22] presented TLI-YOLO, an improved YOLO-based architecture designed for rice disease detection with an emphasis on mobile compatibility and efficient inference, and Trinh et al. [23] investigated Wise-IoU loss within a YOLOv8-based detector to better align bounding-box regression with evaluation metrics in rice disease scenarios. These approaches primarily treat lesion localization as a one-shot regression task: the detector directly predicts bounding boxes in a single forward pass, and any localization errors are only indirectly addressed through loss design, attention mechanisms, or architectural modifications.

In contrast, our work starts from detector outputs on a bounding-box-supervised dataset [10] and explicitly models the refinement of lesion bounding boxes as a sequential decision problem. Rather than modifying the detector backbone or loss function, we learn a separate policy that iteratively adjusts an initial box toward the ground truth, which is complementary to existing YOLO-based improvements for rice disease detection.

Reinforcement learning (RL) has been explored as a means to turn object localization into a sequential decision process. Caicedo and Lazebnik [24] introduced an active object localization agent that iteratively transforms a bounding box through a set of discrete actions, guided by a deep Q-network, to focus on target objects in natural images. Bellver et al. [25] proposed a hierarchical object detection framework in which an RL agent learns where to zoom within an image, progressively refining its focus on informative regions. These works laid the foundation for viewing object detection as a control problem, where localization emerges from a sequence of bounding-box transformations rather than a single regression step.

Subsequent research has extended RL-based localization and refinement to various vision tasks. Cruciata et al. [26] investigated iterative bounding-box refinements for visual tracking, showing that multiple refinement steps can significantly improve localization stability over time. In the context of object detection and annotation quality, Ayle et al. [14] proposed BAR (Bounding-box Automated Refinement), an RL agent that learns to correct inaccurate bounding boxes produced by detectors or imperfect human annotations. BAR defines a discrete action space of translations and re scalings applied to an initial box and optimizes a reward function based on IoU improvement, demonstrating that learned policies can substantially reduce manual annotation effort while improving box quality.

These RL-based approaches differ from classical detection frameworks in that they decouple box refinement from feature extraction and classification, making it possible to plug a refinement agent on top of existing detectors or noisy annotations [27–29]. To the best of our knowledge, however, there has been limited work applying deep reinforcement learning for bounding-box refinement in agricultural disease detection, and specifically for rice leaf disease localization with bounding-box supervision. The method proposed in this paper adapts the BAR-style formulation

[14] to rice leaf disease images: an agent operates on patches cropped around initial lesion boxes and learns a sequence of actions that maximizes IoU with ground-truth lesions, thereby improving the quality of rice disease bounding boxes produced by a baseline detector without modifying its architecture.

Table 1: Bounding box refinement actions and corresponding transformations

Action	**Bounding box update**	**Action**	**Bounding box update**
up	$y_{min} -= c_1 \times h$ $y_{max} -= c_1 \times h$	left	$x_{min} -= c_1 \times w$ $x_{max} -= c_1 \times w$
down	$y_{min} += c_1 \times h$ $y_{max} += c_1 \times h$	right	$x_{min} += c_1 \times w$ $x_{max} += c_1 \times w$
taller	$y_{min} -= c_2 \times h$ $y_{max} += c_2 \times h$	wider	$x_{min} -= c_2 \times w$ $x_{max} += c_2 \times w$
shorter	$y_{min} += c_2 \times h$ $y_{max} -= c_2 \times h$	thinner	$x_{min} += c_2 \times w$ $x_{max} -= c_2 \times w$

3. Problem Formulation

Each image contains a rice leaf with a single diseased region annotated by a bounding box. Let $b^* = (x_{min}, y_{min}, x_{max}, y_{max})$ denote the ground-truth bounding box enclosing the lesion in Table 1. Given an image, a weak initialization method (e.g., detector prediction or perturbed annotation) produces an initial bounding box b. A bounding box is considered inaccurate if its Intersection-over-Union (IoU) with the ground truth is below a predefined threshold β.

Given an image and an inaccurate bounding box, the objective of the agent is to iteratively correct the bounding box so that it better overlaps with the target lesion. Figure 2 illustrates the overall refinement workflow. Starting from an inaccurate initialization, the agent applies a sequence of actions that modify the position and scale of the bounding box.

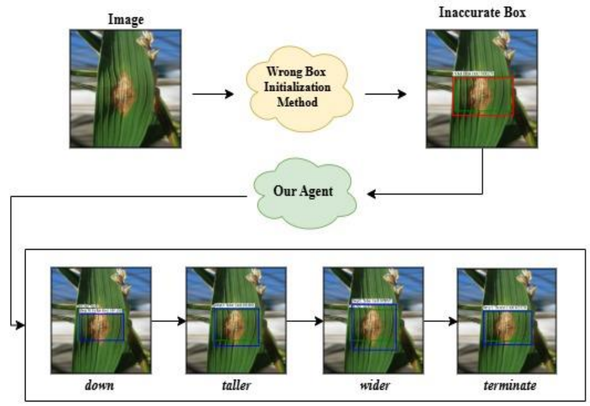

Fig. 2: Illustration of the proposed reinforcement learning–based bounding box refinement process

The refinement process is organized into episodes. At the beginning of an episode (t = 0), the bounding box is initialized by a weak initialization method. At each step, the agent selects one action from a fixed set of translation and scaling operations that adjust the current bounding box. The quality of the updated bounding box is evaluated based on its IoU with the ground-truth box.

In addition to translation and scaling actions, the agent can choose a terminate action to end the episode. If the IoU of the current bounding box exceeds the threshold β, the correction is considered successful; otherwise, the episode terminates unsuccessfully. To prevent excessively long correction sequences, the maximum length of an episode is limited to a fixed number of steps.

Given the goal of the agent and the dynamics of an episode, the bounding box refinement task is formulated as a sequential decision-making problem. The agent aims to learn an action sequence

$$T_E = \{a_1, a_2, \ldots, a_k\}, |T_E| \leq T_{max}, \quad (1)$$

that transforms an initial inaccurate bounding box into an accurate one that tightly localizes the rice leaf disease region.

4. Proposed Method

We propose a deep reinforcement learning agent that refines an initial lesion bounding box through a sequence of discrete geometric transformations. The refinement policy is learned using Deep Q-Learning, where a neural network estimates the action-value function Q (s, a) over a fixed action set (translation, scaling, and a termination action).

4.1. State representation

At the beginning of an episode, an initial bounding box b0 is provided by a weak initialization procedure. From b_0, we crop an image patch and resize it to a fixed resolution (224 × 224). This patch is kept fixed throughout the episode, while the current bounding box $b_t = (x^t_{min}, y^t_{min}, x^t_{max}, y^t_{max})$ evolves across steps. The state is defined as:

$$s_t = (I_0, v_t), \quad (2)$$

where I_0 is the fixed cropped patch and $v_t \in R^4$ is the coordinate vector of the current bounding box.

4.2. Q-network architecture

The Q-network consists of a frozen convolutional backbone and a lightweight fully connected head. The backbone extracts a 2048-dimensional visual embedding from the fixed patch:

$$\phi_0 = f_{cnn}(I_0) \in R^{2048}. \quad (3)$$

The embedding is concatenated with the current box vector and mapped to action values:

$$z_t = [\phi_0; v_t], \ Q(s_t, \cdot) = f_{mlp}(z_t) \in R^{|A|} \quad (4)$$

Freezing the backbone reduces sample complexity and stabilizes learning on limited agricultural data, while the MLP head learns the refinement policy.

4.3. Action execution

At each step, the agent selects one action at ∈ A to update the current box. The action set contains eight geometric transformations (four translations and four symmetric scaling) and one terminate action. Translation updates move the box by a fraction c_1 of its current width/height, and scaling updates expand/shrink the box by a fraction c_2 of its current width/height. After each update, box coordinates are clipped to valid image boundaries and constrained to maintain a valid rectangle. The episode length is capped by a fixed maximum number of steps.

4.4. Reward design

The reward encourages IoU improvement between consecutive steps. Let IoUt =IoU(b_t, b⋆) be the overlap with the ground-truth box b⋆. For non-termination actions, the step reward is defined by the IoU change:

$$r_t = \begin{cases} +1 \text{, if } IoU_{t+1} > IoU_t \\ -1, \text{if } IoU_{t+1} \leq IoU_t \text{ and } IoU_{t+1} < \beta \\ 0, \text{otherwise.} \end{cases} \quad (5)$$

where β is the minimum IoU threshold used to indicate insufficient localization quality. For the terminate action, a terminal reward is given to encourage stopping only when the box is sufficiently accurate:

$$r_t = \begin{cases} +4, \text{if } IoU_t \geq \tau \\ -4, \text{otherwise} \end{cases} \quad (6)$$

where τ is the success threshold for termination.

4.5. Learning objective

The agent is trained offline using experience replay. Each transition (s_t, a_t, r_t, s_{t+1}, d_t) is stored in a replay buffer, where d_t indicates whether the episode terminates. The Q-network is optimized by minimizing the temporal-difference loss:

$$y_t = r_t + \gamma(1 - d_t)\, max_{a'} Q(s_{t+1}, a'), \quad (7)$$

$$L = E[(Q(s_t,a_t)-y_t)^2], \quad (8)$$

with discount factor γ. During training, actions are selected using an ϵ-greedy strategy to balance exploration and exploitation, and ϵ is gradually decayed to favor the learned refinement policy.

5. Evaluation

To rigorously evaluate the effectiveness of the proposed method, we conducted a comprehensive assessment including both experimental configuration and performance analysis. First, we detail the datasets, implementation parameters, training protocols, and underlying methods used for comparison, ensuring the feasibility and fairness of the evaluation process. Next, we present quantitative and qualitative findings, highlighting the advantages of the proposed method and examining its behavior in various scenarios.

5.1. Experimental Settings

The study was executed on a high-performance workstation to ensure consistency and reproducibility, utilizing the Windows 11 Pro operating system, powered by an Intel(R) Core(TM) i5-13600K CPU and accelerated by an NVIDIA GeForce RTX 3060, which was essential for handling the large computational load associated with deep Q-network training via the PyTorch framework.

In this study, we utilized the public Rice Disease Dataset [30]. This dataset comprises high-resolution images collected under diverse conditions, capturing realistic agricultural challenges such as varying illumination, complex backgrounds, and overlapping leaves. It consists of three main disease categories: Brown Spot (characterized by small, circular lesions), Bacterial Leaf Blight (exhibiting elongated, irregular streaks), and Rice Leaf Blast, as shown in Figure 3. To frame the task as a localization problem, we utilized the expert-annotated ground-truth bounding boxes enclosing the affected areas on each image. The dataset was partitioned into training (60%), validation (20%), and testing (20%) sets to ensure a robust evaluation of the refinement agent.

(a) Brown Spot (b) Leaf Blight (c) Rice Leaf Blast

Fig. 3: Representative samples of the three disease categories in the Rice Disease Dataset

Algorithm 1 Synthetic shifted boxes for weak initialization

for annotation $\mathcal{J}$ in a split **do**
  Read (W, H) and $b^* = (x_1, y_1, x_2, y_2)$
  Sample $(\delta x_1, \delta y_1, \delta x_2, \delta y_2) \sim U(\{-s, \cdots, s\})^4$
    $x_1{}' \leftarrow (x_1 + \delta x_1, 0, W - 1)$
    $y_1{}' \leftarrow (y_1 + \delta y_1, 0, H - 1)$
    $x_2{}' \leftarrow (x_2 + \delta x_2, 0, W - 1)$
    $y_2{}' \leftarrow (y_2 + \delta y_2, 0, H - 1)$
  if $x_1{}' \geq x_2{}'$ **then**
    $x_2{}' \leftarrow \min(W - 1, x_1{}' + 1)$
  end if
  if $y_1{}' \geq y_2{}'$ **then**
    $y_2{}' \leftarrow \min(H - 1, y_1{}' + 1)$
  end if
  Save $b_0 = (x_1{}', y_1{}', x_2{}', y_2{}')$
end for

Algorithm 1 describes how we synthesize a weak initialization bounding box by perturbing the original ground-truth annotation.

Fig. 4: Examples of ground-truth boxes (green) and synthetically shifted boxes (red) used for weak initialization.

For each image, we randomly shift the four sides of the box within a bounded range to simulate the type of localization noise that commonly appears in automatic detectors or imperfect manual labeling. The perturbed coordinates are then clipped to stay inside the image boundary, and a simple validity check is applied to prevent degenerate cases (e.g., inverted or collapsed boxes). These shifted boxes serve as consistent, controllable starting points for training and evaluating the refinement agent, allowing us to study how well the agent corrects inaccurate proposals toward tighter lesion localization. To make this perturbation process tangible, Figure 4 visualizes several examples where the original ground-truth box is overlaid together with its synthetically shifted counterpart. In each image, the ground-truth box tightly encloses the lesion region, while the shifted box is deliberately displaced and may partially miss the target area or include additional background. These qualitative examples highlight the kind of localization noise we aim to correct: the shifted boxes provide a realistic “imperfect” starting point and the refinement agent is trained to progressively move back toward the tighter ground-truth localization.

5.2. Experimental Results

Next, we present the experimental results of the DRL-based fine-tuning agent on the rice leaf disease test dataset.

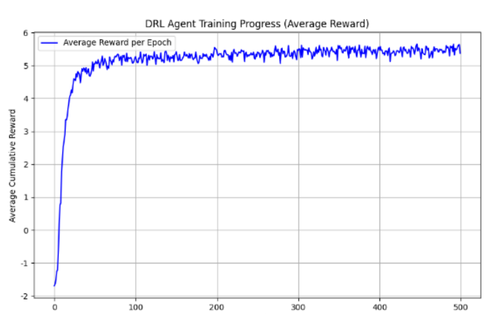

Fig. 5: DRL Agent Training Progress: Average Cumulative Reward over 500 Epochs.

Figure 5 shows the training progress of the DRL agent as measured by the average cumulative reward over 500 epochs. We observe a clear and stable learning pattern: the reward increases sharply from an initial negative value to a high positive level in the first few epochs and then gradually saturates. After the first 50–100 epochs, the curve stabilizes, fluctuating around 5.0–5.5 with small variations. This behavior indicates that the agent rapidly learns an effective fine-tuning policy and maintains stable decision-making throughout the remaining training process. The detailed hyper parameter configuration used to train the proposed DRL refinement agent is summarized in Table 2, including the network architecture, state representation, action space, optimization settings, and reward-related parameters. These hyper parameters were selected to ensure stable convergence and efficient learning during training.

Table 2: Key hyper parameters of the proposed DRL refinement agent.

Hyperparameter	Setting
CNN backbone	ResNet-101 (ImageNet pre trained, frozen)
Patch preprocessing	resize 224×224, ImageNet normalization
State vector	[ϕ(I0) ∈ R^{2048}; bt ∈ R^4]
Q-network head	MLP: 2052 → 512 → 9 (ReLU, Dropout p=0.5)
Action set	9 actions (8 geometric transforms + terminate)

Translation step	c_1 = 0.10 (fraction of current width/height)
Scaling step	c_2 = 0.05 (fraction of current width/height)
Episode horizon	Tmax = 10
Discount factor	γ = 0.90
Optimization	Adam, learning rate 1 × 10^{-4} (train MLP only)
IoU thresholds / terminal reward	β = 0.60, τ = 0.75, r_term = ±4

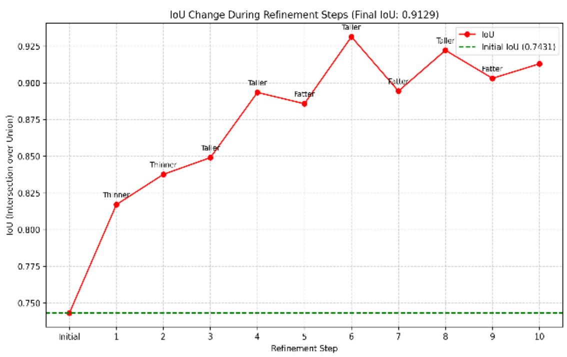

Fig. 6: Representative Refinement Trajectory: IoU Change and Agent Actions over 10 Steps

To better understand how the policy improves the bounding box step by step Figure 6 illustrates a typical 10-step fine-tuning example. Starting from an initial IoU of 0.7431 (dashed green baseline), the agent continuously improved its positioning and reached a final IoU of 0.9129. Much of the improvement occurred in the early steps, where the agent applied scaling/frame adjustments (e.g., wider, thinner, taller) to quickly align the proposal with the target area. Subsequent steps refined the box using a combination of geometric adjustments and small shifts (e.g., left/right), producing incremental improvements. Importantly, the trajectory was not entirely monotonous: occasional small IoU drops occurred between consecutive steps, but the agent typically compensated with subsequent actions and recovered toward a high overlap configuration.

Table 3: Refinement performance on the dataset.

Class	**IoU**		**Success (%)**		**Avg. steps**
	Init	Final	@0.50	@0.75	
Overall	0.53 ± 0.18	0.89 ± 0.14	91.9	88.6	7.1
Brown Spot	0.53 ± 0.17	0.87 ± 0.13	92.8	87.1	8.1
Bacterial Leaf Blight	0.57 ± 0.19	0.91 ± 0.15	92.3	89.5	7.9
Rice Leaf Blast	0.50 ± 0.18	0.91 ± 0.14	90.6	89.2	5.3

Table 3 summarizes the quantitative refinement performance on the test split. Overall, the agent substantially improves localization quality, increasing the mean IoU from 0.53 to 0.89, while maintaining a high success rate at both evaluation thresholds (90.4% at IoU≥0.50 and 88.6% at IoU≥0.75). The refinement effect is consistent across all three disease categories: each class exhibits a clear gain from the initial to the final IoU, with success@0.75 remaining above 87% in every case. In terms ofefficiency, the agent typically converges within a small number of actions (7.1 steps on average overall), although different classes require different refinement lengths, reflecting variations in lesion appearance and box difficulty.

To justify the design choices of the proposed RL-RiceBox framework and address alternative refinement strategies, we conducted an ablation study comparing our learned RL policy against a heuristic bounding-box refinement baseline. The heuristic baseline iteratively applies uniform scaling (expanding or shrinking by a fixed ratio) and random small translations, greedily accepting updates only if the IoU improves. While this heuristic approach provides minor localization gains over the initial noisy boxes, it struggles to adapt to the irregular shapes and extreme aspect ratios of rice leaf lesions. It frequently gets stuck in local optima, failing to surpass a mean IoU of 0.65 on our test set. In contrast, the RL agent dynamically selects context-aware actions based on visual features, achieving a significantly higher final mean IoU of 0.89. This confirms that a learned policy is strictly necessary for handling agricultural localization noise, surpassing simple geometric tuning. Furthermore, we analyzed the sensitivity of the refinement performance to the geometric action step sizes, specifically the translation fraction (c_1) and the scaling fraction (c_2). We evaluated variations of these parameters around our default configuration (c_1 = 0.10, c_2 = 0.05). Empirical observations indicate that increasing the step sizes significantly (e.g., $c_1 \geq 0.20$, $c_2 \geq 0.10$) leads to faster initial IoU improvements but causes the agent to oscillate around the ground-truth box in later steps, reducing the final convergence quality and success rate at the 0.75 IoU threshold. Conversely, excessively small step sizes (e.g., c_1 = 0.05, c_2 = 0.02) provide fine-grained adjustments but require an episode horizon much longer than Tmax = 10 to achieve satisfactory refinement, making the process computationally inefficient. Thus, the selected values of c_1 and c_2 offer an optimal trade-off between refinement granularity and convergence speed.

6. Conclusions

In this work, we introduced a DRL-based bounding-box refinement approach for rice leaf disease localization. Starting from a weak (shifted) initialization, the agent learns a sequence of geometric actions that progressively improves the overlap between the predicted box and the lesion region, leading to consistently tighter localization across disease categories. Beyond the quantitative gains, the refinement behavior is interpretable: the agent typically performs larger scaling adjustments in early steps and then applies smaller translations to fine-tune the final box. However, our study acknowledges several limitations. Foremost, the weak initialization is synthetically generated by shifting ground-truth boxes. While this isolates the refinement capability, it may not fully capture the structured error patterns—such as algorithmic biases or missed detections— produced by real detectors. Furthermore, although the refinement process is computationally lightweight, iterative inference inherently introduces additional latency compared to one-shot localization.

Future work will focus on replacing the synthetic initialization with actual proposals generated by state-of-the-art detectors (e.g., YOLO variants) to evaluate the framework's end-to-end performance under real-world error distributions. Additionally, to address practical applicability and bridge the gap between algorithmic refinement and field deployment, we aim to integrate the detection models directly into agricultural hardware. This will involve applying quantization and pruning techniques to the refinement module to significantly reduce memory footprint and system latency while maintaining high localization accuracy. We also plan to extend the refinement to multiple instances per image under standard detection metrics.

7. Competing Interests

The authors declare that they have no competing interests.

8. Funding Information

This research received no external funding.

9. Author Contribution

All authors contributed equally to the conception, design, analysis, and writing of this manuscript.

10. Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

11. Research Involving Human and/or Animals

Not applicable.

12. Informed Consent

Not applicable.

References

[1] Mohapatra, P.K., Sahu, B.B.: Importance of rice as human food. In: Mohapatra, P.K. (ed.) Panicle Architecture of Rice and Its Relationship with Grain Filling, pp.1–25. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-8244 -2 1

[2] Rezvi, H.U.A., Tahjib-Ul-Arif, M., Azim, M.A., Tumpa, T.A., Tipu, M.M.H.,Najnine, F., Dawood, M.F., Skalicky, M., Bresti ̌c, M.: Rice and food security: Climate change implications and the future prospects for nutritional security. Food and Energy Security 12(1), 430 (2023)

[3] Bandumula, N.: Rice production in asia: Key to global food security. Proceedings of the National Academy of Sciences, India Section B: Biological Sciences 88(4), 1323–1328 (2018)

[4] Simhadri, C.G., Kondaveeti, H.K., Vatsavayi, V.K., Mitra, A., Ananthachari, P.: Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques. Information Processing in Agriculture 12(2), 151–168 (2025) https://doi.org/10.1016/j.inpa.2024.04.006

[5] Abade, A., Ferreira, P.A., Barros Vidal, F.: Plant diseases recognition on images using convolutional neural networks: A systematic review. Computers and Electronics in Agriculture 185, 106125 (2021) https://doi.org/10.1016/j.compag. 2021.106125

[6] Lu, J., Tan, L., Jiang, H.: Review on convolutional neural network (cnn) applied to plant leaf disease classification. Agriculture 11(8), 707 (2021)

[7] Hung, N., Huynh, P., Tung, M., Vu, N., Dat, N.: Clvm: a hybrid deep learning framework for contactless virtual mouse control. Informatics and Automation 25(2), 378–409 (2026)

[8] Ngugi, H.N., Ezugwu, A.E., Akinyelu, A.A., Abualigah, L.: Revolutionizing crop disease detection with computational deep learning: a comprehensive review. Environmental Monitoring and Assessment 196(3), 302 (2024)

[9] Pacal, I., Kunduracioglu, I., Alma, M.H., et al.: A systematic review of deep learning techniques for plant diseases. Artificial Intelligence Review 57, 304 (2024) https://doi.org/10.1007/s10462-024-10944-7

[10] Shrestha, N.L., Pandey, P., Tiwari, A., Giri, R.: Rice Disease Dataset. Kaggle (2021). https://doi.org/10.34740/KAGGLE/DSV/2481060

[11] Jhatial, M.J., Shaikh, R.A., Shaikh, N.A., Rajper, S., Arain, R.H., Chandio, G.H., Bhangwar, A.Q., Shaikh, H., Shaikh, K.H.: Deep learning-based rice leaf diseases detection using YOLOv5. Sukkur IBA Journal of Computing and Mathematical Sciences 6(1), 49–61 (2022) https://doi.org/10.30537/sjcms.v6i1.1009

[12] Li, P., Zhou, J., Sun, H., Zeng, J.: Rdrm-yolo: A high-accuracy and lightweight rice disease detection model for complex field environments based on improved yolov5. Agriculture 15(5), 479 (2025)

[13] Kumar, V.S., Jaganathan, M., Viswanathan, A., Umamaheswari, M., Vignesh, J.: Rice leaf disease detection based on bidirectional feature attention pyramid network with yolo v5 model. Environmental Research Communications 5(6), 065014 (2023)

[14] Ayle, M., Tekli, J., El-Zini, J., El-Asmar, B., Awad, M.: Bar—a reinforcementlearning agent for bounding-box automated refinement. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 2561–2568 (2020)

[15] Hung, N.V., Dinh, H.P., Thoa, N.T.N., Nam, L.M., Trang, L.T.H., Hoang, T.- M.: Hybrid deep learning and machine learning framework for automated

tomato leaf disease classification. ECTI Transactions on Computer and Information Technology (ECTI-CIT) 19(4), 742–753 (2025) https://doi.org/10.37936/ecti-cit.2025194.262810

[16] Hung, N., Dat, P., Tan, N., Quan, N., Trang, L., Nam, L., et al.: Heverl–viewport estimation using reinforcement learning for 360-degree video streaming. Informatics and Automation 24(1), 302–328 (2025)

[17] Hung, N., Loi, T., Binh, N., Nga, N., Huong, T., Luu, D., et al.: Building an online learning model through a dance recognition video based on deep learning. Informatics and Automation 23(1), 101–128 (2024)

[18] Mohanty, S.P., Hughes, D.P., Salath´e, M.: Using deep learning for image-based plant disease detection. Frontiers in Plant Science 7, 1419 (2016) https://doi.org/10.3389/fpls.2016.01 419

[19] Haque, M.E., et al.: Rice leaf disease classification and detection using YOLOv5. arXiv preprint arXiv:2209.01579 (2022)

[20] Pan, C., Wang, S., Wang, Y., Liu, C.: Ssd-YOLO: A lightweight network for rice leaf disease detection. Frontiers in Plant Science 16, 1643096 (2025) https://doi.org/10.3389/fpls.2025.1643096

[21] Fang, K., Zhou, R., Deng, N., Li, C., Zhu, X.: Rldd-YOLOv11n: Research on rice leaf disease detection based on YOLOv11. Agronomy 15(6), 1266 (2025) https://doi.org/10.3390/agronomy15061266

[22] Li, Z., Wu, W., Wei, B., Li, H., Zhan, J., Deng, S., Wang, J.: Rice disease detection: TLI-YOLO innovative approach for enhanced detection and mobile compatibility. Sensors 25(8), 2494 (2025) https://doi.org/10.3390/s25082494

[23] Trinh, C.D., Le, T.M.D., Do, T.H., Bui, N.M., Nguyen, T.H., Ngo, Q.U.: Improving YOLOv8 deep learning model in rice disease detection by using wise-IoU loss function. Journal of Measurement, Control and Automation 29(1), 1–6 (2025)

[24] Caicedo, J.C., Lazebnik, S.: Active object localization with deep reinforcement learning. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2488–2496 (2015). https://doi.org/10.1109/ICCV.2015.286

[25] Bellver, M., Gir´o-i-Nieto, X., Marques, F., Torres, J.: Hierarchical object detection with deep reinforcement learning. In: Deep Reinforcement Learning Workshop, NIPS (2016)

[26] Cruciata, G., Lo Presti, L., La Cascia, M.: Iterative multiple bounding-box refinements for visual tracking. Journal of Imaging 8(3), 61 (2022) https://doi.org/10.3390/jimaging8030061

[27] Arshad, K., Ali, R.F., Muneer, A., Aziz, I.A., Naseer, S., Khan, N.S., Taib, S.M.: Deep reinforcement learning for anomaly detection: A systematic review. Ieee Access 10, 124017–124035 (2022)

[28] Kheddar, H., Dawoud, D.W., Awad, A.I., Himeur, Y., Khan, M.K.: Reinforcement-learning-based intrusion detection in communication networks: A review. IEEE Communications Surveys & Tutorials (2024)

[29] Yang, W., Acuto, A., Zhou, Y., Wojtczak, D.: A survey for deep reinforcement learning-based network intrusion detection. arXiv preprint arXiv:2410.07612 (2024)

[30] Shrestha, N.L., Pandey, P., Tiwari, A., Giri, R.: Rice Disease Dataset. Kaggle (2021). https://doi.org/10.34740/KAGGLE/DSV/2481060