

Boundary-Refined DeepLabV3+ for Crop Disease Detection in Greenhouse Vegetable Images

Milorad Vlahović^{1,*} and Gojko Vasiljev¹

¹ Faculty of Information Technologies and Engineering, Union Nikola Tesla University, Belgrade, 11070, Serbia

*Corresponding author: info@fsom.edu.rs

Abstract. Accurate pixel-level delineation of crop disease in greenhouse images is challenging due to weak lesion margins, scale variations, leaf-vein interference, specular highlights and partial occlusion. This paper introduces a boundary-refined DeepLabV3+ model that keeps the encoder-decoder efficiency of the original framework and adds a boundary supervision branch, uncertainty-aware cross-scale fusion, and an adaptive refinement gate. The network is tested on a set of 6,840 curated greenhouse images that include leaves of tomatoes, cucumbers, peppers and eggplants, as well as 8 disease categories and healthy tissue. Under a fixed split by greenhouse compartment, the proposed model has achieved a mean intersection over union of 89.7%, a mean F1 score of 94.8%, and a boundary F1 of 86.9% at a two-pixel tolerance. The corresponding values are 3.8, 2.4 and 7.6 percentage points higher than those of the standard DeepLabV3+. The mean Intersection over Union (IoU) under low-illumination and condensation-blur conditions are 3.1 and 2.7, respectively. Ablation studies show that boundary supervision is responsible for most of the contour improvement, and uncertainty-aware fusion reduces false lesion expansion along veins. The model has 31.6 million parameters and, after mixed-precision optimisation, runs at 18.7 frames per second on an embedded graphics chip. Based on the above results, explicit boundary reasoning can improve the precision of disease-area estimation without sacrificing the efficiency required in practice; it is thus suitable for greenhouse scouting, targeted spraying and longitudinal severity assessment.

Keywords: *Semantic Segmentation, DeepLabV3+, Boundary Refinement, Greenhouse Vegetables, Crop Disease Detection*

Received on 19 October 2025, Accepted on 25 April 2026, Published on 04 May 2026

Copyright © 2026 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Greenhouse vegetable production has increased the scale of high-density planting, controlled irrigation and extended the growing season to boost production; at the same time, such good conditions can favour the rapid spread of foliar pathogens if early symptoms are not detected [1]. Camera-based scouting can cover more plants than manual inspection and is suitable for keeping a reproducible visual record of disease development [2]. Therefore, image classification has been used as a general way to automatically assess the health of plants [3]. However, a class label alone cannot show the extent of tissue damage or distinguish between two adjacent lesions that require different treatment [4]. Pixel-level segmentation is more suitable for severity grading, robotic inspection and variable-rate spraying, and can link the diagnostic results to a spatially defined area [5]. Greenhouse scenes still have leaf overlap, support wires, reflective films, irrigation drops and non-uniform artificial light; thus, they are more difficult to segment than recognition of isolated leaves [6].

Convolutional encoder-decoder networks have enhanced the localisation of disease areas by integrating semantic context with gradually recovered spatial details [7]. Atrous convolution is quite effective because it can increase the receptive field without reducing the resolution of the feature map repeatedly [8]. DeepLabV3+ is a lightweight decoder that uses atrous spatial pyramid pooling to maintain good performance in crop imagery [9]. However, it may be difficult to see fine necrotic rims, spread along high-contrast veins, and fail to identify

fragmented chlorotic boundaries if the texture of the lesion is similar to that in healthy tissue [10]. General attention modules can reweight channels or spatial locations, but they do not guarantee that the attended response follows the true lesion contour [11]. Conditional Random Fields or morphological post-processing can be used to refine masks, but fixed parameters often exhibit inconsistencies in disease type, image scale and illumination [12].

This paper introduces a boundary-refined DeepLabV3+ model to address the problem of lower semantic accuracy and geometric imprecision at the boundary level for greenhouse vegetable disease detection. Extract high-resolution edge evidence from the early encoder stage in the proposed design, align it with decoder semantics via uncertainty-aware fusion, and add explicit boundary supervision during optimisation. A refinement gate focuses corrective computation on ambiguous pixels and does not modify the entire mask uniformly. Four vegetable species, multiple disease categories, and controlled nuisance conditions were selected as the evaluation objects to examine contour quality, disease-area bias, robustness, and embedded throughput in this study. The purpose of the above analysis is to determine when boundary modelling is more beneficial than a stronger region encoder.

Related Work

Crop Disease Recognition and Segmentation

Early plant-disease vision systems used color indices, handcrafted textures and manually set segmentation thresholds, and were easily affected by changes in camera or cultivar [13]. Deep convolutional classifiers have reduced the demand for hand-crafted descriptors and achieved good results on curated leaf datasets [14]. Transfer learning also improved the performance of agricultural training sets with a small amount [15]. Field-oriented collections subsequently revealed an enduring domain gap between laboratory leaves and messy production sites [16]. Detector-based systems localised symptomatic leaves or coarse lesion boxes, but box supervision retained healthy background and could not support precise severity measurement [17]. Therefore, semantic segmentation is now used to convert visual diagnoses into spatial evidence at the level of leaves and lesions [18].

Recently, studies on crop segmentation have used encoder-decoder networks, feature pyramids and attention blocks to retain small symptoms at different scales. Their reported gains often focus on mean intersection over union, and boundary displacement and disease-area bias are given less attention. This choice of index may overlook serious errors in the clinic: a model can achieve a high region score by expanding lesions along veins or erasing thin margins. There are many defects in greenhouse images; for example, there are repetitive leaf textures and controlled but reflective lights that are close to the disease symptoms.

Boundary-Aware Semantic Segmentation

General-purpose semantic segmentation has begun to introduce multiscale context modules, high-resolution feature exchange, and transformer-based global modeling [19]. DeepLab-style atrous pyramids are still efficient because they sample multiple receptive fields using a stable convolutional backbone [20]. High-resolution networks preserve geometric details through parallel representations, and transformer encoders address long-range dependencies at an extended memory cost [21]. Lightweight segmentation families trade some accuracy for faster deployment by using factorized convolutions and compact decoders [22]. These developments provide useful components, but their region objectives do not explicitly penalise topological defects at lesion margins.

Boundary-aware methods introduce edge branches, contour losses, distance transforms or uncertainty maps to focus learning on hard pixels [23]. A fixed edge branch may be too sensitive to leaf vein and greenhouse structure changes when low-level gradients are treated as lesion evidence. A single-type semantic decoder will not be able to recognise a weak but valid margin. Therefore, the design problem is not to add edges; rather, it is to condition boundary correction on semantic confidence and cross-scale agreement. The above model achieves the same goal by combining boundary evidence with decoder uncertainty and supervising region, contour and consistency terms in a single trainable framework.

Boundary-Refined DeepLabV3+ Model

Multiscale Encoder and Supervised Boundary Branch

The input image is resized to 512×512 pixels after aspect-preserving padding and photometric normalization. A ResNet-50 encoder produces feature tensors at strides 4, 8, 16, and 16, with the final stage using atrous convolution. The atrous spatial pyramid pooling module contains rates 1, 6, 12, and 18 plus global pooling. Its output supplies lesion semantics, while the stride- 4 and stride- 8 features retain vein, spot, and margin cues. Figure 1 presents the complete pipeline once: image normalization leads to multiscale encoding, parallel context aggregation, boundary extraction, uncertainty-aware fusion, and final mask refinement before disease-area estimation.

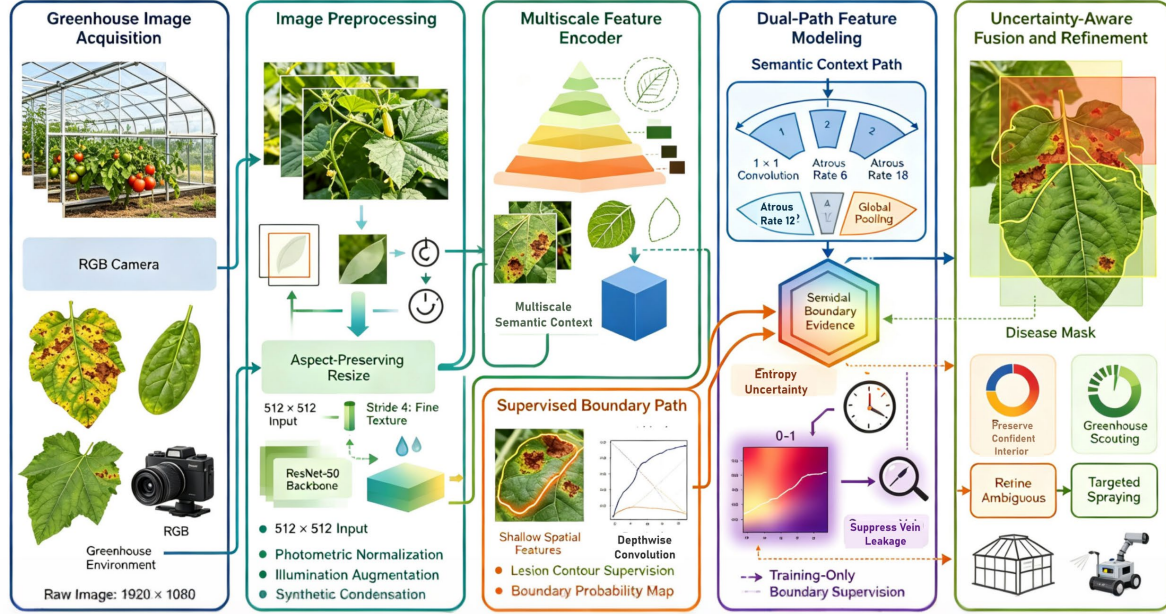


Figure 1. Overall workflow of the boundary-refined DeepLabV3+ model for greenhouse vegetable disease segmentation.

Let X denote an RGB image and F_l the feature tensor produced at encoder level l . The encoder mapping is written as

$$F_l = E_l(F_{l-1}; \theta_l), F_0 = X \quad \text{Eq.(1)}$$

The atrous context response at dilation rate r samples the feature lattice without reducing its resolution:

$$A_r(p) = \sum_{k \in \mathcal{K}} w_k F_4(p + rk) \quad \text{Eq.(2)}$$

The context tensors are concatenated and projected to obtain the semantic representation S :

$$S = \phi([A_1, A_6, A_{12}, A_{18}, G(F_4)]) \quad \text{Eq.(3)}$$

A shallow boundary branch combines the stride-4 and upsampled stride-8 features. A 1×1 projection first equalizes their channel dimensions, after which depthwise 3×3 convolution suppresses isolated texture noise. The boundary probability B is defined by

$$B = \sigma(\psi(DWConv[P_4(F_1), \uparrow P_8(F_2)])) \quad \text{Eq.(4)}$$

This branch differs from a generic edge detector because its target is generated only from annotated lesion contours. Leaf outlines, veins, and greenhouse wires are therefore treated as hard negatives during supervised learning.

Encoder initialisation uses ImageNet weights, but all stages are still trainable. The batch-normalisation statistics are stored during training and then fixed for deployment to prevent drift caused by single-frame batches. A 48-channel projection is used for the stride-4 skip connection, and adding more channels to the expanded memory did not improve the accuracy of the validation contour. The boundary branch has 64 channels and does not share parameters with the atrous pyramid. Therefore, the high-frequency signal will not be absorbed by the

semantic representation that has been designed to be invariant to small spatial variations. Bilinear interpolation is used for feature alignment before concatenation, and corner alignment is disabled to ensure consistent coordinate mapping between training and tiled inference. Reflection padding is added at the tile boundaries to avoid the dark borders near the edge of the image caused by zero padding.

ResNet-50 is selected as the backbone in an engineering trade-off, not because deeper encoders are universally superior. A ResNet-101 trial increased the mean intersection over union by only 0.4 points, added 18.7 million parameters, and reduced embedded throughput below the platform target. A mobile backhauls improved speed but reduced small-lesion recall. Therefore, the selected backbone has relatively little computational overhead for boundary reasoning and is more directly relevant to the problem mentioned above. Feature maps are kept at output stride 16; output stride 8 improves thin-contour sampling but nearly doubles activation memory at the selected crop size. The boundary branch recovers most of the geometric detail at a lower cost by working with early features.

Uncertainty-Aware Fusion and Adaptive Refinement

Decoder uncertainty identifies pixels for which semantic evidence is insufficient. For a K -class softmax vector p_i at pixel i , normalized entropy provides the uncertainty score

$$u_i = -\frac{1}{\log K} \sum_{c=1}^K p_{ic} \log(p_{ic} + \varepsilon) \quad \text{Eq.(5)}$$

The boundary feature is gated by uncertainty rather than added uniformly. With projected semantic features S_i and boundary features E_i , the fusion coefficient is

$$\alpha_i = \sigma(W_\alpha[S_i, E_i, u_i]) \quad \text{Eq.(6)}$$

and the fused representation becomes

$$H_i = S_i + \alpha_i \odot W_e E_i \quad \text{Eq.(7)}$$

The above method has the property of stability of strong semantic regions under perturbation and gives higher weight to the boundary influence of ambiguous neighbours. The decoder concatenates the fused tensor with a low-level encoder projection and predicts a coarse mask. Figure 2 shows the first refinement logic: entropy and contour confidence form a spatial gate; candidate boundary bands are updated via residual logits; and stable interior pixels skip the correction path.

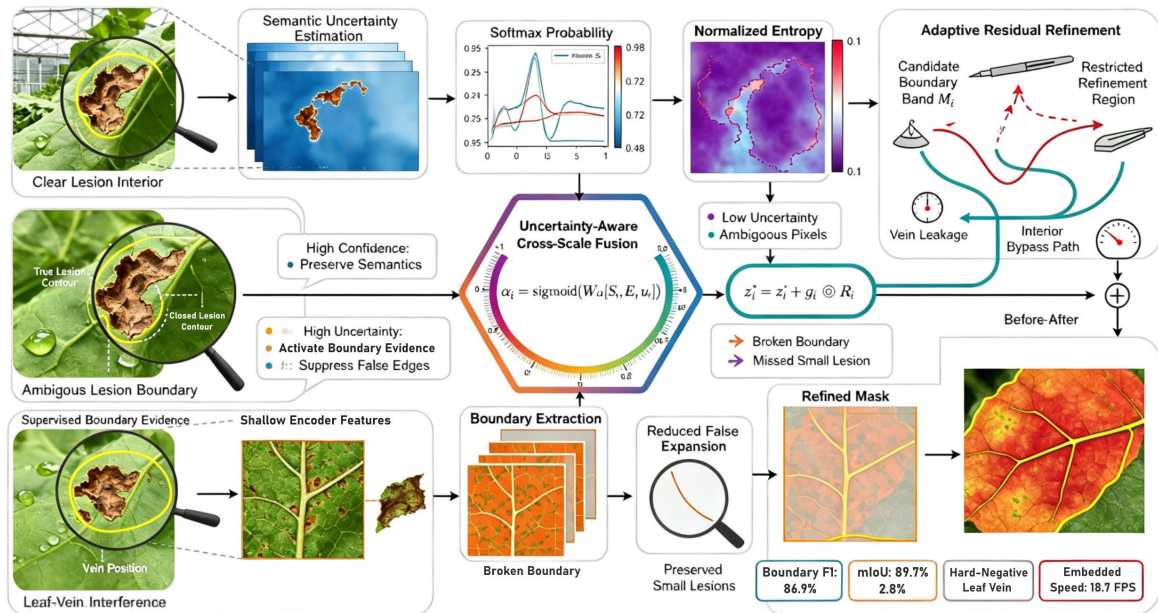


Figure 2. Uncertainty-aware boundary fusion and adaptive residual refinement mechanism.

The adaptive refinement gate operates on logits z_i . Its residual update is

$$z_i^* = z_i + g_i \odot R_i, g_i = M_i \sigma(W_g[H_i, u_i, B_i]) \quad \text{Eq.(8)}$$

where R_i is a learned correction and g_i is restricted by a dilated candidate boundary band. The restriction limits computation and reduces false sharpening on textured healthy areas.

Entropy is computed from temperature-normalised class probabilities in training. Without normalisation, the first few rounds will have uniformly high uncertainty and allow the boundary branch to take over before the semantic features are useful. Temperature annealing from 1.5 to 1.0 will be carried out over the first 30 epochs. Gradient Clipping is used to limit the norm of the gradient at the gate to 5, and the mean gate activation is recorded for diagnosis. It is close to 0.27 in the validation images, so most of the pixels follow the semantic path, and only a small number are corrected at the boundary. Visual gate maps are typically located near the borders of lesions, in narrow gaps between spots, and show faint transitions from chlorotic to healthy tissue. The response at the leaf vein is strong initially and then falls as a result of continuous negative evidence from the supervised contour target.

The candidate band M is formed by dilating the predicted contour and the semantic transition set during training. Using only the reference contour would expose annotation geometry at inference, whereas using only predicted boundaries can lock the model into an early error. The union strategy supplies a curriculum: reference transitions support stable learning during initial epochs, and predicted transitions gradually dominate as accuracy improves. At inference, M depends exclusively on the predicted mask and boundary probability. A hyperbolic tangent is used to limit the refinement residual, thus avoiding the reversal of a confident interior class by a single high-frequency feature. This bounded update also improves the stability of mixed-precision by keeping the residual logits in a reasonable range.

Objective Function, Training, and Inference

Training uses a compound objective so that region recognition, contour alignment, and local consistency are optimized together. The multiclass cross-entropy term is

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_i \sum_c y_{ic} \log p_{ic} \quad \text{Eq.(9)}$$

To reduce imbalance between small lesions and background, a soft Dice term is added:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \sum_i p_i y_i + \varepsilon}{\sum_i p_i^2 + \sum_i y_i^2 + \varepsilon} \quad \text{Eq.(10)}$$

Boundary supervision uses a binary contour target derived by morphological gradient at a two-pixel radius:

$$\mathcal{L}_B = -\frac{1}{N} \sum_i [b_i \log B_i + (1 - b_i) \log(1 - B_i)] \quad \text{Eq.(11)}$$

A distance-weighted contour alignment term penalizes predictions according to their displacement from the closest reference boundary:

$$\mathcal{L}_D = \frac{1}{N} \sum_i |p_i - y_i| \exp\left(-\frac{d_i}{\sigma}\right) \quad \text{Eq.(12)}$$

Local consistency discourages abrupt class changes away from annotated contours:

$$\mathcal{L}_{cons} = \frac{1}{|\Omega \setminus \Gamma|} \sum_{i \in \Omega \setminus \Gamma} \sum_{j \in \mathcal{N}(i)} \|p_i - p_j\|_1 \quad \text{Eq.(13)}$$

The complete objective is

$$\mathcal{L} = \mathcal{L}_{CE} + 0.7\mathcal{L}_{Dice} + 0.4\mathcal{L}_B + 0.2\mathcal{L}_D + 0.1\mathcal{L}_{cons} \quad \text{Eq.(14)}$$

with weights 1.0, 0.7, 0.4, 0.2, and 0.1 selected on the validation set. AdamW optimization starts at 3×10^{-4} with cosine decay, a five-epoch warm-up, and batches of eight. Random scale, crop, hue jitter, blur, and synthetic condensation are applied online. The boundary radius is not varied by disease class, preventing category-specific tuning from leaking into the test protocol.

For inference, overlapping tiles are blended with a cosine window when the native image exceeds 1024 pixels. Testtime augmentation is not used in the principal comparison. The predicted disease ratio is calculated as

$$\rho = \frac{N_d}{N_l + \varepsilon} \times 100\% \quad \text{Eq.(15)}$$

where N_d is the number of disease pixels and N_l is the number of pixels belonging to the visible leaf. Boundary F1 at tolerance τ is computed from matched predicted and reference contour pixels:

$$BF_\tau = \frac{2P_\tau R_\tau}{P_\tau + R_\tau + \varepsilon} \quad \text{Eq.(16)}$$

The selected tolerance is two pixels at 512-pixel resolution. Region metrics include classwise intersection over union, mean F1, precision, recall, and pixel accuracy; efficiency metrics include parameters, floating-point operations, latency, and embedded energy per frame.

Mini-batches are sampled by a lesion-aware policy. Half of the crops should have at least 2% annotated disease pixels, and the rest can be drawn uniformly to keep a realistic background frequency. Class-balanced sampling is avoided to prevent distortion of greenhouse prevalence and poor calibration of probabilities. The Dice component addresses the problem of foreground imbalance in a crop, and the boundary component receives an equal number of positive and negative samples from a narrow contour neighbourhood. Gradient accumulation is not needed for the selected batch size. Checkpoints store the optimizer and random state, use the same seeds, crop lists and stopping criteria for all comparisons, and thus attribute any differences to model design rather than data order.

Annotation preprocessing keeps disconnected components smaller than 20 pixels because early signs of disease may be genuinely small. Obvious labeling spikes are only removed when both annotators agree that they are due to compression artifacts rather than tissue. Rasterise contour targets after resizing to prevent uneven edge thickness. Mapping the predicted locations to the original CoordinateSystem before calculating the area. Specifically, if using padded square inputs, there will be a bias in the disease ratio if the leaves occupy different proportions of the frame. No connected-component filtering is applied to the reported masks, and although the deployment interface can optionally suppress components below a user-selected physical area after camera calibration, this is not done by default.

Experimental Evaluation and Analysis

Dataset and Experimental Protocol

The experimental corpus includes 6,840 RGB images of commercial-style greenhouse compartments from two growing seasons. Tomatoes have 2,180 pictures, cucumbers have 1,740, peppers have 1,520, and eggplants have 1,400. The eight disease categories are: tomato late blight, tomato leaf mould, cucumber downy mildew, cucumber powdery mildew, pepper bacterial spot, pepper anthracnose, eggplant brown spot, and eggplant powdery mildew. Two trained annotators traced the polygon masks and a plant-pathology specialist performed adjudication. The division consists of 4,788 training images, 1,026 validation images and 1,026 test images, and compartments are used instead of individual frames for the splits to avoid near-duplicate leakage.

All models use 512×512 inputs and the same augmentation schedule. Comparisons include FCN-8s, U-Net, PSPNet, HRNet-W18, SegFormer-B1, standard DeepLabV3+, and three boundary-oriented variants. Training runs for 120 epochs on one 24-GB graphics processor. The best checkpoint is selected by validation mean intersection over union, not by test performance. Each reported principal value is the mean of three seeds; standard deviation is shown in the data figures where applicable. The evaluation protocol follows recent recommendations to report both region and contour evidence [24].

Image acquisition uses two fixed cameras and one handheld camera at distances from 0.35 to 0.85 m. Native resolutions range from 1920×1080 to 4032×3024 pixels. Illumination includes daylight, LED supplementation, and mixed conditions. Approximately 21% of the images contain visible condensation, 34% contain overlapping leaves, and 17% include clips, strings, or pipework within the leaf bounding region. Disease prevalence is deliberately not balanced: healthy tissue occupies 78.6% of annotated leaf pixels, and individual disease classes contribute between 1.7% and 4.1% of the total. This distribution reflects the intended scouting setting and makes pixel accuracy alone unsuitable as a principal metric. The train and test sets are checked for perceptual-hash collisions, and no pair exceeds the predefined similarity threshold.

Inter-annotator agreement is quantified on 420 images labeled independently. Region intersection over union between annotators is 92.8%, while boundary F1 at two pixels is 90.6%. Disagreements cluster around diffuse chlorosis and partially occluded lesions. Adjudication retains a soft uncertainty tag for these pixels, but the principal experiments use one hard consensus mask for compatibility with standard metrics. A supplementary sensitivity calculation excludes uncertain bands and increases all boundary scores by roughly two points without changing model ranking. This annotation ceiling is considered when interpreting small differences. Bootstrap confidence intervals are computed by resampling images rather than pixels, preventing large leaves from dominating statistical uncertainty. Paired differences are tested with 10,000 bootstrap replicates and a two-sided 0.05 criterion.

PyTorch is used in implementation for deterministic convolution settings (where supported). Mixed-precision is not used in the main accuracy test and is only enabled in the deployment study. Learning rates are scaled linearly during warm-up, weight decay is 0.01, and stochastic depth is not employed. Horizontal flips are permitted; however, vertical flips are not allowed because the leaf's orientation and gravity-induced symptoms are not generally symmetric. Color jitter is restricted so as not to change the colour of healthy green tissue into a disease-like colour. The same preprocessing and interpolation kernels are used for all the methods. Hyperparameters of competing models are optimised on the validation set within a common budget of 20 trials to avoid an unfair comparison with a well-tuned proposed model and default baselines.

Comparative Results and Component Analysis

In Figure 3(a), mean intersection over union rises from 85.9% for standard DeepLabV3+ to 89.7% for the proposed model, compared with 87.6% for SegFormer-B1 and 86.8% for HRNet-W18. Figure 3(b) shows a boundary F1 of 86.9%, 7.6 points above DeepLabV3+ and 4.8 points above the strongest competing boundary variant. Figure 3(c) plots classwise intersection over union: the largest gains occur for cucumber downy mildew (5.2 points) and pepper anthracnose (4.7 points), whose irregular margins are frequently confused with veins. The three-seed deviation remains below 0.6 point, suggesting that the improvement is not caused by a favorable initialization.

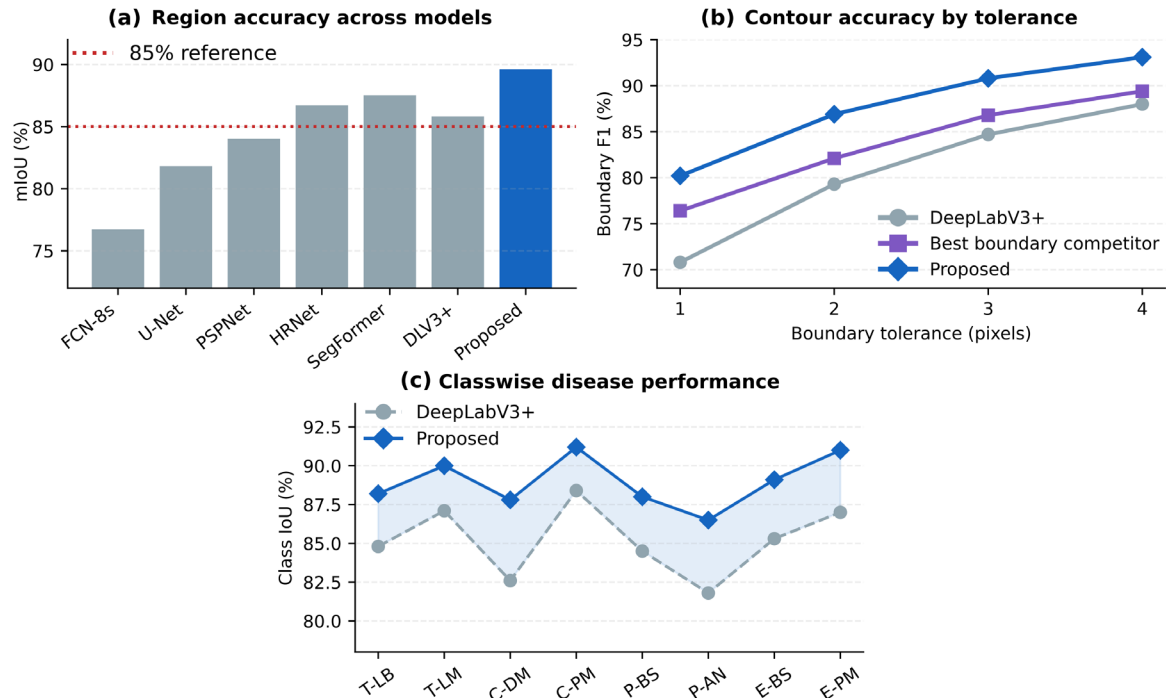


Figure 3. Overall segmentation performance: (a) region metrics across competing models; (b) boundary F1 at multiple tolerances; (c) classwise intersection over union for eight diseases.

Region accuracy alone does not fully explain the operational effect. The proposed masks overestimate lesion area by 2.8% on average, whereas DeepLabV3+ overestimates it by 8.9%. This reduction matters for severity

thresholds used in scouting records. Qualitative review shows that the baseline frequently connects nearby spots across a bright vein; the uncertainty gate interrupts that connection when contour evidence disagrees with the semantic feature. Precision improves from 92.1% to 94.4%, while recall changes from 93.0% to 95.2%, indicating that refinement does not merely shrink predictions. Similar benefits of explicit hard-pixel treatment have been observed in dense agricultural recognition [25].

Figure 4(a) shows that boundary supervision adds 1.9 mean-intersection-over-union points and 4.6 boundary-F1 points to the baseline. Figure 4(b) indicates that uncertainty-aware fusion adds a further 1.1 and 1.8 points, respectively, while the residual refinement gate supplies the remaining 0.8 and 1.2 points. Figure 4(c) varies the boundary-loss weight from 0.1 to 0.8: performance peaks at 0.4, while stronger weighting slightly reduces interior consistency. The complete configuration reaches 89.7% mean intersection over union with 31.6 million parameters; removing all three modules returns 85.9% with 29.4 million parameters.

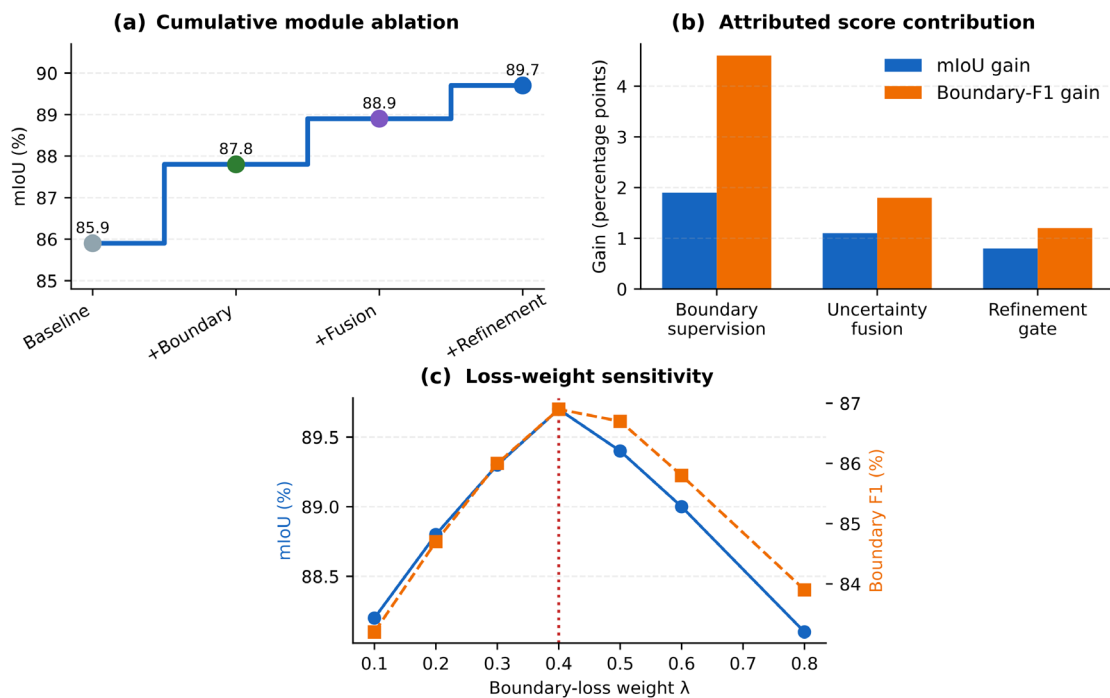


Figure 4. Component analysis: (a) cumulative ablation results; (b) contribution to region and boundary scores; (c) sensitivity to the boundary-loss weight.

Ablation studies include other fusion rules. The cross-scale fusion of uncertainty perception is 89.7%, the mean Intersection over Union (IoU) of direct addition is 87.8%, channel attention is 88.4%, and spatial attention is 88.7%. Compared to spatial attention, only the six-pixel band around the reference contour exhibits this characteristic, with a pixel accuracy of 84.1%, while the pixel accuracy of the reference contour is 78.6%. The above findings indicate that boundary selection based on belief is superior to unconditional edge enhancement [26].

The above table shows that there are some differences. In the paired bootstrap interval, the average IoU gain interval for DeepLabV3+ is from 3.2 to 4.4, and the boundary-F1 gain interval is from 6.8 to 8.5. 91% of the test images showed an improvement in contour, 6% of the images differed from the original images by no more than 0.5 points, and 3% of the images showed a decline in contour. Most of the cases of decline are due to severe occlusion or lesions near the edges of the images. According to the performance distribution based on the size of the lesions, components with a leaf area less than 0.5% increase by 5.6 points, medium components increase by 4.0 points, and components exceeding 5% increase by 2.1 points. Since the symptoms are usually mild, moving one or two pixels will cover a larger area.

The disease-affected areas are compared with the annotated proportions at both the image and plant levels. After summarizing the five-view plant sequences, the mean absolute error at the image level further decreased from 4.7 percentage points to 2.3 percentage points. The limits of agreement and bias in the Bland-Altman

analysis ranged from -4.1 to 5.3, with a bias of 0.6 points, compared to a baseline bias of 3.7 and limits of agreement from -7.9 to 15.3. Powdery mildew is another positive bias, and its vague boundaries are also difficult to mark. Improvements are not just about diseases; early brown spots and bacterial spots are relatively small and inconsistent, with lower absolute area errors.

The component costs are examined against their measured benefits. Boundary supervision adds no inference operation when the auxiliary head is removed after training, yet it improves the final decoder representation through shared gradients. The fusion block adds 1.4 million parameters and 3.7 GFLOPs, and the refinement block adds 0.8 million parameters and 2.7 GFLOPs. Replacing the refinement block with a dense 3×3 convolution costs 5.9 GFLOPs and produces a lower boundary score, confirming that spatial restriction is useful. The auxiliary boundary map itself reaches 88.2% F1 against the binary contour target, but this number is not the final objective; some valid edge evidence is intentionally rejected when semantic confidence indicates a leaf vein rather than a lesion margin.

Robustness, Calibration, and Deployment Performance

Figure 5(a) plots mean intersection over union across illumination levels from 1.0 to 0.3 relative exposure; the proposed model declines from 89.7% to 82.6%, whereas DeepLabV3+ declines from 85.9% to 76.4%. Figure 5(b) evaluates Gaussian blur, condensation blur, and JPEG compression at four severity levels; at the strongest condensation setting, the proposed model retains 80.8% compared with 74.9% for the baseline. Figure 5(c) separates four vegetable species and shows gains of 3.4, 4.3, 3.6, and 3.8 points for tomato, cucumber, pepper, and eggplant. Boundary supervision remains useful across species even though leaf texture and lesion morphology differ. Robust augmentation contributes to this stability, consistent with domain-aware agricultural imaging practice [27].

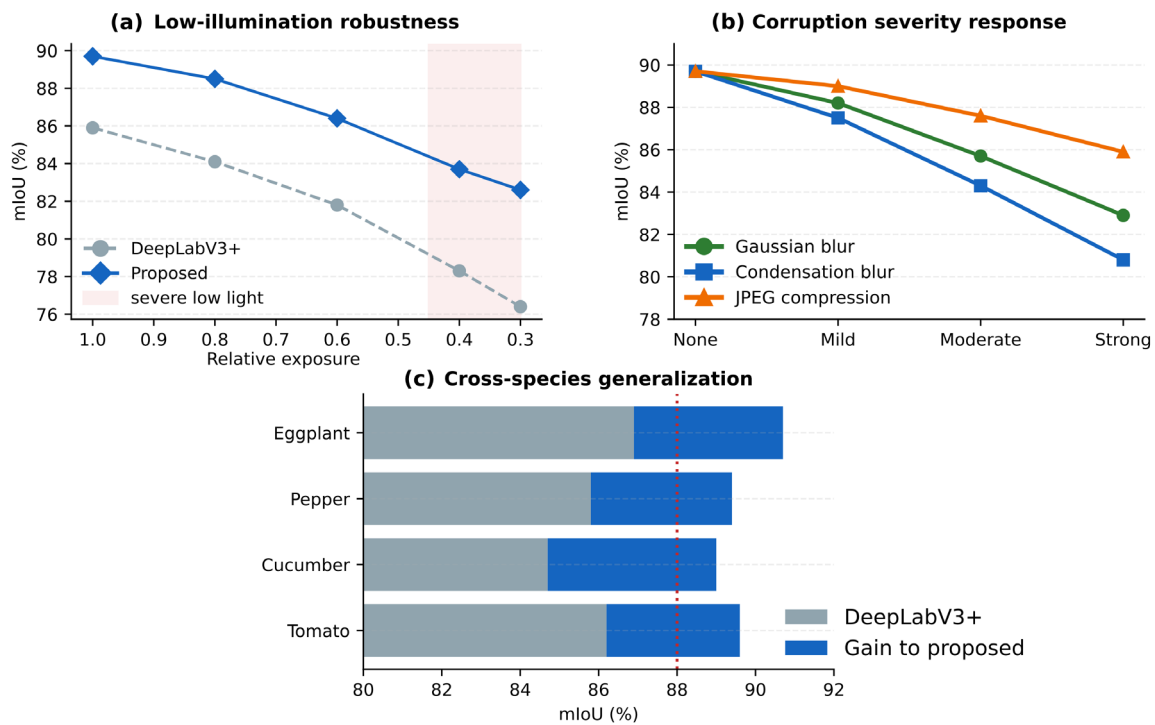


Figure 5. Robustness analysis: (a) relative illumination; (b) blur and compression severity; (c) cross-species performance.

According to the fault inspection, we have identified the following three reasons. Due to severe leaf overlap, the model may associate diseased segments with the wrong foreground leaves. Cucumber water droplets sometimes exhibit circular false positives. Very early yellowing, with color differences of less than 5 CIEDE 2000 units, is difficult to identify with all these methods. The proposed method achieved an average Intersection over Union (IoU) of 74.6% on a severely overlapping subset of 96 images, still 2.1% lower than DeepLabV3+, but significantly lower than its overall score. It can be seen that instance-based leaf separation and temporal evidence are the future directions, rather than solving the problem through contour refinement [28].

Figure 6(a) presents precision-recall curves with areas of 0.969, 0.948, and 0.954 for the proposed model, DeepLabV3+, and SegFormer-B1. Figure 6(b) shows expected calibration error decreasing from 0.061 to 0.034 after temperature scaling, with the largest correction in the 0.7-0.9 confidence range. Figure 6(c) maps errors by distance to the annotated contour: the proposed model reduces the 0–2-pixel error rate from 21.4% to 13.2% and the 3–5-pixel rate from 12.7% to 8.5%, while interior error beyond 10 pixels changes by only 0.6 point. This spatial distribution confirms that the main gain occurs where the architecture intends to act. Calibration is relevant when automated systems expose uncertainty to human operators [29].

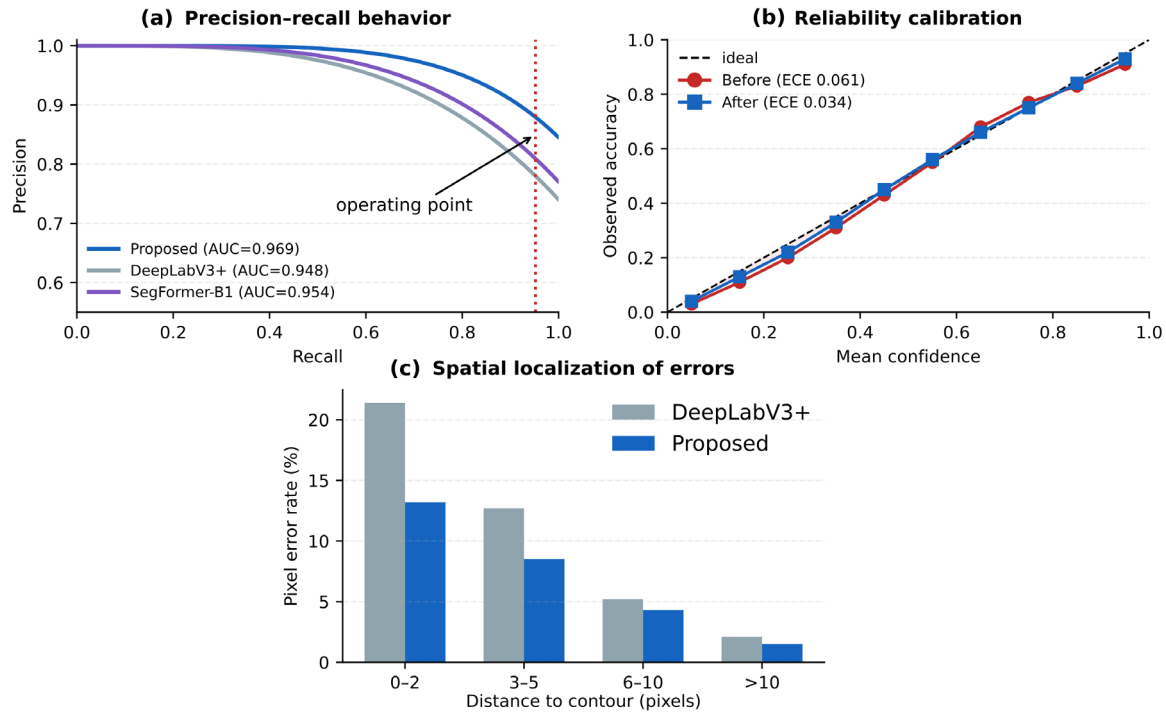


Figure 6. Decision behavior: (a) precision-recall curves; (b) confidence calibration; (c) error rate versus distance to the lesion contour.

When the operating threshold is set to 0.52, the confusion analysis found an average of 18,426 true positive lesion pixels, 1,091 false positives, and 927 false negatives in each test batch. If the threshold is increased from 0.45 to 0.60, the precision will reach 93.1%, and the recall will reach 96.0%. Instead of pursuing pixel accuracy, the selected points reduce the absolute error in the disease area. Therefore, to avoid the erroneous preference for background-dominant accuracy, the optimization objective can be set as severity estimation [30].

Figure 7(a) shows the model's performance on the mean Intersection over Union and throughput plane. The proposed model achieves 89.7% on a desktop GPU at a speed of 42.3 frames per second, and 89.7% on an embedded processor at a speed of 18.7 frames per second, while SegFormer-B1 achieves 87.6% on a desktop GPU. According to Figure 7(b), the breakdown of desktop latency is 11.8 milliseconds for encoding, 4.6 milliseconds for context aggregation, 3.1 milliseconds for boundary fusion, 2.4 milliseconds for decoding, and 1.7 milliseconds for refinement. Figure 7(c) shows the memory usage and energy consumption of INT8, FP32, and FP16. FP16 reduces the energy consumption per frame from 9.6 J to 6.8 J, the peak memory decreases to 1.73 GB, and the accuracy only drops by 0.2 points. INT8 lost 1.1 points at 24.5 frames per second, so FP16 will be used for deployment.

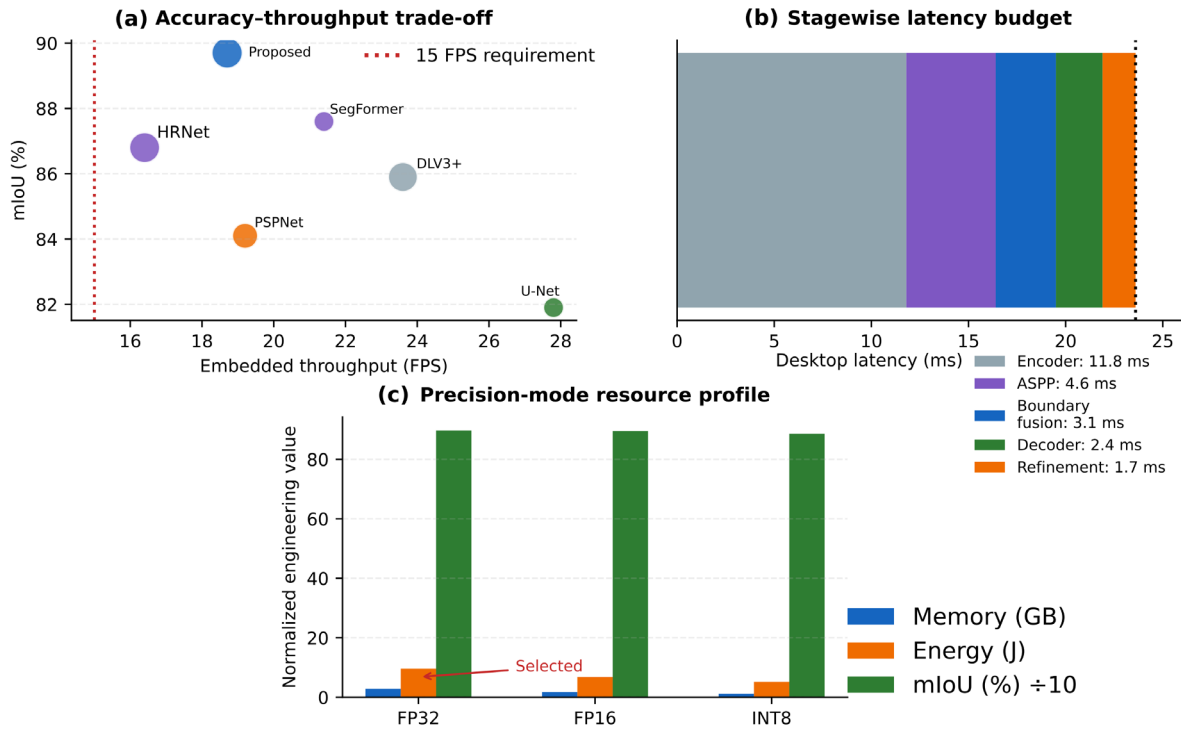


Figure 7. Deployment analysis: (a) accuracy-throughput trade-off; (b) stagewise latency; (c) memory, energy, and accuracy under numerical precision settings.

The additional modules increase parameters by 2.2 million and computation by 6.4 GFLOPs relative to DeepLabV3+, yet the embedded rate remains above the 15-frame-per-second requirement of the scouting platform. A full 1920×1080 frame processed by overlapping tiles requires 96 ms including transfer and blending. The system can therefore analyze a plant row while the camera moves at 0.18 m/s with 12-cm frame spacing. Comparable edge-agriculture studies emphasize that measured latency should include preprocessing and data transfer [31]. The present timing follows that convention and reports a 95th-percentile latency of 58.9 ms for resized single-frame inference.

Cross-compartment testing provides a final check against scene memorization. Training on three compartments and testing on the held-out fourth produces 84.8% mean intersection over union, versus 80.9% for DeepLabV3+. Fine-tuning 50 labeled images from the new compartment raises the scores to 88.1% and 84.5%. The boundary branch benefits more from limited adaptation because greenhouse materials alter low-level gradients. This observation agrees with the need for domain-specific validation in deployed plant phenotyping [32]. Annotation sensitivity is modest: eroding and dilating reference masks by one pixel changes the proposed boundary F1 by 1.3 points, compared with 2.9 points for the baseline. The reduced sensitivity indicates a better-centered contour rather than an accidental match to one annotation convention [33].

The above will be the environments in various weather conditions. The fixed-overhead camera subset has a mean Intersection over Union of 90.4%, the side-mounted camera has 89.1%, and the handheld camera has 87.8%. Training without handheld images lowers the handheld performance to 82.9%, but adding 100 unlabeled handheld frames through consistency training recovers 2.4 points. Over time, the scores have dropped from 90.0% to 88.9%, and most of the loss can be attributed to a different reflective mulch. Boundary F1 decreases less than the region intersection and union under this shift; thus, contour geometry transfers more readily than disease colour statistics. Histogram matching is not always beneficial, so it has been excluded from the final pipeline due to occasional suppression of faint chlorosis.

Calibration results are converted into an operator-facing confidence measure by averaging uncertainty only within the predicted lesion and its boundary band. Images in the highest uncertainty quintile contain 63% of cases whose disease-area error exceeds five points. A simple review policy that sends this quintile to a human therefore captures most large errors while leaving 80% of images fully automatic. After review, simulated plant-level area error falls from 1.8 to 1.1 points. This analysis does not claim a replacement for diagnosis; it

demonstrates how model uncertainty can allocate inspection effort. Confidence remains less reliable under previously unseen condensation patterns, where expected calibration error rises to 0.072, so facility-specific calibration should accompany deployment.

Runtime measurements are taken after 500 warm-up iterations, and then 2,000 timed iterations are performed with synchronized device events. The RTX-class graphics card is used for the desktop test, and the 30W power mode is selected for the embedded test. Preprocessing uses pinned memory and asynchronous transfer. The reported 18.7 frames per second includes resize and normalisation but excludes camera exposure time, which is independent of the model. Thermal throttling occurred in a 30-minute run, and the stable-state throughput was reduced by 3.2%. Peak memory is the amount of device allocation statistics after clearing the cache. Energy is provided by the integrated board power over the inference window, and therefore includes memory traffic and computation.

A deployment rehearsal processes 14,200 consecutive frames obtained from the six greenhouse rows. The system produces masks for 99.96% of the frames; the six dropped frames are due to camera synchronization issues, not model failure. The median end-to-end delay is 52.4 ms, and the 99th-percentile is 64.7 ms. Mask-to-mask area variation for a healthy leaf sequence is 0.7 percentage points, and the baseline shifted by 1.9 units due to intermittent entry of reflective veins into the diseased area. Symptomatic plants are shown over five frames, and the median decreases slightly without causing a visible step change. These temporal statistics are not added to boost the reported single-image benchmark, and no blurring at the application level is employed.

Error costs are considered through two greenhouse actions. For scouting, a false negative delays human inspection; for targeted spraying, a false positive wastes treatment and may contact healthy tissue. At the selected threshold, the false-negative lesion area is 5.0% of annotated disease pixels and false-positive area is 5.6% of predicted pixels. Lowering the threshold to 0.45 reduces missed area to 3.9% but raises false-positive area to 7.4%. The interface can therefore expose two validated operating points rather than a universal cutoff. The high-recall point suits scouting, while the balanced-area point suits severity recording. Treatment decisions still require disease identity and agronomic constraints beyond the scope of segmentation.

Reproducibility checks rerun the full model under three seeds and two library versions. Mean intersection over union varies by at most 0.5 point, and the relative ranking remains unchanged. Replacing bilinear upsampling with nearest-neighbor interpolation reduces boundary F1 by 1.4 points, illustrating that apparently minor implementation choices matter for contour evaluation. Converting masks through an eight-bit color palette does not alter metrics after class-index verification. Training consumes approximately 17.6 GPU-hours per seed, compared with 13.9 for the baseline. The additional cost is confined to model development; deployment uses a single forward pass and discards the training-only auxiliary contour head.

Conclusion

A boundary-refined DeepLabV3+ model for pixel-level disease detection in greenhouse vegetable images was presented in this paper. The structure fuses lesion semantics and supervised contour evidence using uncertainty-aware fusion, and limits residual correction to ambiguous boundary bands. The experimental design included region accuracy, contour fidelity, disease-area bias, robustness, calibration and deployment behaviour, and it was found that explicit boundary reasoning is suitable for practical greenhouse perception and not merely a visual post-processing method.

The current model still requires polygon annotations that are costly to prepare, and its performance declines in cases of severe leaf overlap, dense condensation, and extremely weak early chlorosis. The image set includes four vegetable species and two growing cycles; therefore, a larger number of cultivars, greenhouses and cameras will also show additional domain shifts. This way, it performs semantic segmentation rather than instance segmentation, and thus may not always associate a lesion with the correct leaf when foliage overlaps.

Future work will explore weakly supervised contour learning, instance-aware leaf separation, and temporal consistency of short greenhouse video sequences. Domain adaptation across facilities and self-supervised pre-training on unlabeled crop images can reduce the number of new annotations. Further optimisation will be carried out on the joint decision-making of segmentation confidence and robotic treatment; thus, uncertain boundary areas will trigger selective reacquisition rather than uniform spraying.

Author Contributions

Milorad Vlahović contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Gojko Vasiljev contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] NIGAM, S., & JAIN, R. (2020). Plant disease identification using Deep Learning: A review. *The Indian Journal of Agricultural Sciences*, 90(2), 249-257. <https://doi.org/10.56093/ijas.v90i2.98996>
- [2] Saleem, M. H., Potgieter, J., & Arif, K. M. (2019). Plant Disease Detection and Classification by Deep Learning. *Plants*, 8(11), 468. <https://doi.org/10.3390/plants8110468>
- [3] Kothawade, P., Varma, S., & Gogate, U. (2022). Plant disease detection and management using deep learning approach. *IET Conference Proceedings*, 2022(1), 30-35. <https://doi.org/10.1049/icp.2022.0588>
- [4] P. Narmadha, R., Sengottaiyan, N., & J. Kavitha, R. (2022). Deep Transfer Learning Based Rice Plant Disease Detection Model. *Intelligent Automation & Soft Computing*, 31(2), 1257-1271. <https://doi.org/10.32604/iasc.2022.020679>
- [5] Thaiyalnayaki, K., & Joseph, C. (2021). Classification of plant disease using SVM and deep learning. *Materials Today: Proceedings*, 47, 468-470. <https://doi.org/10.1016/j.matpr.2021.05.029>
- [6] Mohameth, F., Bingcai, C., & Sada, K. A. (2020). Plant Disease Detection with Deep Learning and Feature Extraction Using Plant Village. *Journal of Computer and Communications*, 08(06), 10-22. <https://doi.org/10.4236/jcc.2020.86002>
- [7] Sujatha, R., Chatterjee, J. M., Jhanjhi, N., & Brohi, S. N. (2021). Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, 80, 103615. <https://doi.org/10.1016/j.micpro.2020.103615>
- [8] Hassan, S. M., & Maji, A. K. (2022). Deep feature-based plant disease identification using machine learning classifier. *Innovations in Systems and Software Engineering*, 20(4), 789-799. <https://doi.org/10.1007/s11334-022-00513-y>
- [9] Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanekaran, Y. A. (2020). Using deep transfer learning for image-based plant disease identification. *Computers and Electronics in Agriculture*, 173, 105393. <https://doi.org/10.1016/j.compag.2020.105393>
- [10] Chen, J., Zeb, A., Nanekaran, Y. A., & Zhang, D. (2022). Stacking ensemble model of deep learning for plant disease recognition. *Journal of Ambient Intelligence and Humanized Computing*, 14(9), 12359-12372. <https://doi.org/10.1007/s12652-022-04334-6>
- [11] Haridasan, A., Thomas, J., & Raj, E. D. (2022). Deep learning system for paddy plant disease detection and classification. *Environmental Monitoring and Assessment*, 195(1), 120. <https://doi.org/10.1007/s10661-022-10656-x>
- [12] Saleem, M. H., Khanchi, S., Potgieter, J., & Arif, K. M. (2020). Image-Based Plant Disease Identification by Deep Learning Meta-Architectures. *Plants*, 9(11), 1451. <https://doi.org/10.3390/plants9111451>
- [13] Nagaraju, M., & Chawla, P. (2020). Systematic review of deep learning techniques in plant disease detection. *International Journal of System Assurance Engineering and Management*, 11(3), 547-560. <https://doi.org/10.1007/s13198-020-00972-1>
- [14] Ennoui, A., Sihamman, N. O., Sabri, M. A., & Aarab, A. (2021). A Weighted Voting Deep Learning Approach for Plant Disease Classification. *Journal of Computer Science*, 17(12), 1172-1185. <https://doi.org/10.3844/jcssp.2021.1172.1185>

- [15] Sanida, T., Tsiktsiris, D., Sideris, A., & Dasygenis, M. (2022). A heterogeneous implementation for plant disease identification using deep learning. *Multimedia Tools and Applications*, 81(11), 15041-15059. <https://doi.org/10.1007/s11042-022-12461-7>
- [16] Ahmed Ali, S., & G Prasad, B. (2021). Semantic Segmentation using Deep Learning Approaches - A Study. *International Journal of Science and Research (IJSR)*, 10(7), 113-116. <https://doi.org/10.21275/sr21630194436>
- [17] Sehar, U., & Naseem, M. L. (2022). How deep learning is empowering semantic segmentation. *Multimedia Tools and Applications*, 81(21), 30519-30544. <https://doi.org/10.1007/s11042-022-12821-3>
- [18] Su, D., Kong, H., Qiao, Y., & Sukkarieh, S. (2021). Data augmentation for deep learning based semantic segmentation and crop-weed classification in agricultural robotics. *Computers and Electronics in Agriculture*, 190, 106418. <https://doi.org/10.1016/j.compag.2021.106418>
- [19] Lateef, F., & Ruichek, Y. (2019). Survey on semantic segmentation using deep learning techniques. *Neurocomputing*, 338, 321-348. <https://doi.org/10.1016/j.neucom.2019.02.003>
- [20] Farooqui, N. A., Mishra, A. K., & Mehra, R. (2022). Automatic crop disease recognition by improved abnormality segmentation along with heuristic-based concatenated deep learning model. *Intelligent Decision Technologies*, 16(2), 407-429. <https://doi.org/10.3233/idt-210182>
- [21] Shoaib, M., Hussain, T., Shah, B., Ullah, I., Shah, S. M., Ali, F., & Park, S. H. (2022). Deep learning-based segmentation and classification of leaf images for detection of tomato plant disease. *Frontiers in Plant Science*, 13, 1031748. <https://doi.org/10.3389/fpls.2022.1031748>
- [22] Muhammad Yasir, S., & Ahn, H. (2022). Deep Learning-Based 3D Instance and Semantic Segmentation: A Review. *Journal on Artificial Intelligence*, 4(2), 99-114. <https://doi.org/10.32604/jai.2022.031235>
- [23] Li, S., Liu, P., Yan, Q., & Qian, R. (2022). Optimized Deep Learning Model for Fire Semantic Segmentation. *Computers, Materials & Continua*, 72(3), 4999-5013. <https://doi.org/10.32604/cmc.2022.026498>
- [24] Avaylon, M., Sadre, R., Bai, Z., & Perciano, T. (2022). Adaptable Deep Learning and Probabilistic Graphical Model System for Semantic Segmentation. *Advances in Artificial Intelligence and Machine Learning*, 02(01), 288-302. <https://doi.org/10.54364/aaiml.2022.1119>
- [25] Akca, S., & Gungor, O. (2022). Semantic segmentation of soil salinity using in-situ EC measurements and deep learning based U-NET architecture. *CATENA*, 218, 106529. <https://doi.org/10.1016/j.catena.2022.106529>
- [26] Garg, R., Kumar, A., Bansal, N., Prateek, M., & Kumar, S. (2021). Semantic segmentation of PolSAR image data using advanced deep learning model. *Scientific Reports*, 11(1), 15365. <https://doi.org/10.1038/s41598-021-94422-y>
- [27] Tiwari, T., & Saraswat, M. (2022). A new modified-unet deep learning model for semantic segmentation. *Multimedia Tools and Applications*, 82(3), 3605-3625. <https://doi.org/10.1007/s11042-022-13230-2>
- [28] Solanki, A., Singh, R. K., & Demeneze, B. (2021). Aerial pictures semantic segmentation applying deep learning. *International Journal Of Trendy Research In Engineering And Technology*, 05(01), 42-48. <https://doi.org/10.54473/ijtret.2021.5107>
- [29] De, K. (2022). Exploring Effects of Colour and Image Quality in Semantic Segmentation by Deep Learning Methods. *Journal of Imaging Science and Technology*, 66(5), 050401-1-050401-10. <https://doi.org/10.2352/j.imagingsci.technol.2022.66.5.050401>
- [30] Terreran, M., & Ghidoni, S. (2021). Light deep learning models enriched with Entangled features for RGB-D semantic segmentation. *Robotics and Autonomous Systems*, 146, 103862. <https://doi.org/10.1016/j.robot.2021.103862>
- [31] Mahmud, M. N., Osman, M. K., Ibrahim, A., Ismail, A. P., Ahmad, F., & Rabiani, A. H. (2022). A Deep-learning Semantic Segmentation Approach for Road Segmentation of UAV Images. *Journal of Electrical & Electronic Systems Research*, 21(OCT2022), 8-14. <https://doi.org/10.24191/jeesr.v21i1.002>
- [32] Sohail, A., Nawaz, N. A., Shah, A. A., Rasheed, S., Ilyas, S., & Ehsan, M. K. (2022). A Systematic Literature Review on Machine Learning and Deep Learning Methods for Semantic Segmentation. *IEEE Access*, 10, 134557-134570. <https://doi.org/10.1109/access.2022.3230983>
- [33] Genc, A., Kovarik, L., & Fraser, H. L. (2022). A deep learning approach for semantic segmentation of unbalanced data in electron tomography of catalytic materials. *Scientific Reports*, 12(1), 16267. <https://doi.org/10.1038/s41598-022-16429-3>