Embedded Vision AI for UAV-Based Weed Detection in Sugarcane Field Using Semantic Segmentation

Robson Rogério Dutra Pereira[1], **Sandro Elias Graziosi**[1], **Roberto Santos Inoue**[1]

[1]Federal University of São Carlos, Department of Computing (UFSCar-DC)
Rod. Washington Luís, km 235 – 13565-905 – São Carlos, SP – Brazil

robson.pereira@ifmt.edu.br, rsinoue@ufscar.br

Abstract. *Weed infestation in sugarcane increases herbicide use, costs, and environmental pressure. This paper benchmarks 394 custom dense-prediction Semantic Segmentation variants across nine architectures (U-Net, PSPNet, FPN, SegNet, DeepLabV3+, FCN-ASPP, LinkNet, HRNet, SegFormer) for onboard UAV-based weed detection using high-resolution RGB images from commercial fields annotated into four foreground classes plus background. A rapid workflow was developed combining field validation, aerial imaging, manual polygon annotation, and tiled partitioning (128–224 px). Beyond standard metrics, we introduced lightweight inference-time uncertainty diagnostics to filter miscalibrated models. The pipeline enable practical edge for precision agriculture.*

Keywords: semantic segmentation, edge AI, UAVs, precision agriculture, uncertainty diagnostics, TensorRT.

1. Introduction

Unmanned Aerial Vehicles (UAVs) enable high-resolution sensing for monitoring and decision support in field environments [Mohsan et al. 2022]. In precision agriculture, this capability is valuable because spatially variable management depends on timely maps of crop condition and weed occurrence [Mukherjee et al. 2023, Portz et al. 2023]. For large sugarcane production areas, UAV imagery combined with onboard Artificial Intelligence (AI) can reduce latency, communication dependence, and post-processing bottlenecks [Mukherjee et al. 2019, Koubaa et al. 2023]. Unlike a ground weeding robot, the mapping UAV need not actuate immediately. Its advantage is to convert full-resolution images locally into compact GeoPackage, Shapefile, GeoJSON, or KML layers that can be transferred when connectivity becomes available and consumed by tractors, ground sprayers, or spraying UAVs.

Semantic Segmentation is well suited to this task because dense prediction networks produce a class label for every pixel, supporting weed patch delineation, density estimation, and variable-rate spraying [Shelhamer et al. 2017, Zhao et al. 2017, Milioto et al. 2018]. However, UAV deployment imposes stricter requirements than offline validation: models must satisfy limits on memory, power, temperature, and throughput while remaining reliable on full-resolution scenes [Mohsan et al. 2023]. Tile-level validation alone may hide class collapse, overconfidence, or minority-class blindness during full-image inference.

Model calibration is therefore central to deployment. Neural networks can be accurate but poorly calibrated, meaning softmax confidence does not necessarily match

the probability of correctness; overconfidence is especially common in modern models [Guo et al. 2017]. For agricultural edge inference, simple diagnostics computed from full-image softmax outputs—prediction distribution, mean confidence, and confidence standard deviation—can expose models that obtain high validation scores but fail to process spatial evidence at inference time. This is particularly important for weed mapping, where the objects of interest are minority classes distributed in small, irregular patches. A model that predicts a dominant vegetation class with unrealistic certainty may still appear competitive on tile-level metrics, but it is not useful for site-specific weed management.

Sugarcane is economically important for sugar, ethanol, and bioenergy production [Maldaner et al. 2021b], while weed infestation reduces yield and increases herbicide demand [Moriya et al. 2023, Maldaner et al. 2021a]. Rather than claiming novelty from edge inference alone, this work contributes: (i) a controlled screening of 394 dense-prediction configurations under one rapid-dataset protocol; (ii) a multi-stage selection procedure joining tile accuracy, full-image prediction-distribution checks, confidence diagnostics, and embedded profiling; (iii) evidence that configurations ranked highly by tile mIoU can collapse during full-image inference in this small, spatially dependent dataset; and (iv) an end-to-end sugarcane case study that converts accepted predictions into interoperable vector layers for downstream treatment machinery. The novelty is therefore deployment-oriented model selection and geospatial product generation.

The organization of this paper is structured as follows. Section 2 reviews related work on dense prediction networks for agricultural segmentation and edge AI for UAV perception. Section 3 describes the proposed methodology, including dataset construction, class definition, training protocol, evaluation metrics, uncertainty diagnostics, and deployment pipeline. Section 4 presents the experimental results, covering dataset characterization, model benchmarking, uncertainty filtering, TensorRT inference profiling, and GIS output generation. Section 5 discusses architectural trade-offs, the tile versus full-image discrepancy, and provides practical deployment recommendations. Finally, Section 6 concludes the paper with a summary of findings, limitations, and directions for future work.

2. Related Work

Dense Prediction Networks originated from Fully Convolutional Networks, which replaced dense layers with convolutions for pixel-wise prediction. U-Net introduced encoder–decoder skip connections; PSPNet added pyramid pooling for multi-scale context [Zhao et al. 2017]; FPN exploited top-down hierarchical features [Lin et al. 2017]. Other families include SegNet with pooling indices [Badrinarayanan et al. 2017], DeepLabV3+ and FCN-ASPP with atrous multi-scale context [Chen et al. 2018], LinkNet for fast segmentation [Chaurasia and Culurciello 2017], HRNet for high-resolution representations, and SegFormer for transformer-based dense prediction [Xie et al. 2021].

Dense semantic segmentation has been widely investigated for crop–weed discrimination. Previous studies evaluated segmentation in corn fields [Garibaldi-Márquez et al. 2023, Martins et al. 2024] and identified domain-shift limitations between ground-level and UAV imagery [Gao et al. 2024]. Other works demonstrated onboard UAV inference feasibility [Koubaa et al. 2023], including resource-adaptive architectures [Machidon et al. 2023].

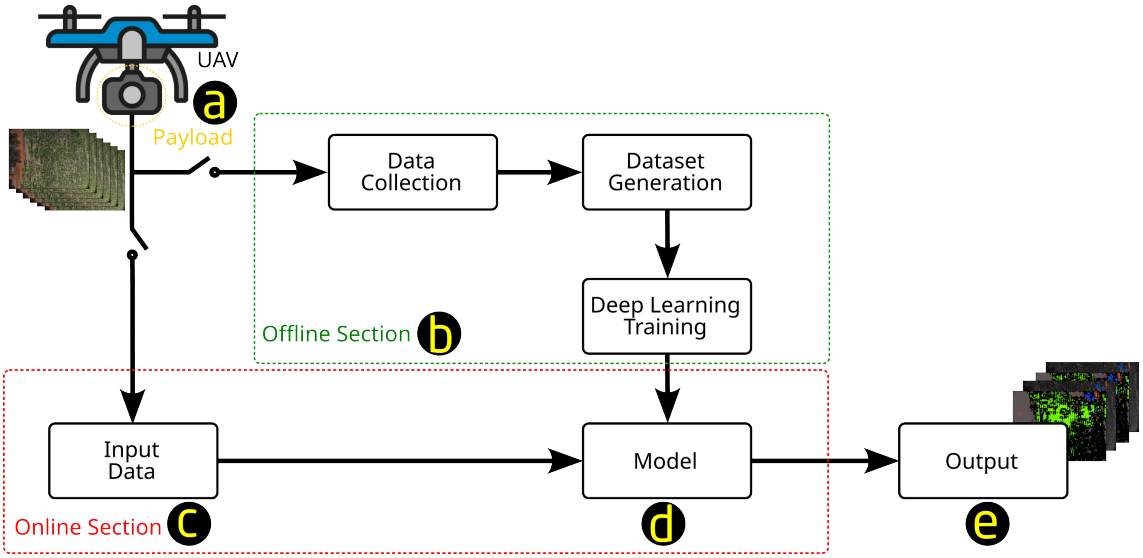


Figure 1. Edge-computing framework: (a) UAV payload; (b) offline data preparation; (c) model validation; (d) embedded deployment; (e) georeferenced output maps.

However, these studies generally focus on segmentation performance, domain adaptation, or onboard execution separately. The present work integrates large-scale architecture screening, full-image reliability assessment, TensorRT profiling, and species-specific GIS vector layer generation within a unified UAV-based sugarcane workflow.

3. Methodology

The proposed framework (Figure 1) separates offline dataset/model construction from online embedded inference. The offline stage includes UAV image acquisition, annotation, tiling, training, and validation. The online stage runs the selected TensorRT model on an edge computer integrated with the UAV and generates georeferenced weed maps for GIS and variable-rate workflows. During field operation, the embedded computer monitors a directory for newly acquired images, performs tiled inference, reconstructs the semantic mask, preserves spatial metadata, and exports map products compatible with downstream decision-support tools. This design avoids cloud dependence and is therefore suitable for connectivity-limited agricultural areas.

High-resolution UAV imagery was collected over a commercial sugarcane field in Brazil using a DJI Mavic 3 Multispectral platform. The survey acquired 270 RGB images with approximately 2 cm Ground Sample Distance (GSD) over 5 ha. Four images (5280×3956 pixels) were selected because together they contained all target classes confirmed by GNSS observations. The remaining 266 images were reserved for operational map generation.

3.1. Class Definition and Dataset Generation

Four foreground classes were defined (Figure 2): sugarcane (*Saccharum officinarum*), *Ricinus communis*, *Amaranthus* spp., and ground. Pixels outside these categories were assigned to background. Manual polygon annotation generated the LBSTDIO dataset. Images and masks were split into non-overlapping tiles of 128×128, 192×192, and 224×224 pixels.

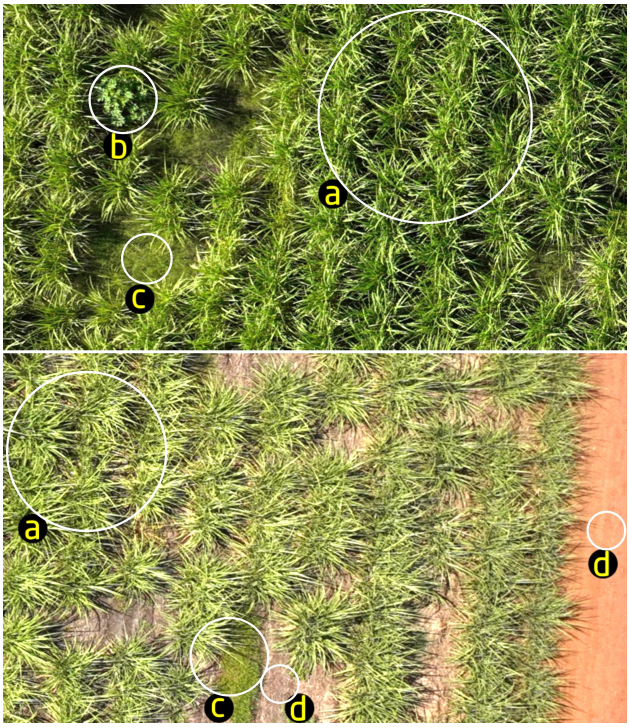


Figure 2. Representative aerial samples: (a) Sugarcane; (b) *Ricinus communis*; (c) *Amaranthus* spp.; (d) Ground.

3.2. Training Protocol and Evaluation Setup

The tiled dataset was split using a 70/20/10 ratio for training, validation, and test sets, respectively, with a fixed random seed of 42 for reproducibility. All metrics reported—mIoU and F1-score—are computed over the four foreground classes only, excluding the background class to avoid inflating scores due to the large background proportion (32.10%). Data augmentation during training included random horizontal and vertical flips, random rotations ($\pm 15^{\circ}$), and brightness/contrast adjustments ($\pm 20\%$). Input images were normalized to zero mean and unit variance per channel using ImageNet statistics. The AdamW optimizer was used with ReduceLROnPlateau scheduling and early stopping.

A systematic TensorFlow training protocol with automatic batch-size fallback was implemented. Training began with batch size 32 and learning rate 0.001; GPU memory constraints triggered progressive reduction down to minimum 1. ReduceLROnPlateau scheduling used factor 0.5 with patience 5, while early stopping monitored validation loss with patience 15 over maximum 100 epochs. A total of 394 model variants were trained across nine architecture families—U-Net, PSPNet, FPN, SegNet, DeepLabV3+, FCN-ASPP, LinkNet, HRNet, and SegFormer—using a modular backbone pool from VGG, ResNet [He et al. 2016], SE-ResNet/SENet [Hu et al. 2018], ResNeXt [Xie et al. 2017], DenseNet [Huang et al. 2017], Inception, MobileNet, and EfficientNet [Tan and Le 2019] families.

Pixel-level evaluation used mIoU and F1-score. To address class imbalance, a Combo Loss combined categorical focal loss and Dice loss:

$$\mathcal{L}_{\text{combo}} = \alpha \mathcal{L}_{\text{focal}} + \beta \mathcal{L}_{\text{dice}}, \quad \alpha = \beta = 0.5. \tag{1}$$

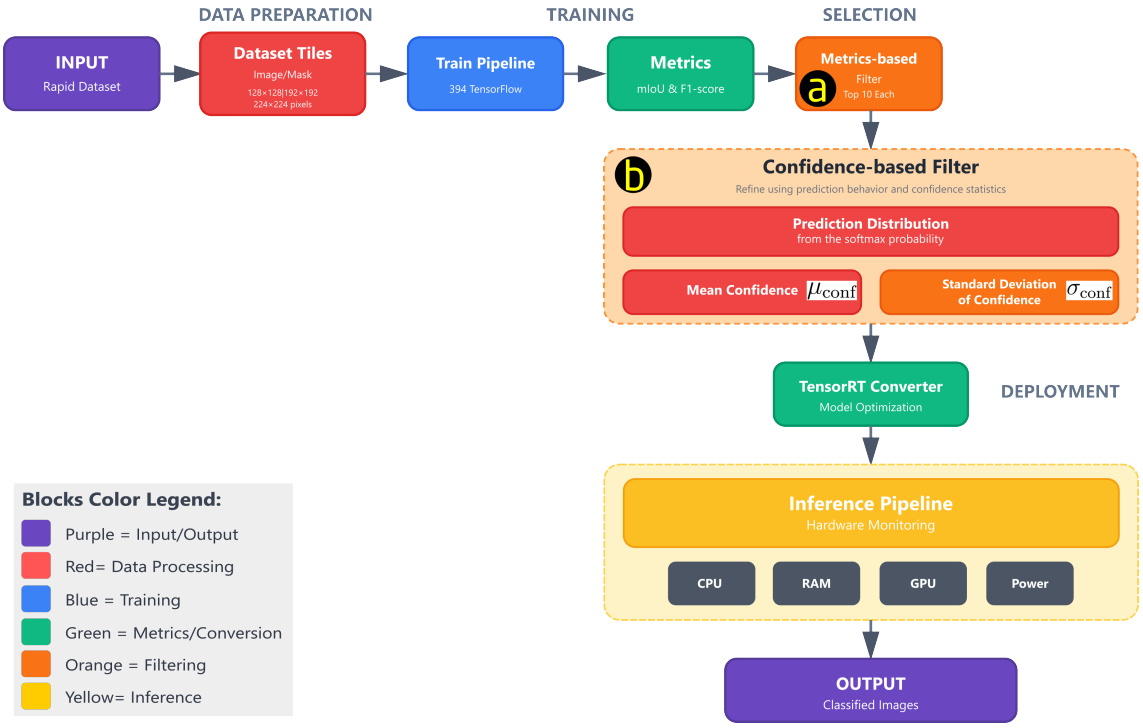


Figure 3. Model protocol pipeline: tiled dataset preparation, TensorFlow training (394 variants), (a) Metric-based Filtering (mIoU/F1-score), (b) Confidence-based Filtering (μ_{conf}, σ_{conf}), TensorRT optimization, hardware-monitored inference, and classified-image output.

The Metric-based Filter in Figure 3 ranked configurations by mIoU and F1-score, retaining the top 10 performers for further screening.

Beyond per-pixel accuracy, three complementary softmax-probability-derived metrics assessed model reliability. Prediction distribution quantifies per-class pixel proportions, revealing bias or class collapse. Mean confidence (μ_{conf}) measures average certainty: values > 0.95 indicate severe overconfidence, while values < 0.60 suggest poor class discrimination. Confidence standard deviation (σ_{conf}) captures the spread of per-pixel maximum probabilities: values < 0.05 indicate uniform, severely miscalibrated predictions; values between 0.10 and 0.25 indicate appropriate uncertainty on boundaries, edges, and shadows; and values > 0.35 suggest unstable predictions.

These diagnostics are not intended to replace formal calibration metrics such as reliability diagrams, Expected Calibration Error, or Brier score. Instead, they provide a low-cost deployment screen for identifying collapsed or implausibly overconfident models during full-image inference. The Confidence-based Filter (Figure 3) applies σ_{conf} as a deployment gate, rejecting implausibly overconfident models regardless of mIoU/F1-score ranking.

Accepted models underwent TensorRT FP16 conversion and profiling on Jetson AGX Orin using 15 unseen images (5280×3956 pixels), recording warm/cold inference time, GPU/CPU/RAM utilization, power draw, and junction temperature. The final full-image benchmark included loading, overlapping tiling, inference, mask reconstruction, and GeoTIFF writing with coordinate metadata preservation.

Table 1. Tile quantities and split distribution.

Tile Size	Total	Training	Validation	Test
128×128	4,920	3,444	984	492
192×192	2,160	1,512	432	216
224×224	1,564	1,512	432	216

Table 2. Pixel distribution for LBSTDIO.

Class	Percentage
Background	32.10%
Sugarcane	60.17%
Ricinus communis	1.01%
Amaranthus spp.	2.58%
Ground	4.15%
Label Density	**67.91%**

4. Results

This section presents the findings of the proposed edge intelligence framework, structured to reflect the sequential stages of model development, validation, and deployment. First, we characterize the LBSTDIO dataset, including tiling strategy, class distribution, and label density. Second, we report the quantitative evaluation of 394 dense prediction network variants, analyzing trade-offs between segmentation accuracy and architectural choices. Third, we present inference-time uncertainty diagnostics used to filter miscalibrated models, followed by system-level benchmarking of deployment-ready models using TensorRT on the NVIDIA Jetson AGX Orin. Finally, we demonstrate practical applicability by showcasing georeferenced weed mapping outputs generated for a commercial sugarcane field.

4.1. Dataset Characterization

Tiles were extracted without overlap at three sizes and split using a 70/20/10 ratio (Table 1). The exploratory analysis in Table 2 reveals severe imbalance: sugarcane accounts for 60.17% of all pixels, whereas *Ricinus communis* and *Amaranthus* spp. account for only 1.01% and 2.58%. Sugarcane therefore contains approximately 59.6 times more pixels than *R. communis* and 23.3 times more than *Amaranthus*. Foreground label density is 67.91%, and the sample count decreases from 4,920 to 1,564 as tile size increases from 128 to 224 pixels. These characteristics motivate background exclusion from macro metrics, the focal–Dice loss, and class-specific reporting.

4.2. Dense Prediction Network Benchmark

Validation distributions were left-skewed: mIoU mean 0.364, median 0.275, std 0.231; F1-score mean 0.401, median 0.340, std 0.190. Table 3 summarizes top configurations. mIoU ranking favored 192×192 tiles, ImageNet initialization, and EfficientNet/ResNeXt backbones. F1-score ranking favored PSPNet or U-Net with ResNet/SE-ResNet backbones, often without ImageNet pre-training.

Table 3. Top-10 configurations by validation mIoU and F1-score (192×192 tiles).

Rank	Arch.	Backbone	Batch	BN	ImageNet	mIoU
Top-10 by Validation mIoU						
1	FPN	EfficientNet-B5	8	False	yes	0.670
2	FPN	EfficientNet-B3	12	False	yes	0.664
3	LinkNet	EfficientNet-B2	20	True	yes	0.662
4	PSPNet	EfficientNet-B7	10	False	yes	0.661
5	FPN	EfficientNet-B4	10	False	yes	0.658
6	PSPNet	EfficientNet-B5	16	False	yes	0.657
7	U-Net	EfficientNet-B4	12	True	yes	0.656
8	PSPNet	ResNeXt-101	20	False	yes	0.654
9	U-Net	EfficientNet-B6	8	True	yes	0.654
10	PSPNet	ResNeXt-50	20	False	yes	0.653
Top-10 by Validation F1-Score						
1	PSPNet	ResNet-18	32	False	no	0.705
2	PSPNet	EfficientNet-B2	32	False	yes	0.703
3	PSPNet	SE-ResNet-18	32	False	no	0.701
4	DeepLabV3+	SE-ResNet-18	32	False	no	0.700
5	U-Net	ResNet-18	32	True	yes	0.697
6	U-Net	ResNet-34	32	True	yes	0.697
7	U-Net	SE-ResNet-34	32	True	no	0.696
8	U-Net	SE-ResNet-18	32	True	yes	0.695
9	PSPNet	SE-ResNet-34	32	False	no	0.693
10	PSPNet	ResNet-34	32	False	no	0.691

4.3. Uncertainty Diagnostics and Edge Inference

The confidence screen changed deployment ranking (Table 4). ResNet, SE-ResNet, and ResNeXt variants retained useful confidence spread ($\sigma_{\text{conf}} = 0.118$–$0.143$) and moderate mean confidence ($\mu_{\text{conf}} = 0.84$–0.94). EfficientNet variants showed near-uniform confidence ($\sigma_{\text{conf}} < 0.03$) and extreme mean confidence ($\mu_{\text{conf}} > 0.997$), indicating class collapse.

After filtering, 11 deployment-ready models were converted to TensorRT FP16 (Table 5). PSPNet with ResNet18 was fastest: 1.08 ± 0.07 s/image, 25.3% GPU, 13.15 W. Compared with TensorFlow FP32 (17.01 W), this represents $2.4\times$ speedup and 22.7% lower power.

4.4. GIS Outputs

PSPNet ResNet18 processed the 266 high-resolution images not used for annotation. After 23 minutes of GIS processing, the system generated an orthomosaic and vector weed layers (Figure 4). Extracted polygons represented 2.2% for *Ricinus communis* and 1.91% for *Amaranthus* spp. over the 5.16 ha field.

5. Discussion

The benchmark demonstrates architecture selection is not secondary. The best-worst configuration difference reached 0.648 mIoU and 0.608 F1-score. Under the rapid, sparse, single-field annotation setting, EfficientNet-based variants ranked highly by validation

Table 4. TensorFlow inference and uncertainty metrics.

Arch.	Backbone	t_{infer} (s)	μ_{conf}	σ_{conf}	Status
DeepLabV3+	seresnet18	4.72 ± 1.48	0.86	0.143	✓ USE
PSPNet	resnet18	2.64 ± 0.08	0.91	0.130	✓ USE
PSPNet	resnet34	2.75 ± 0.09	0.90	0.135	✓ USE
PSPNet	resnext50	4.98 ± 1.41	0.94	0.118	✓ USE
PSPNet	resnext101	4.90 ± 1.31	0.90	0.141	✓ USE
PSPNet	seresnet18	2.74 ± 0.10	0.91	0.135	✓ USE
PSPNet	seresnet34	2.96 ± 0.12	0.89	0.143	✓ USE
U-Net	resnet18	3.71 ± 0.16	0.84	0.140	✓ USE
U-Net	resnet34	3.82 ± 0.16	0.87	0.129	✓ USE
U-Net	seresnet18	3.46 ± 0.12	0.90	0.124	✓ USE
U-Net	seresnet34	4.07 ± 0.17	0.86	0.135	✓ USE
PSPNet	efficientnetb2	3.13 ± 0.11	1.0000	0.0000	× not USE
PSPNet	efficientnetb5	4.50 ± 0.14	0.9997	0.010	× not USE
PSPNet	efficientnetb7	6.00 ± 0.60	0.999	0.013	× not USE
LinkNet	efficientnetb2	5.21 ± 1.21	0.9999	0.0055	× not USE
FPN	efficientnetb3	8.66 ± 0.92	0.9998	0.0073	× not USE
FPN	efficientnetb4	8.41 ± 0.82	0.9999	0.0053	× not USE
FPN	efficientnetb5	9.06 ± 0.80	0.997	0.028	× not USE
U-Net	efficientnetb4	7.45 ± 0.98	0.9999	0.0058	× not USE
U-Net	efficientnetb6	9.28 ± 1.08	0.9998	0.0072	× not USE

Table 5. TensorRT inference benchmark on Jetson AGX Orin.

Rk.	Arch.	Backbone	t_{infer} (s)	GPU (%)	Power (W)	Temp (°C)
1	PSPNet	resnet18	1.08 ± 0.07	25.3	13.15	54.8
2	PSPNet	seresnet18	1.13 ± 0.07	25.8	13.36	55.1
3	PSPNet	resnet34	1.13 ± 0.07	26.5	13.52	54.9
4	PSPNet	seresnet34	1.20 ± 0.05	27.8	13.98	55.1
5	U-Net	resnet18	1.32 ± 0.05	32.8	15.72	55.2
6	DeepLabV3+	seresnet18	1.42 ± 0.09	32.4	15.73	54.5
7	U-Net	seresnet18	1.45 ± 0.04	33.4	16.00	56.0
8	U-Net	resnet34	1.56 ± 0.05	36.7	17.09	55.8
9	U-Net	seresnet34	1.79 ± 0.04	37.1	17.37	56.1
10	PSPNet	resnext50	2.59 ± 0.04	30.7	14.93	55.1
11	PSPNet	resnext101	2.62 ± 0.06	30.3	14.82	55.1

mIoU but exhibited severe full-image confidence collapse. This is not a general EfficientNet limitation but indicates high tile-level validation performance did not imply reliable behavior on complete UAV images.

ResNet and SE-ResNet backbones were less dominant by mIoU but more reliable during full-image inference, maintaining plausible confidence variation. PSPNet was the most consistent architecture family, as pyramid-pooling captured contextual information at different spatial scales. U-Net with ResNet encoders was also robust. FPN was highly dependent on its backbone: EfficientNet-based FPN obtained the highest validation mIoU but failed the full-image reliability gate.

A second issue concerns tile-level validation vs. full-image inference discrepancy.

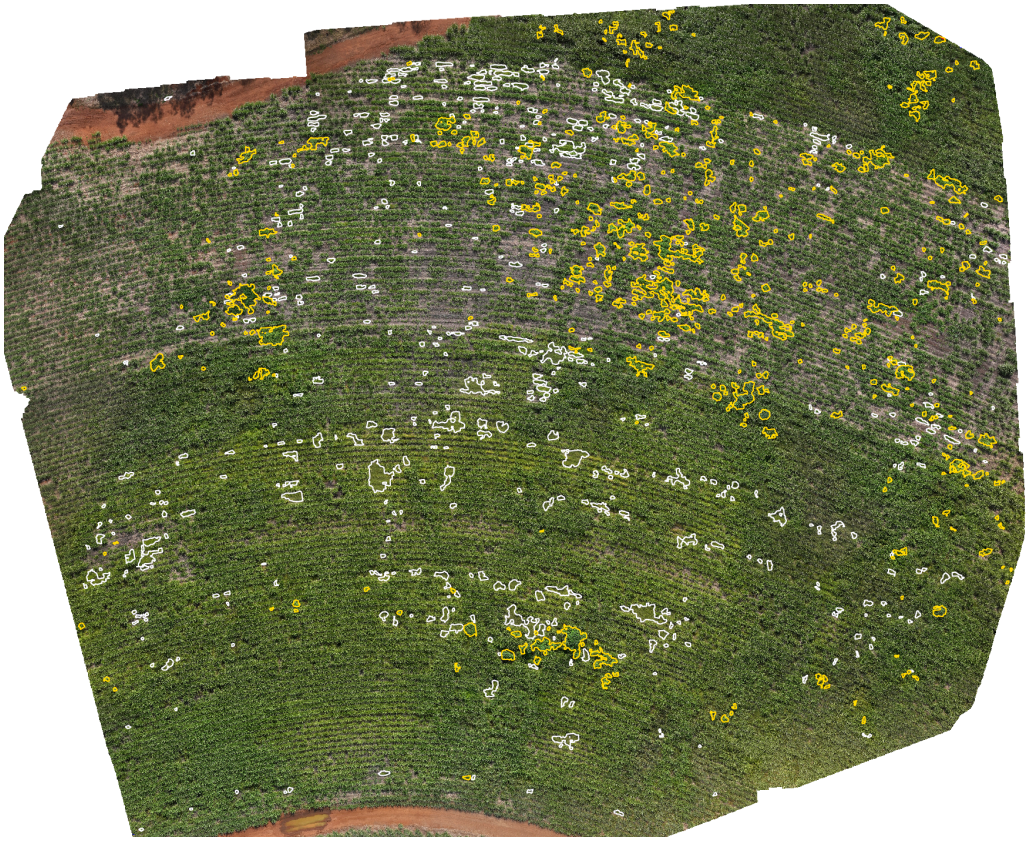


Figure 4. Vector weed layers: *Ricinus communis* (yellow) and *Amaranthus* spp. (white).

The 70/20/10 partition was generated after tiling four source images with a fixed random seed. Validation tiles were not geographically independent—neighboring tiles could share illumination, crop-row orientation, and vegetation structure. Spatial autocorrelation may inflate tile-level metrics. Full-image uncertainty diagnostics provide an additional deployment screen.

Comparison with previous work provides context. Gao et al. [Gao et al. 2024] reported lower UAV IoU for weeds than for maize and soil. Martins et al. [Martins et al. 2024] obtained corn IoU near 0.78, while Garibaldi-Márquez et al. [Garibaldi-Márquez et al. 2023] reported mIoU 0.742. The selected model's F1-score of 0.705 is plausible in this context.

For deployment, PSPNet–ResNet18 is recommended, combining the highest F1-score among reliable candidates (0.705), plausible confidence distribution ($\mu_{\text{conf}} = 0.91$, $\sigma_{\text{conf}} = 0.130$), and lowest TensorRT latency (1.08 ± 0.07 s/image).

5.1. Operational Real-Time Feasibility

The operational feasibility of the selected PSPNet–ResNet18 model was evaluated for a DJI Matrice 350 RTK equipped with a Zenmuse L1 20 mega pixels RGB ($N_x = 5280$ pixels and $N_y = 3956$ pixels) and a Jetson AGX Orin. Under the adopted flight configuration of 70 m altitude (H), 7 m/s speed (v), and 60% forward overlap $O_f = 0.6$, the effective along-track image spacing d_f was calculated as:

$$d_f = \frac{N_y H}{36.5 \times 100}(1 - O_f) = \frac{3956 \times 70}{36.5 \times 100}(0.4) = 27.98 \text{ m} \quad (2)$$

This produced an image-acquisition interval of $t_{\text{photo}} = 27.98/7 = 3.998$ s/image, exceeding the Zenmuse L1 minimum capture interval of 2.5 s for forward overlap 60%.

TensorRT FP16 PSPNet–ResNet18 processed each image in $t_{\text{inf}} = 1.08 \pm 0.07$ s, satisfying $t_{\text{inf}} \leq t_{\text{photo}}$ with a 2.92 s/image margin. The inference throughput of approximately 0.93 images/s was $3.7\times$ greater than the required acquisition rate of 0.25 images/s. The 1.17 s cold-start latency remained below both capture limits.

This establishes inference-level real-time feasibility. However, the complete onboard pipeline must also accommodate image transfer, preprocessing, georeferencing, vectorization, and storage overhead. An asynchronous pipeline could mitigate acquisition delays. Field validation remains essential to determine whether variable image-transfer latency affects the available margin.

6. Conclusion

This paper presented an edge-AI framework for UAV-based sugarcane weed mapping integrating rapid dataset construction, systematic evaluation of 394 dense-prediction configurations, full-image uncertainty screening, TensorRT benchmarking, and GIS-layer generation. Results demonstrate PSPNet with ResNet18 provides the best operational trade-off, achieving F1-score 0.705 and TensorRT FP16 inference time 1.08 ± 0.07 s on Jetson AGX Orin.

The main methodological finding is validation metrics alone are insufficient for agricultural edge deployment. Several EfficientNet variants achieved high tile-level mIoU but failed full-image reliability diagnostics. The proposed diagnostic gate provides a lightweight deployment screen.

Hardware-in-the-loop analysis confirmed PSPNet–ResNet18 satisfies the inference-level real-time constraint. At 70 m altitude, 7 m/s speed, and 60% forward overlap, the sensor captured images every 3.998 s, while the model required only 1.08 ± 0.07 s per image—providing 2.92 s processing margin per acquisition and $3.7\times$ required throughput. The 1.17 s cold-start latency remained below sensor limits.

GIS products confirmed practical relevance: 266 unseen images were converted into georeferenced vector layers identifying *Ricinus communis* and *Amaranthus* spp. across a 5.16 ha commercial sugarcane field.

Future work should validate across fields, seasons, and crop stages; employ geographically separated test areas; investigate formal calibration and out-of-distribution detection; evaluate lower-power Jetson platforms; and assess complete onboard pipelines during operational flights.

References

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.

Chaurasia, A. and Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing*, pages 1–4. IEEE.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 801–818. Springer.

Gao, J., Liao, W., Nuyttens, D., Lootens, P., Alexandersson, E., and Pieters, J. (2024). Cross-domain transfer learning for weed segmentation and mapping in precision farming using ground and UAV images. *Expert Systems with Applications*, 237:122980.

Garibaldi-Márquez, F., Flores, G., and Valentin-Coronado, L. M. (2023). Segmentation and classification networks for corn/weed detection under excessive field variabilities. In *Pattern Recognition*, pages 135–146. Springer Nature Switzerland.

Guo, C., Pleiss, G., Sun, Y., and Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 70:1321–1330.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.

Koubaa, A., Ammar, A., Abdelkader, M., Alhabashi, Y., and Ghouti, L. (2023). AERO: AI-enabled remote sensing observation with onboard edge computing in UAVs. *Remote Sensing*, 15(7):1873.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2117–2125.

Machidon, O. M., Krašovec, A., Machidon, A. L., Pejovic, V., et al. (2023). AgriAdapt: Towards resource-efficient UAV weed detection using adaptable deep learning. In *Proceedings of the 1st International Workshop on Advances in Environmental Sensing Systems for Smart Cities*.

Maldaner, L. F., Molin, J. P., Canata, T. F., and Martello, M. (2021a). Monitoring weed infestation in sugarcane fields using uav-based multispectral imagery. *Biosystems Engineering*, 212:45–57.

Maldaner, L. F., Molin, J. P., Canata, T. F., and Martello, M. (2021b). Sugarcane yield mapping using high-resolution imagery and machine learning. *Precision Agriculture*, 22:1671–1691.

Martins, J. A. C., Higuti, A. Y. H., Pellegrin, A. O., et al. (2024). Assessment of UAV-based deep learning for corn crop analysis in midwest brazil. *Agriculture*, 14(11):2029.

Milioto, A., Lottes, P., and Stachniss, C. (2018). Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2229–2235. IEEE.

Mohsan, S. A. H., Khan, M. A., Noor, F., Ullah, I., and Alsharif, M. H. (2023). Towards the unmanned aerial vehicles (uavs): A comprehensive review. *Drones*, 7(6):391.

Mohsan, S. A. H., Othman, N. Q. H., Li, Y., Alsharif, M. H., and Khan, M. A. (2022). Unmanned aerial vehicles (uavs): Practical aspects, applications, open challenges, security issues, and future trends. *Intelligent Service Robotics*, 16:109–137.

Moriya, É. A. S., Imai, N. N., Tommaselli, A. M. G., Berveglieri, A., and Honkavaara, E. (2023). Detection of weeds in sugarcane using uav-based hyperspectral imaging and deep learning. *Remote Sensing*, 15(9):2322.

Mukherjee, A., Misra, S., and Raghuwanshi, N. S. (2019). A survey of unmanned aerial sensing solutions in precision agriculture. *Journal of Network and Computer Applications*, 148:102461.

Mukherjee, A., Misra, S., Raghuwanshi, N. S., and Mitra, S. (2023). A comprehensive survey of uav-based precision agriculture: Current status and future perspectives. *IEEE Access*, 11:52626–52662.

Portz, G., Molin, J. P., and Jasper, J. (2023). Precision agriculture in sugarcane: A review on sensing technologies for crop and soil monitoring. *Precision Agriculture*, 24:1150–1181.

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully convolutional networks for semantic segmentation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 39, pages 640–651.

Tan, M. and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 6105–6114.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 12077–12090.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2881–2890.