On-Edge Weed Detection Using Unmanned Aerial Vehicles

Mohamed Elmohandes[1], **Diaa Addeen Abuhani**[1], **Maya Haj Hussain**[1], **Jowaria Khan**[1], **Imran Zualkernan**[1]

[1]Computer Science and Engineering Department, American University of Sharjah, Sharjah, United Arab Emirates
b00083108@aus.edu, b00086137@aus.edu, g00085636@aus.edu, g00084343@aus.edu, izualkernan@aus.edu

Abstract

Weed detection is a critical task in precision agriculture, as weeds significantly reduce crop yields and increase production costs. This work presents an efficient, low-cost, and low-power UAV-deployable weed detection pipeline that relies solely on consumer grade RGB cameras. We propose novel lightweight U-Net and Attention U-Net architectures optimized for real-time semantic segmentation on an edge device. To enhance segmentation accuracy, we integrate additional RGB derived features. Experiments on the CoFly-WeedDB detection dataset demonstrate that both architectures perform effectively on RGB imagery, with further improvements when incorporating hue and edge detection features. The proposed lightweight U-Net architectures, made more efficient through quantization, achieves IoU scores above 50% and Dice scores exceeding 60% on the CoFly-WeedDB dataset across multiple augmentation levels. These findings highlight the practicality of deploying lightweight deep learning models for precision weed detection in resource-constrained agricultural environments.

Code — https://github.com/elmohandes2002/SeniorDesign

[1]

1 Introduction

Precision Agriculture (PA) is an agricultural management strategy that aims to enhance agronomic output and reduce resource waste through the integration and utilization of technology and farming management principles in agricultural fields. Image segmentation is used in PA to identify different features or components of a field, such as crops, weeds, soil types, and water content. By segmenting an image, farmers can obtain detailed information about the spatial distribution and characteristics of these features, which can be used to inform decisions about planting, fertilizing, irrigating, and harvesting. There are various methods for image segmentation in PA, including traditional techniques such as thresholding, clustering, and edge detection, as well as more advanced methods such as deep learning-based approaches. Deep learning-based approaches such as convolutional neural networks (CNNs) have shown promising results in crop segmentation tasks, particularly in complex and variable environments. By combining image segmentation with other PA technologies such as sensors, GPS, and drones, farmers can obtain real-time information about their crops and fields, which can be used to make data driven decisions and improve crop yields, efficiency, and sustainability. This agrarian data can then be used by farmers to optimize and tailor their farming practices to their fields' requirements, allowing farmers to apply the right resources exactly where needed and at the time needed (Bongiovanni and Lowenberg-DeBoer 2004). Therefore, PA aids in reducing resource waste and agrochemical usage, increasing crop yield and quality, and reducing human labor.

One of agriculture's most pressing problems today is weed infestation in fields. Due to the severity and impact of the problem, PA has gained attention from agricultural researchers. PA technologies offer cost efficiency by reducing blanket herbicide applications, improve treatment efficacy, and promote environmental sustainability (Bongiovanni and Lowenberg-DeBoer 2004). Notably, recent interest has focused on deep learning approaches using remote sensing data and real-time processing, as these technologies enable large-scale field monitoring and timely decision-making.

Weeds are undesirable, persistent plants that infiltrate fields and hamper the growth of surrounding crops. Weeds mainly cause problems by consuming resources required for keeping crops alive and sustaining their growth. To be precise, weeds compete with crops over water, soil nutrients, space, and even sunlight. Hence, unsurprisingly, when fields are infested with weeds, crop yield loss is noted. In fact, weeds have been the cause behind 35% of global crop yield loss annually, presenting one of the most significant factors causing yield loss (Khan et al. 2021). Therefore, the main goal of our work is to implement such a system, which can be described as a comprehensive deep learning-based UAV deployable system for weed detection and control. The main contributions of this paper are summarized in the following points:

- Implement a cost-effective end-to-end drone-based weed detection system that utilizes consumer grade cameras to achieve real-time weed detection in agricultural fields.
- Develop two lightweight U-Net models.

[1]Accepted at the First International Workshop on AI in Agriculture (Agri AI), co-located with AAAI 2026.

- Analyze UAV imagery to evaluate how image augmentation techniques and RGB-derived features and edge detectors influence overall model performance.

2 Related Works

Recently, there has been a significant focus on the problem of weed detection in agricultural fields, leading to extensive research efforts. Numerous proposals have emerged, presenting complete pipelines to tackle this issue. Among the most notable are those that address the problem by deploying semantic segmentation techniques. Given its ability to accurately classify and localize objects through precise boundary detection, this technique has been the subject of much prior work, specifically in tackling the problem of weed and crop detection, where the objects (weeds and crops) are sometimes indistinguishable and, in many cases, interwoven. This ability makes segmentation models an essential component in effective, comprehensive weed detection systems (Saini and Nagesh 2025).

In cases where multispectral imagery is not available, filters and combinations of RGB channels can be used to perform accurate segmentation of weeds. Several Vegetation Indices (VIs) have been derived to allow accurate estimation of vegetation properties from RGB images (Liang et al. 2022). Various architectures have been proposed for the task of semantic segmentation of crops and weeds in agricultural fields, with the U-Net being the most popular architecture. U-Nets can be easily customized to better fit the problem they are being used for. This adaptability has allowed many researchers working on semantic segmentation in agriculture to tailor their architectures to their specific needs, and hence achieve optimal performance (Saini and Nagesh 2025). MSEA-Net is a recent architecture that currently has the best perfomance on the dataset we are using which is CoFly-WeedDB. U-Net and MESA-Net are encoder–decoder segmentation architectures, with U-Net being the traditional baseline and MESA-Net representing a more recent, high-performing design (Syed et al. 2025).

3 Methodology and Proposed Solution

3.1 Dataset and Preprocessing

The images used in this work are obtained from the publicly available CoFly-WeedDB dataset. The dataset contains 366 RGB images captured through a DJI Phantom 4 drone from a cotton field in Larissa, Greece during the first stage of growth (Krestenitis et al. 2022). All images were captured at a 5 m altitude and 3 m/s speed with a resolution of 1280 × 720 pixels. The original dataset includes three types of weeds namely Johnson grass, Field bindweed, and Purslane. While the dataset provides original segmentation labels, we have found that the masks provided are imprecise as they extend to contain soil and crop areas. Hence, we have manually labeled the dataset using Label Studio software after dropping out empty images resulting in a dataset of 201 images that was split using an 80%, 10%, 10%, training, validation, and testing split respectively. Figure 1 shows a sample image, along with the corresponding pixel-wise mask annotation.

(a) (b)

Figure 1: An example from the dataset. (a) Original image captured showing Johnson Grass weeds (b) The corresponding mask annotation.

Vegetation Index	Equation
ExG	$2G - R - B$
ExR	$1.4R - G$
CIVE	$0.441R - 0.881G$ $+0.385B + 18.78745$
Modified CIVE	$-$CIVE
NDI	$(G - R)/(G + R)$

Table 1: RGB-based VIs and their equations.

3.2 Multispectral Bands Extraction

To extract a richer set of features from RGB images, we divided the features into three categories: Hue, Saturation, and Value (HSV) alternate color representation, VIs derived from light spectra, and Edge Detector (ED) features. Table 1 summarizes a selected set of VIs, including the Excess Green (ExG), Excess Red (ExR), Color Index of Vegetation (CIVE) (Liang et al. 2022), and Normalized Difference Index (NDI) (Meyer and Neto 2008). While many VIs exist, these represent a few commonly used indices. The original CIVE formula subtracts the green channel, so we invert the formula by multiplying it by -1 to make the green values positive, as shown in Table 1.

3.3 Augmentation

Image augmentation has been employed to increase the size of the training dataset. Taking into consideration that the images are captured through a UAV, the type of augmentation applied was chosen in a way such that it can reflect an image captured in the same method. Thereby, we decided to augment our images by horizontally flipping the images, randomly cropping certain segments of the image, and finally by applying grid distortion on the image at hand.

3.4 Proposed Architecture

We propose a custom model that follows a U-Net based architecture to perform image segmentation of weeds in the CoFly-WeedDB dataset. The model is tested in two different configurations in which the first is a straightforward U-Net model with five levels of convolutions while the second configuration contains attention blocks as shown in Figure 2. The attention mechanism applied implements a gating layer in which spatial information extracted from the encoder layers are added with the current decoder layers which tend to contain deeper features (Oktay et al. 2018). This way,

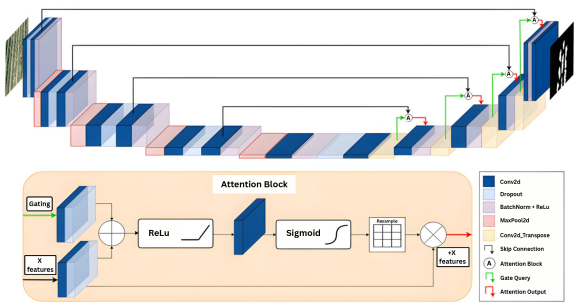

Figure 2: The proposed U-Net based model along with the attention block mechanism.

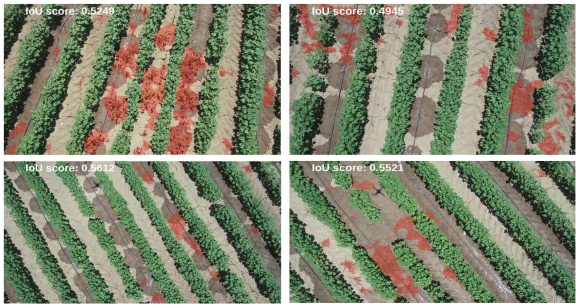

Figure 3: Weed detection results using U-Net + Hue multispectral bands.

both the spatial and feature information are retained, normalized, resampled, and fed to the following layer. The attention block in Figure 2 provides the procedure followed in further detail.

3.5 Optimization and Deployment

The Jetson Nano model chosen for this paper had a 16 GB storage and an NVIDIA Maxwell GPU with 128 NVIDIA CUDA cores. It was equipped with 4 GB 64-bit LPDDR4 memory running at a clock frequency of 1600MHz. Given the clear RAM and disk space limitations of our edge computing device, in addition to the latency constraints of our real-time system, we needed to optimize the selected models before deployment. As a result, we converted the chosen models into two optimized compressed formats using the TensorFlow Lite and Tensor-RT libraries.

3.6 Power Consumption, Drone speed, and Inference time

This section details the feasibility of our model's deployment on a drone in a real-time setting. The camera used in our experiment is an RPi camera module V2. Therefore, our Ground Sample Distance (GSD) calculation is done using this camera's specifications. The GSD is the distance, measured on the ground, between the centers of two consecutive pixels. It is a metric for measuring accuracy in remote sensing and aerial mapping activities (Enterprise 2022). Equation (1) illustrates the GSD formula.

$$\text{GSD} = \frac{\text{Sensor Width} \times \text{Drone Altitude}}{\text{Image Width} \times \text{Camera Focal Length}} \quad (1)$$

Given the RPi module's sensor width and focal length of 2.76mm and 3.04mm respectively, and our image width and drone altitude of 720 pixels and 5m respectively, the GSD calculates to be 0.63 cm/pixel. Therefore, the linear length captured by the drone in a single shot at 5m altitude given our image length of 1280 pixels is 8.06m. An altitude of 5m is estimated here since most farming applications that employ drones also maintain the same resolution. Furthermore, the dataset over which our model is trained is also acquired at a 5m resolution. Therefore, this altitude estimation is chosen to maintain consistency. Ideal drone speed can be calculated using Equation (2).

$$\text{Ideal Speed} = \text{Linear Length (at 5m altitude)} \times \text{FPS} \quad (2)$$

Power consumption was determined after measuring the current drawn by the Nvidia Jetson Nano during model inference. To measure the current, a Yocto-Amp device was connected between the power source and the power jack of the Nvidia Jetson Nano. The measured current values were then saved in a CSV file. These values were iterated over, averaged, and subsequently used to calculate the power consumption during inference.

4 Results and discussion

4.1 Weed Detection Results - Pre-Optimization

Weed detection results were obtained using different number of training samples; specifically, using 160 samples, 320 samples, 480 samples, and 640 samples. The weights were stored using Float-32 decimal points. Forward variable selection was implemented at each augmentation level and the top five performing combinations of multispectral bands were recorded. Figure 3 shows some sample outputs of the weed detection results obtained from using RGB channels along with an additional hue band along with the corresponding IoU score. Table 2 summarizes the results of our custom U-Net, U-Net with attention and the best performing multispectral bands added.

As can be seen in Table 2, U-Net which employed the RGB bands with a single additional multispectral band resulted in the best performing model at three different augmentation levels. Hue spectrum and Laplacian edge detector appeared to be more informative bands than others. In general, adding an additional spectral band appears to be as influential as adding attention blocks to our architecture with a slight difference in performance in favor of the additional band and less computational complexity. A further look into the features extracted shows that while the attention model captures more shape features on the decoder end of the U-Net, the addition of a multispectral band adds more spatial information on the encoder end of the U-Net. We argue that both methods boost the U-Net performance in two different ways leading to almost equal enhancement in the overall performance. It is worth mentioning that combining multispectral bands with attention did not result in an improvement in the performance of our models. We believe that this

Bands Used	# of Images	Model	IoU F32	Dice Coefficient F32
RGB (3)	160	U-Net	0.2436	0.3696
RGB (3)	160	U-Net + Attention	0.4707	0.6226
RGB + H (4)	**160**	**U-Net**	**0.5066**	**0.6605**
RGB (3)	320	U-Net	0.4475	0.6039
RGB (3)	320	U-Net + Attention	0.4705	0.6261
RGB + LAP (4)	**320**	**U-Net**	**0.4992**	**0.6522**
RGB (3)	**480**	**U-Net**	**0.5192**	**0.6721**
RGB (3)	480	U-Net + Attention	0.5181	0.6714
RGB + LAP (4)	480	U-Net	0.5018	0.6544
RGB (3)	640	U-Net	0.4740	0.6271
RGB (3)	640	U-Net + Attention	0.4926	0.6473
RGB + H (4)	**640**	**U-Net**	**0.5212**	**0.6736**

Table 2: Results obtained using different multispectral bands.

Bands Used	# of Images	Model	IoU F16	Dice Coefficient F16
RGB (3)	160	U-Net	0.2437	0.3698
RGB (3)	160	U-Net + Attention	0.4704	0.6223
RGB + H (4)	**160**	**U-Net**	**0.5069**	**0.6608**
RGB (3)	320	U-Net	0.4475	0.6038
RGB (3)	320	U-Net + Attention	0.4709	0.6265
RGB + LAP (4)	**320**	**U-Net**	**0.4994**	**0.6523**
RGB (3)	**480**	**U-Net**	**0.5192**	**0.6721**
RGB (3)	480	U-Net + Attention	0.5182	0.6714
RGB + LAP (4)	480	U-Net	0.4201	0.5698
RGB (3)	640	U-Net	0.4742	0.6273
RGB (3)	640	U-Net + Attention	0.4925	0.6471
RGB + H (4)	**640**	**U-Net**	**0.5212**	**0.6735**

Table 3: Results obtained using different multispectral bands after quantization.

is because the information gain is the same even though both techniques provide enhancement in opposite ends of the U-Net model. Besides, attention model can be slightly improved by modifying the spatial to feature information gain ratio within the attention block. Nonetheless, our empirical results showed an improvement of around 0.7% and 1.5% in terms of IoU and Dice only, hence, it is not significant.

Generally speaking, RGB bands seemed to be enough given having around a threshold of 500 images. In fact, the simple RGB model slightly outperformed both attention and multispectral U-Net models using 480 samples with a mean IoU (mIoU) score of 51.92% and a Dice score of 67.21%. To further verify these findings, the top five best performing multispectral models were plotted along with the RGB model and attention model with a 10% margin of error. As can be seen in Figure 4 the addition of a multispectral band and attention were significant as long as the number of images was below the second augmentation level (480 images) in terms of IoU and Dice scores. After crossing this threshold, the RGB bands provide enough information for the U-Net model to perform well. This is important to point out as UAVs flight time ranges between 10 to 30 minutes capturing a small number of images at that time. Additionally, the Kruskal-Wallis test for significance was carried out for the FPS rates at each augmentation level at a significance level of 0.05. For FPS at FP32, the Kruskal-Wallis H test indicated that there is a non-significant difference between the different augmentation levels, with a mean rank score of 5.17, 8.0, 4.5, and 8.33 for each of the four augmentation levels respectively. Given that we aim towards a low power consuming approach, adding multispectral bands seems to be the optimal approach as it requires a smaller number of images, using RGB images only and has lower computational complexity than that required by attention models.

4.2 Weed Detection Results - Post-Optimization

Table 3 summarizes the results obtained after optimization. In order to deploy our models on a system-on-chip (SoC), it was necessary to optimize the model given that SoCs are resource limited in terms of memory, speed, and power. For that, we have decided to optimize our models by storing the weights using Float-16 decimal points and on an Nvidia Jetson Nano microcomputer. The models were optimized using TensorRT and TensorFlow Lite to evaluate how processor type affects inference time, power consumption, and post-optimization performance.

4.3 Power Consumption, Inference Time, and Speed

Table 4 shows the average performance loss due to optimization and the obtained Power consumption and inference time

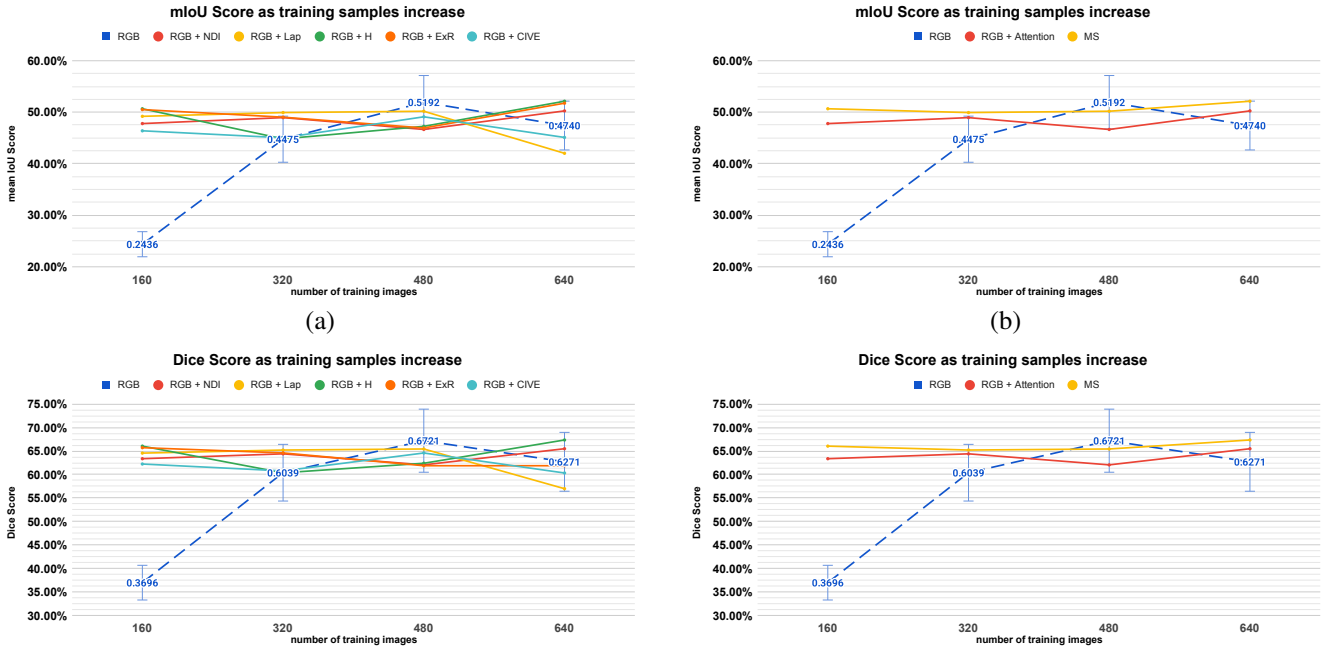

Figure 4: Performance comparison between standard RGB models, attention-enhanced RGB models, and multispectral models. (a) IoU scores of the RGB model versus multispectral models. (b) IoU scores of the best multispectral models versus RGB and attention-enhanced RGB models. (c) Dice scores of the RGB model versus multispectral models. (d) Dice scores of the best multispectral models versus RGB and attention-enhanced RGB models.

Optimization Platform	**Tensorflow Lite**			**Tensor-RT**		
Model	**Inference time (s)**	**FPS (1/s)**	**Power (W)**	**Inference time (s)**	**FPS (1/s)**	**Power (W)**
U-Net	3.6858	0.2713	4.684	0.5366	1.864	4.806
U-Net + Attention	3.5850	0.2789	4.658	0.5370	1.862	4.760
U-Net + MS	3.6415	0.2746	4.656	0.5422	1.844	4.861

Table 4: Inference time, FPS, and power consumption of different models on Tensorflow Lite and Tensor-RT optimization platforms.

results using the two libraries. Both methods resulted in a negligible loss in the overall performance with an average loss of 0.01% for TF-Lite and 0.005% for Tensor-RT. However, utilizing a GPU using Tensor-RT resulted in a lower inference time of almost 3 seconds with a slight increase in power consumption which is approximately about 0.2 Watts. For FPS at FP16, the Kruskal-Wallis H test indicated that there is a non-significant difference between the different augmentation levels, with a mean rank score of 4.67, 8.5, 5.0, and 7.83 for each of the four augmentation levels respectively. Given that real time detection requires fast inference time, it is safe to say that utilizing a GPU would be the optimal choice as it improves FPS without consuming much additional power. Our custom U-Net model was able to achieve maximum FPS rate of approximately 2. Accordingly, the ideal drone speed given 2 FPS for inference would be 16.08 m/s. However, drones generally have a limitation of 5 m/s on its maximum speed. Consequently, the maximum FPS rate required given the speed of 5 m/s is 0.62, and the minimum FPS rate required given a minimum drone speed of 1 m/s is 0.12. By capturing non-overlapping frames, these estimated values therefore elucidate our model's high FPS rate and subsequently its feasibility of conducting real-time inference. Given the low power consumption of our model, the inference can be executed directly onboard the UAV without significantly affecting flight time.

4.4 Comparison With State of the Art Models

We did a comparison between our lightweight model and the state of the art models that are reviewed in the MSEA-Net research paper. The authors show that their proposed MESA-Net is the current state-of-the-art evaluated on the same CoFly-WeedDB dataset. Their method achieved a mIoU of 71.35%, outperforming several models (Syed et al. 2025). However, as illustrated in Figure 5, this accuracy comes at a significant computational and memory cost. MSEA-Net is approximately 9.53× larger than our proposed quantized custom U-Net model. Despite the drop in IoU, our

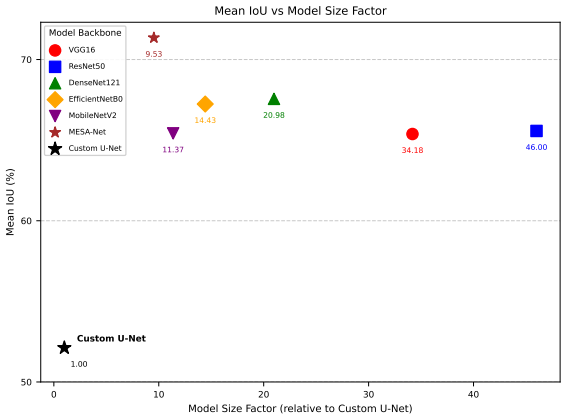

Figure 5: Mean IoU performance versus model size

model provides a substantial reduction in model size and complexity. This highlights its suitability for deployment on resource-constrained platforms, where minimizing power consumption, and onboard compute load is essential for extending flight time and improving operational efficiency.

5 Future Work

For future work, our model will be evaluated through onboard simulations in real-world environments. Moreover, power consumption can be further reduced by using LoRaWAN network modules instead of Wi-Fi networks. The reason for this is that we require an energy efficient method to send the processed images from the drone to the farmer's device. Research shows that LoRaWAN has a lower power consumption and can cover a large range when compared to Wi-Fi (de Carvalho Silva et al. 2017). Additionally, the system sets the floor for further improvements such as real-time herbicide spraying, whereby weeds are sprayed by the drone as soon as they are detected. With this approach, we reduce the farmer's workload of planning spraying routes. Another option would be to generate optimal field spraying routes using the generated scan reports, where an agricultural drone can use these routes at any time to perform efficient herbicide spraying of the field. These two potential expansions of the system would render our project a comprehensive weed management and control system. Our image dataset could also be enhanced further to reduce pixelation and improve results using an Enhanced Super Resolution GAN (ESRGAN) but at the cost of higher memory footprint. Additionally, NIR estimations from similar fields can be obtained to derive more weed sensitive indices.

6 Conclusion

In this work, we proposed an efficient real time weed detection pipeline using UAVs while maintaining a lightweight model and utilizing low power consumption. The proposed model proved to be highly effective on the CoFly-WeedDB dataset. We were able to achieve promising results using a very small model compared to the state of the art and other models on the same dataset. The effect of incorporating attention blocks and multiple multispectral bands on model performance was also analyzed. We found that features derived from RGB images and edge detectors can rapidly improve the model's performance especially using small number of images. Moreover, the models were optimized to be suitable for on-drone deployment and the inference time, power consumption, and required speed were computed.

Acknowledgements

This work was supported by the American University of Sharjah (AUS) through the Undergraduate Research Grant. The authors gratefully acknowledge the financial and institutional support provided by AUS, which made this research possible.

References

Bongiovanni, R.; and Lowenberg-DeBoer, J. 2004. Precision agriculture and sustainability. *Precision agriculture*, 5(4): 359–387.

de Carvalho Silva, J.; Rodrigues, J. J. P. C.; Alberti, A. M.; Solic, P.; and Aquino, A. L. L. 2017. LoRaWAN — A low power WAN protocol for Internet of Things: A review and opportunities. In *2017 2nd International Multidisciplinary Conference on Computer and Energy Science (SpliTech)*, 1–6.

Enterprise, D. 2022. Ground Sample Distance — DJI Enterprise. https://enterprise-insights.dji.com/blog/ground-sample-distance. Accessed: 2023-06-25.

Khan, S.; Tufail, M.; Khan, M. T.; Khan, Z. A.; Iqbal, J.; and Alam, M. 2021. A novel semi-supervised framework for UAV based crop/weed classification. *PLOS ONE*, 16(5): 1–18.

Krestenitis, M.; Raptis, E. K.; Kapoutsis, A. C.; Ioannidis, K.; Kosmatopoulos, E. B.; Vrochidis, S.; and Kompatsiaris, I. 2022. CoFly-WeedDB: A UAV image dataset for weed detection and species identification. *Data in Brief*, 45: 108575.

Liang, Y.; Kou, W.; Lai, H.; Wang, J.; Wang, Q.; Xu, W.; Wang, H.; and Lu, N. 2022. Improved estimation of aboveground biomass in rubber plantations by fusing spectral and textural information from UAV-based RGB imagery. *Ecological Indicators*, 142: 109286.

Meyer, G. E.; and Neto, J. C. 2008. Verification of color vegetation indices for automated crop imaging applications. *Computers and Electronics in Agriculture*, 63(2): 282–293.

Oktay, O.; Schlemper, J.; Folgoc, L. L.; Lee, M. J.; Heinrich, M. P.; Misawa, K.; Mori, K.; McDonagh, S. G.; Hammerla, N. Y.; Kainz, B.; Glocker, B.; and Rueckert, D. 2018. Attention U-Net: Learning Where to Look for the Pancreas. *ArXiv*, abs/1804.03999.

Saini, P.; and Nagesh, D. 2025. A review of deep learning applications in weed detection: UAV and robotic approaches for precision agriculture. *European Journal of Agronomy*, 168: 127652.

Syed, A.; Chen, B.; Abbasi, A. A.; Butt, S. A.; and Fang, X. 2025. MSEA-Net: Multi-Scale and Edge-Aware Network for Weed Segmentation. *AgriEngineering*, 7(4).