



Article

Detection of Artificial Seed-like Objects from UAV Imagery

Yanuar A. Bomantara, Hasib Mustafa * , Harm Bartholomeus and Lammert Kooistra

Laboratory of Geo-Information Science and Remote Sensing, Wageningen University and Research, 6700 AA Wageningen, The Netherlands

* Correspondence: hasib.mustafa@wur.nl; Tel.: +31-(0)-317-485-399

Abstract: In the last two decades, unmanned aerial vehicle (UAV) technology has been widely utilized as an aerial survey method. Recently, a unique system of self-deployable and biodegradable microrobots akin to winged achene seeds was introduced to monitor environmental parameters in the air above the soil interface, which requires geo-localization. This research focuses on detecting these artificial seed-like objects from UAV RGB images in real-time scenarios, employing the object detection algorithm YOLO (You Only Look Once). Three environmental parameters, namely, daylight condition, background type, and flying altitude, were investigated to encompass varying data acquisition situations and their influence on detection accuracy. Artificial seeds were detected using four variants of the YOLO version 5 (YOLOv5) algorithm, which were compared in terms of accuracy and speed. The most accurate model variant was used in combination with slice-aided hyper inference (SAHI) on full resolution images to evaluate the model's performance. It was found that the YOLOv5n variant had the highest accuracy and fastest inference speed. After model training, the best conditions for detecting artificial seed-like objects were found at a flight altitude of 4 m, on an overcast day, and against a concrete background, obtaining accuracies of 0.91, 0.90, and 0.99, respectively. YOLOv5n outperformed the other models by achieving a mAP0.5 score of 84.6% on the validation set and 83.2% on the test set. This study can be used as a baseline for detecting seed-like objects under the tested conditions in future studies.

Keywords: unmanned aerial vehicles; object detection; deep learning; flying height; light conditions; background type



Citation: Bomantara, Y.A.; Mustafa, H.; Bartholomeus, H.; Kooistra, L. Detection of Artificial Seed-like Objects from UAV Imagery. *Remote Sens.* **2023**, *15*, 1637. <https://doi.org/10.3390/rs15061637>

Academic Editors: Andrzej Staczny and Danfeng Hong

Received: 8 February 2023

Revised: 10 March 2023

Accepted: 15 March 2023

Published: 17 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last 15 years, technology related to unmanned aerial vehicles (UAV) has improved significantly, and it has been adopted for photogrammetric survey applications. In comparison to conventional methods, such as terrestrial surveys, UAV-based photogrammetry has continuous image-based coverage of tens of hectares in a single flight, has a higher spatial resolution (in millimeters) compared to satellite imagery, and is less expensive to operate than manned airborne missions [1,2]. Object detection in optical remote sensing images is a method to determine part of an image containing one or more objects belonging to the class of interest. In follow-up, the location of the detected objects can be determined [3]. There are at least four main approaches to object detection, namely template matching-based methods, knowledge-based methods, object-based image analysis (OBIA)-based methods, and machine learning methods [4]. As part of the machine learning category, deep learning methods have gained remarkable accuracy in object detection [5]. Deep learning algorithms for UAV imagery classification and segmentation have been widely used for various applications, including but not limited to palm tree counting [6], tree seedling detection [7], land use classification [8], moving vehicle detection and tracking [9], mammal detection [10], and bird detection [11]. In recent years, remote sensing image analysis has benefited greatly from deep learning [12]. A wide range of remote sensing tasks, including image classification, object detection, semantic segmentation, and change detection, have seen significant advancements thanks to the

utilization of deep neural networks. Quite a number of interesting novel approaches to remote sensing image analysis have emerged as a result of recent advances in artificial intelligence (AI) and deep learning, including but not limited to Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN), Graph Convolutional networks (GCN) and multimodal deep learning. CNN is a popular deep learning architecture particularly well-suited for image analysis tasks and can learn useful features from images automatically [13]. Attention mechanisms, such as RNN, selectively focus on certain parts of an input, such as regions of an image [14]. GCNs are specifically designed to operate on graphs to exploit spatial information in images, as well as the relationships between different spectral bands to improve the classification accuracy in hyperspectral images. Using GCN, Hong et al. introduced a method that effectively classifies images with wide range of spectral bands [15]. Multimodal approach involves the training of a deep neural network on multiple modalities of data, such as both hyperspectral and panchromatic images. It uses a diverse set of data augmentations which improve the overall classification performance. For example, Hong et al. proposed a method that involves training a neural network on both panchromatic and hyperspectral images and demonstrated that the proposed method outperforms other state-of-the-art methods for remote-sensing imagery classification. [16].

Taking inspiration from plant seeds' morphology and dispersion capabilities, the EU H2020 I-Seed project (see Figure 1) aims to develop biodegradable and self-deployable soft miniaturized artificial seeds with fluorescence sensors (termed 'I-Seed' in this work) for monitoring environmental parameters, such as temperature, humidity, carbon dioxide, and mercury in topsoil and the air above soil [17]. Within the project, winged seeds such as Samara (length 4–5 cm, width 1–2 cm) are one of the selected species for biomimicking (see Figure 2a,b). Initially, these artificial seeds will be dispersed using UAVs, and the seeds will be spread over the target area following their natural-like pathways (see Figure 1). One of the potential ways to geo-localize these artificial seeds is using UAV RGB imagery.

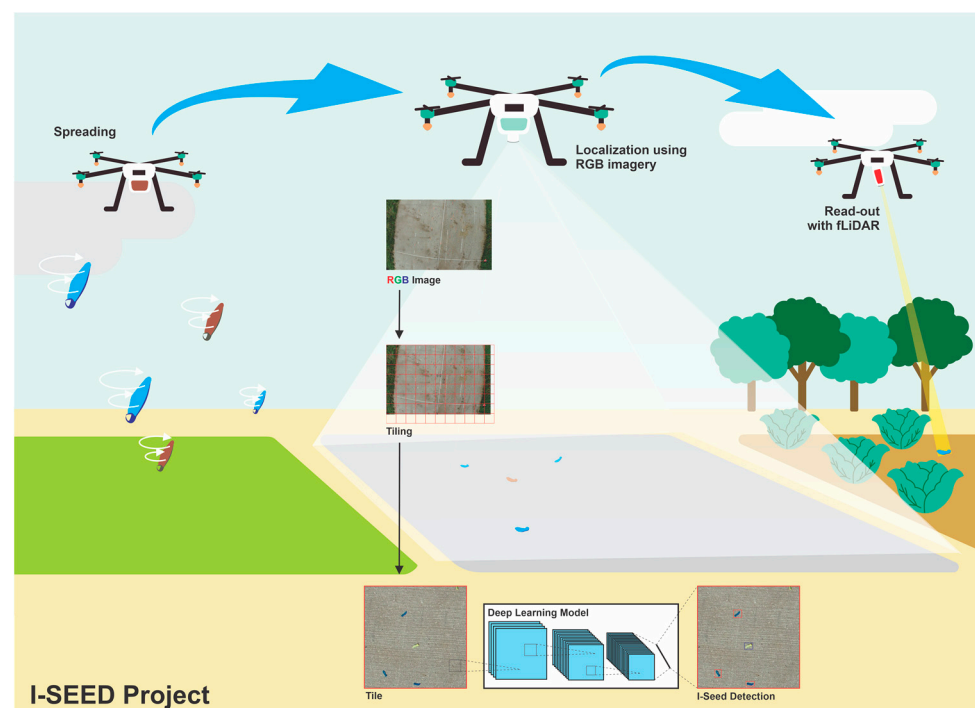


Figure 1. Overview of UAV scenario within I-Seed project in different phases: (from left to right) spreading, detection, and read-out of I-Seeds.

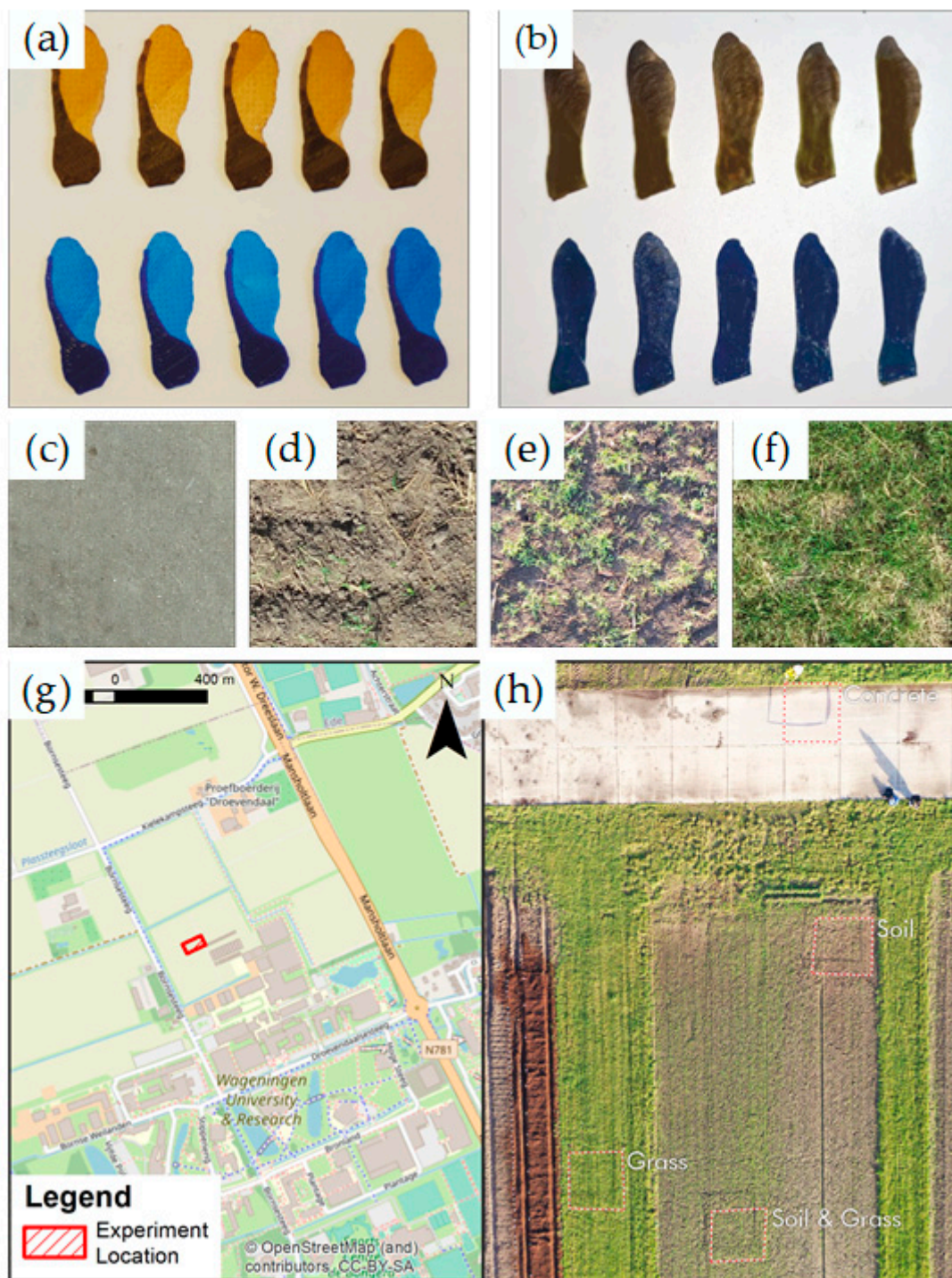


Figure 2. Seeds used in this work—(a) I-Seed prototypes and (b) colored Samara. Different backgrounds of interest—(c) concrete, (d) soil, (e) soil and grass, and (f) grass. (g) Location of the test area and (h) overview photo of the experimental setup.

Prior work on the detection of seed-like objects from aerial imagery is limited, mainly due to the definition of small object detection in aerial imagery. There are two definitions for small objects within the scope of object detection [18]. The first definition is based on the physical size of the object in the real world. For example, Liu et al. classified humans as small objects in aerial imagery using the COCO dataset [19], Chen et al. classified vehicles as small objects in aerial imagery [20], and Zhao et al. classified wheat spikes as small objects in aerial imagery [21]. The second definition is based on the image occupancy area

of the object (less than 32×32 pixels), known as the MS-COCO evaluation metric [22]. For example, Rui et al. classified UAVs as small objects in standard images (frontal perspective), as they have an image occupancy size of less than 322 pixels [23]. Samara seeds are relatively small in size (physical dimension) and will be viewed from an aerial perspective after landing on an uneven surface at various angles (small image occupancy area). By both definitions, I-Seeds fall within the category of small object detection.

Although there has been significant progress in small object detection with deep learning in recent years, improvements are still needed in the accuracy of small object detection in comparison with normal-sized object detection [5,24]. The reasons for the lower accuracy of small object detection lie within the design of the state-of-the-art detectors, which are (a) intended for frontal perspective images and not for aerial perspective images and (b) optimized to detect large and normal-sized objects, resulting in the receptive field not being robust enough for small objects [19]. To improve the performance of small object detection in aerial images, conventional state-of-the-art detectors can be specialized by reducing anchor sizes, using multi-scale feature learning, and using data augmentation [25]. State-of-the-art object detection models can be divided into two types based on their working principle, namely two-stage detection frameworks and one-stage detectors [19]. The two-stage detection frameworks first generate region proposals and then perform classification and bounding box regression for each proposal. Examples of two-stage detection frameworks are Fast R-CNN, Faster R-CNN, Feature Pyramid Network, and Mask R-CNN. In contrast, one-stage detectors perform a one-pass regression of class probabilities and bounding box locations. Examples include You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD). The main difference between these two types of one-stage detector lies in the trade-off between accuracy and processing time [19,26]. The comparison between SSD and YOLO as well as different versions of YOLO have been carried out by numerous authors. For example, Liu et al. developed UAV-YOLO from YOLOv3 and defined humans as small objects (in terms of image occupancy area) from UAV perspective [14]. They also compared their model with two variants of Single Shot MultiBox Detector (SSD) and found that SSD variants perform poorly in terms of mAP, IoU and processing time in comparison to YOLOv3 and UAV-YOLO. Zhao et al. defined small aircrafts as small objects (in terms of image occupancy area), and exploited YOLOv3 without any modification and showed that its performance in computational time is 3.4 and 27 times faster than SSD and Faster R-CNN, respectively [27]. Nina et al. compared YOLO and You Only Look Twice (YOLT) and found better performance of YOLO for Mini Ship and High-Resolution Ship datasets [28]. In the context of the I-Seed project, the artificial seeds need to be detected and localized in real time for the subsequent activity of read-out of the sensors. For remote read-out of in situ sensor information (see Figure 1), an active laser-induced fluorescence observation system onboard UAV will excite the I-Seeds for fluorescence emission that is the function of measured environmental parameters on topsoil and air above soil, e.g., temperature, humidity, CO₂, and mercury [29]. In addition, the real-time operation is aimed to be performed onboard UAV with small form-factor computational unit. On top, remote uninhabited areas have poor cellular coverage to meet real-time data streaming to a remote server/computational unit. Therefore, neural networks requiring offline/remote processing, longer processing time, and heavy computational power are disregarded for our investigation. To support the real-time detection of small objects such as I-Seeds, one-stage detectors, such as YOLO, are a suitable option. Several studies have utilized the modified YOLO model for small object detection in remote sensing images. For example, Liu et al. developed UAV-YOLO by concatenating two equally sized residual blocks of the network backbone and reported better performance of UAV-YOLO compared to YOLOv3 and SSD using an optimized training dataset [19]. Pham et al. proposed YOLO-fine, which can give better results in small and very small objects using finer detection grids [26]. To detect small objects from aerial imagery, Ali et al. modified a YOLOv4 neck structure by passing the output of the fourth Convolution + Batch Normalization + Mesh (CBM) layer into an up-sampling layer with a factor of 4, which generates more refined and

fine-grained features of the small object [30]. Zhao et al. enhanced YOLOv5 by adding a microscale layer and a custom anchor box setting and adapting the confidence loss function of the detection layer based on the intersection over union (IoU) [21].

Another noteworthy reason for the lower accuracy of small object detection is the diverse and complex background, which leads to false positive predictions [31]. To counteract the effects of a complex background, either the settings of the camera can be optimized during acquisition or the deep learning model can be adapted to varying background conditions during the post-processing phase. However, commercially available UAV RGB cameras are seldom configurable and generally capture images in automatic mode, making it inaccessible to finetuning hardware settings. Alternatively, deep learning models have been engineered to detect objects with complex backgrounds [31–33]. In the case of YOLO, the algorithm reasons globally to avoid false positives on complex backgrounds and encodes contextual information about the image [34].

To extract positional information of the distributed I-Seeds on the ground from high-resolution UAV imagery in an efficient and accurate manner, the required object detection model should be capable of small object detection in a complex image background, such as concrete, soil, or grass, from the aerial perspective. In this paper, we propose and evaluate a suitable deep learning-based detection model for small object detection with high accuracy at effectively detecting I-Seed objects in each aerial image. The novelty of this work is two-fold and considers primarily (a) effect of environmental parameters in data acquisition on the detection of small seed-like objects from aerial perspective, and (b) implementation and evaluation of a one-stage detector in combination with SAHI to reduce processing time and computational load for full-sized images in order to implement on UAV platform for real-time object detection and localization. We prepared an experimental dataset consisting of imagery acquired at different flying altitudes with varying daylight and background conditions to evaluate the generalizability of the model and to determine the detection limits of the deep learning algorithm. Subsequently, we carried out a performance analysis for the detection of I-Seeds in full-size images and the subsequent processing time requirements.

2. Materials and Methods

2.1. Sample Preparation

Generally, deep learning models require a large number of observations as input for the data labelling process. With respect to the requirement of model training, the available number of I-Seed prototypes were 10 brown and 10 blue objects (see Figure 2a). Therefore, real Samara seeds (*Acer platanoides*) of similar dimensions to the I-Seed prototypes were used in addition. To mimic the prototypes' colors, the real Samara seeds were used in their original color (brown), as well as being painted in blue (see Figure 2b). The total number of seed-like objects in this study was therefore 160 brown objects and 160 blue objects. The seeds were distributed by hand within delineated plots of 3 m × 3 m within the study area over four different land cover types: concrete, bare soil, soil with grass patches, and grass only (see Figure 2c–f).

2.2. Data Acquisition

The study site is located at Unifarm, the agricultural experimental research farm of Wageningen University and Research, the Netherlands, at latitude 51°59'21.45"N and longitude 5°39'38.97"E (see Figure 2g,h). The images for the dataset were captured by two UAV platforms, a Phantom 4 Pro RTK and a DJI Mavic 2 Pro, for which the camera specifications are listed in Table 1. Both the UAVs have a maximum flight time of 30 min. Since the spatial resolution of the UAV camera is inversely proportional to the target distance, the UAV flying altitude is limited for resolvable data acquisition. The minimum flying altitude was found to be 4 m, where the downwash airflow from the rotors of the UAV did not cause location shifts of the seeds. The images were captured at flying altitudes of 4 m to 10 m with altitude increments of 2 m. To account for the contrast and exposure differences arising from the variation of daylight condition during UAV image capture, the

UAV missions were carried out on different days with sunny and cloudy skies (Table 2). The solar irradiation intensities were collected from the nearby Veenkampen weather station of Wageningen University and Research, the Netherlands. The image acquisition time for all the datasets was between 12:00 and 15:00 local time, and the corresponding solar zenith angles are listed in Table 2. Note that clouds diffract the incoming solar radiation, resulting in higher diffuse irradiation than direct solar irradiation. Two different image capturing methods were utilized, namely capture at equal distance/time interval and Hover and Capture. The former is employed in autonomous flight missions and is prone to motion blur, whereas the latter is used in non-automated flight missions and is more time-consuming.

Table 1. Camera specifications of UAV platforms used in the experiments.

Parameter	Phantom 4 Pro RTX	Mavic 2 Pro
Field of View	84°	77°
Minimum Focus Distance	3.3'/1 m	3.3'/1 m
Still Image Support	DNG/JPEG	JPEG
Mechanical Shutter Speed	1/2000 to 8 s	N/A
Electronic Shutter speed	8–1/8000 s	8–1/8000 s

Table 2. Overview of image datasets acquired over the experimental plots (Figure 2) and the related environmental information in the period of acquisition.

Dataset ID	Date	Light Condition	Solar Intensity [W/cm ²]			Solar Zenith Angle (°)	Drone Platform	Flight Method	Total Images	Category			Full Size Dataset (FSD)
			Specular	Diffuse	Tiled Dataset (TD)								
					Training					Validation	Testing		
DS1	21 Sep. 2021	Overcast	157 ± 60	305 ± 27.8	52.2 ± 0.8	Phantom 4 Pro RTX	Auto	359	P	P			
DS2	11 Nov. 2021	Overcast	0.98 ± 0.1	75 ± 14	71.1 ± 1.4	Phantom 4 Pro RTX	Auto	816	P	P			
DS3	21 Dec. 2021	Sunny	652 ± 41.7	32.5 ± 3.3	76.4 ± 1.1	Phantom 4 Pro RTX	Auto	816	P	P			
DS4	6 Jan. 2022	Sunny	11 ± 25.8	42.4 ± 34.1	77.6 ± 1.0	Mavic 2 Pro	Manual	139		P	P	P	
DS5	9 Jan. 2022	Sunny	60.4 ± 91.7	38.3 ± 4.6	78.1 ± 1.1	Mavic 2 Pro	Manual	131		P	P	P	
DS6	17 Jan. 2022	Overcast	0.8 ± 6.3	59.8 ± 7.8	75.3 ± 1.7	Mavic 2 Pro	Manual	202		P	P	P	

2.3. Datasets

For this research, seven datasets were acquired within the period of September 2021 to January 2022 (Table 2). In each dataset, an image is captured at each flight altitude and each background. For the creation of the dataset, at least two shooting methods were utilized. The first was a self-directed flight with 70–95% overlap between photographs. It seeks to reproduce the photographs taken so that there is a large amount of data to include in the training model. The second type of data acquisition was performed by flying manually over a bounded area of 3 m × 3 m with 14 I-Seed blue and 14 I-Seed original color. This second strategy focuses on the evaluation of the deep learning method.

The datasets are divided into two categories (Table 2). The first one is the tiled dataset (TD) that is used in the training (70%), validation (15%), and testing (15%) of the transfer learning model YOLOv5. Note that, there were no identical images used in training, validation, or testing. The tiled as well as full-sized images used in training, validation and testing were unique, although they come from the images taken on same day. Such practice is quite common in dataset design and analysis, for example, see the works of Liu et al. [19], Chen et al. [20], and Zhao et al. [21]. The images are tiled, sub-images with background information are discarded, and sub-images containing I-Seeds are selected and labelled manually. The second dataset is the full-size dataset (FSD) that contains full resolution images (5472 px × 3648 px) taken by the drone without tiling. The purpose of

this dataset is to assess the identification of I-Seed YOLOv5 at full size. FSD consists of 192 full-size images. Since the design of the datasets influences the accuracy of the model being trained, datasets containing images captured in equal distance/time interval mode including image blur are allotted to the training set of TD. Meanwhile, datasets using the Hover and Capture method are allocated for the FSD and test sets of TD. These datasets include images without motion blur and ground truth data.

2.4. Data Pre-Processing

Since the full resolution images captured by the UAV contain a considerable amount of background information less relevant to the training process (42%), tiling is utilized as a pre-processing step to filter out irrelevant portions of the image. In the tiling process, the full resolution image is divided into tiles with a size of 512×512 pixels, highlighting the I-Seed characteristics while avoiding the loss of image information. A total of 12,300 tiles was generated for TD, of which 42% did not contain any I-Seed objects (background information).

Labelling is an essential step for training in supervised learning. Roboflow [35], a web-based image annotation platform for computer vision algorithms, was used to manually label objects and to create bounding boxes around each object in the image. One of the major issues in object detection is class imbalance arising from the large differences in objects per class. As is shown in Table A1, a total of 17,760 objects was annotated for the tiled dataset (TD), where 59% of the annotations belong to I-Seed blue and 41% of the annotations belong to I-Seed original color (brown). This division makes class imbalance insignificant for our case. Since tiles containing a single object and multiple objects fare equally, care was taken to avoid annotating objects at the margin of the tile. The maximum number of objects per tile was less than 16, and in practice, each tile is expected to contain between 1 and 8 items at a resolution of 512×512 pixels. For the full-size dataset (FSD), a total of 192 images was labelled and 6526 objects were annotated to assess the accuracy of the model while running inference with slicing-aided hyper inference (SAHI). Ground truth on number of seeds over an area was ensured by dispersing seeds over a bounded area of $3 \text{ m} \times 3 \text{ m}$. Therefore, each image in this dataset should ideally comprise 34 items, 17 I-Seed blue objects and 17 I-Seed original color objects. However, two photos contain only 33 objects, including 16 I-Seed objects in their original color. The reason for the reduced number of items in these images is because they contain 1 less I-Seed original color object. This approach was taken to ensure that the model evaluation can be performed against the ground truth of seed numbers, and human error in the data labelling step can be avoided. The TD test data tiles are generated from FSD to maintain the identical condition and quality of data for training and testing.

2.5. Data Processing

2.5.1. Object Detection Model

YOLOv5, proposed by Glenn Jocher in 2020, has been developed using the PyTorch library of Python and has various versions depending on the network size, e.g., small (YOLOv5s), medium (YOLOv5m), large (YOLOv5l), and extra-large (YOLOv5x) [36]. Generally, the performance of the model increases with increasing network size at the cost of processing time [36]. That is, larger models can handle complex problems with large datasets with good accuracy, but a longer processing time is required compared to smaller models. In terms of integration, Python-based YOLOv5 is simpler than the previous C-based YOLO versions and is one of the state-of-the-art techniques for object detection [37]. In this work, four variants of the YOLOv5 model, namely nano, small, medium, and large, were utilized and compared for training the dataset for I-Seed detection. The key differences between the variants are listed in Table 3, where depth and width multiples represent the addition of layers and channels to the neural network, respectively. The hyper parameters in a deep learning model play a key role in the learning process by controlling

and deciding the model parameters that a model learns. The hyper parameters for this work are listed in the Appendix A (see Table A2).

Table 3. Comparison of YOLOv5 variants in terms of crucial features.

Variant	Depth Multiple	Width Multiple	Parameters
Yolov5x	1.33	1.25	86.7 M
Yolov5l	1.0	1.0	46.5 M
Yolov5m	0.67	0.75	21.2 M
Yolov5s	0.33	0.50	7.2 M
Yolov5n	0.33	0.25	1.9 M

2.5.2. Performance Evaluation

The performance of the YOLOv5 model for I-Seed detection was evaluated using standard performance measures such as average precision (AP) and mean average precision (mAP). Average precision AP is calculated as the area under the precision–recall curve $p(r)$ for a given intersection over union (IoU) α and mean average precision mAP is the average of AP over n observations as follows:

$$AP@{\alpha} = \int_0^1 p(r) dr, \quad (1)$$

$$mAP@{\alpha} = \frac{1}{n} \sum_{i=1}^n AP_i. \quad (2)$$

IoU α evaluates the overlap between the ground truth area GT with the predicted area PD as follows:

$$\alpha = \frac{area(GT \cap PD)}{area(GT \cup PD)}. \quad (3)$$

Precision p depicts model performance when identifying relevant objects only, while recall r measures the model performance when finding all relevant cases (all ground truth) in terms of true positive TP , false positive FP , and false-negative FN as follows:

$$p = \frac{TP}{TP + FP} \quad (4)$$

$$r = \frac{TP}{TP + FN}. \quad (5)$$

For example, when $\alpha > 65\%$ on $AP50$, the detection is considered a true positive because the IoU value exceeds the threshold, whereas for $AP75$, it is a false positive. Alongside accuracy, model performance also considers the time needed for the model inference to account for the real-time detection of the I-Seeds.

2.5.3. Inference in the Full-Size Dataset

The full-size dataset (FSD) is provided as an input to the trained YOLOv5 model for detecting the I-Seeds. YOLOv5 is trained on $512\text{px} \times 512\text{px}$ input images, whereas each image in FSD is $5472\text{px} \times 3648\text{px}$. On the one hand, if the FSD is fed directly as input into YOLOv5, the details in the image will be lost due to the size reduction. On the other hand, if YOLOv5 is trained for the FSD size, the required computational resources and memory for training and interference will increase significantly. Slicing-aided hyper inference (SAHI) was utilized for detecting smaller objects without retraining the model or requiring larger graphical processing units (GPUs). Initially, the original image I is sliced into N number of $K \times L$ overlapping patches before progressing to object detection forward pass. The results of the overlapping predictions and optimal full-inference (FI) are merged using non-maximum suppression (NMS) [38]. SAHI gives the predictions of the full-size image in COCO JSON format. Due to the fact that the FSD has a unique

name indicating the state of the three parameters considered to influence object detection, the SAHI and ground truth detection results are classified according to the class/value of each parameter or the combination of the three parameters, namely light conditions, flight height, and backgrounds. This separation is accomplished through the use of Python to parse the COCO JSON format for both SAHI detection results and ground truth labelling. Afterwards, the effect of each parameter and its combination on the model's detection capability for I-Seed is evaluated.

3. Results

3.1. Training Comparison

I-Seed TD is a small dataset that contains only two classes of I-Seed: I-Seed blue and I-Seed original color. Due to the model's complexity, training the YOLOv5 variants from scratch with small datasets is ineffective for obtaining high-accuracy models. Because deep learning models require more data than conventional machine learning methods due to the increased complexity of the models, using pre-trained weights is a well-known way to train data. Pre-trained YOLOv5 models trained on the COCO dataset were employed and transfer learning strategies were used to improve the model's performance while using less data and time in this research. As can be seen in Figure 3a, the mAP0.5 of all the model variants converge on near-optimal values around the 100th epoch due to fine-tuning. Convergence is often related to the size of the dataset. YOLOv5 is trained for 250–300 epochs on a big dataset, as is the case for the COCO dataset. Since this model has previously been fine-tuned on the COCO dataset, and the I-Seed dataset is small with only two classes, it does not require a large number of epochs to learn. If training is continued for a large number of epochs, the model will over-fit and good generalization will be unlikely. As can be seen in Figure A1, the losses for YOLOv5 models' variants over epochs consistently decrease for all the variants, indicating inherently that YOLOv5 is a good fit to model this problem. Moreover, the loss rates of YOLOv5n and YOLOv5s are lower due to the lower number of parameters resulting in less overfitting than bigger variants. After training all YOLOv5 variants on the I-Seed dataset, the mAP0.5 of the best weight for each YOLOv5 variant was calculated for both the TD validation and test sets. As can be seen in Figure 3b, the trends for TD test and validation datasets for all the categories are similar, demonstrating that the model has sufficient generalization capabilities and does not over-fit the training set. For both datasets, YOLOv5n outperforms the other three models, YOLOv5m performs worst for the test dataset, and YOLOv5l performs worst for the validation dataset.

Table 4 lists the variant-specific information of YOLOv5 trained with our I-Seed dataset. The time it takes for pre-processing and non-max suppression (NMS) is roughly the same for all models, whereas the major variation in speed lies in the inference. As the size of the model increases, the inference time and the number of layers and parameters increase. Although it is assumed that the size of the model has no bearing on the success of deep learning, larger models appear to fare poorly in terms of mAP0.5, as is shown in Table 4. Therefore, as the complexity of the model rises, more widespread networks tend to underfit in the computational sense [39]. In addition, when objects are small and simple in size, there are typically inadequate features to aid in the training process regardless of whether the nodes are activated [40].

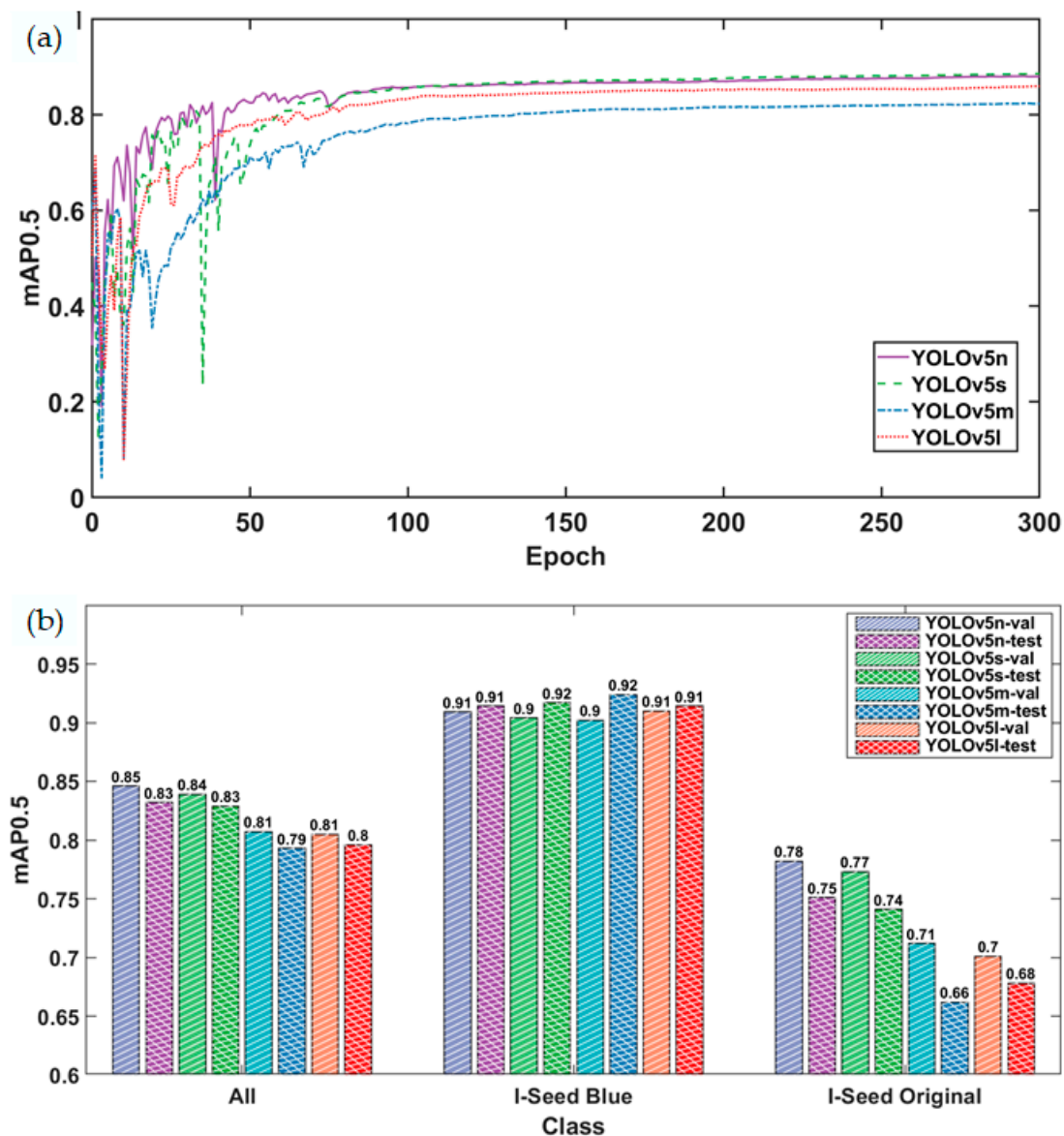


Figure 3. (a) Mean average precision along I-Seed YOLOv5 model training. (b) Mean average precision from each YOLOv5 variant's best weight for the validation and test datasets.

Table 4. Details of variants of YOLOv5. The layers extract meaningful information from the input and the parameters indicate the number of weights and biases in the whole model.

YOLOv5 Variants	Training Accuracy			Speed (ms)			Processing		
	Precision	Recall	mAP _{0.5}	Pre-Process	Inference	NMS	Total	Layers	Parameters
YOLOv5n	0.84	0.86	0.88	1.2	1	0.8	3	213	1,761,871
YOLOv5s	0.86	0.84	0.89	1.2	2.3	0.8	4.3	213	7,015,519
YOLOv5m	0.83	0.77	0.82	1.2	2.3	0.8	7.2	290	20,856,975
YOLOv5l	0.85	0.81	0.86	1.2	9.1	0.9	11.2	367	46,113,663

3.2. Inference on Full-Size Image

Figure 4 illustrates examples of the detection results obtained after SAHI on different backgrounds (concrete and grass). When the detection results in Figure 4a are compared to the prediction results in Figure 4b, it can be observed that the red labels indicate true positives within the marked area, and almost all the seeds are detected with a score higher than 0.77. Outside the area, the red labels indicate all the false positive results obtained,

where the detected seeds are not present. In the case of a concrete background, there are very few false positives, resulting in a high accuracy. However, the results obtained on a grass background are not as accurate as those obtained on a concrete background, as can be observed from Figure 4c,d. Not only are the number of false positives higher (the red labels outside the marked area) but some of the seeds present in the ground truth image are not detected (green labels), leading to a lower accuracy of the model with a grass background. In Figure 4e, the results of SAHI for the three size categories are shown, namely all sizes, small, and medium. Large objects with a size greater than $96\text{px} \times 96\text{px}$ were not categorized because there are no detections made in that size. As can be observed from Figure 4e, the best mAP0.5 across all classes was obtained for medium-sized objects. Additionally, the mAP0.5 of the small I-Seed is only slightly lower than the mAP0.5 of the whole I-Seed, whereas the mAP0.5 of the medium-sized I-Seed is significantly higher. This indicates that the detection performance of small I-Seeds is not significantly superior to that of medium-sized I-Seeds. The number of medium I-Seeds, on the other hand, is not as large as the quantity of small I-Seeds.

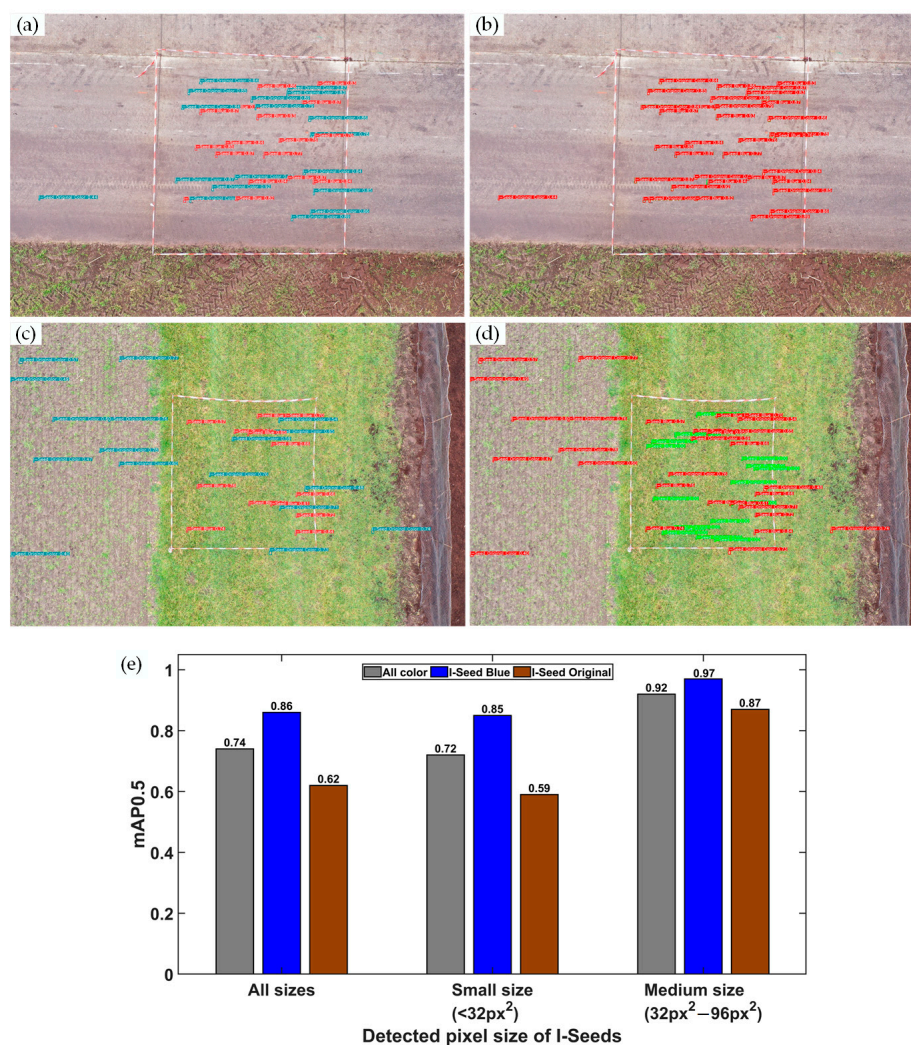


Figure 4. Example of detection and comparison in overcast conditions from 6 m flying altitude. (a) Detection of I-Seed blue (red) and I-Seed original color (blue) for concrete background. (b) Detection (red) vs. ground truths (green) for concrete background, where detection outside the boundary box is a false positive. (c) Detection of I-Seed blue (red) and I-Seed original color (blue) for grass background. (d) Detection (red) vs. ground truths (green) for grass background, where green indicates a false negative. (e) mAP0.5 of I-Seed YOLOv5 Nano with SAHI.

3.3. Model Assessment

The performance of YOLOv5n was evaluated under different environmental conditions, namely daylight condition, flying altitude, and type of background. In Figure 5a, the effect of different daylight conditions (overcast and sunny) is visualized. It can be seen that of the two conditions, greater accuracy is achieved on an overcast day. For both light conditions, the class representing blue I-Seeds has the highest mAP0.5 score. The relationship between flying altitude and mAP0.5 can be seen in Figure 5b, where mAP0.5 increases with decreasing altitude, with comparable behavior for both classes. The best and worst performance for both classes was achieved at an altitude of 4 m and 10 m, respectively. In Figure 5c, the influence of different backgrounds on the mAP0.5 score is plotted. Out of the four different backgrounds, the highest mAP scores were obtained for concrete background in both classes (>0.94). Similar to the pattern seen whilst analyzing the other parameters, the worst performing class was the original, with the lowest score being 0.06 on grass. The overall worst performing background was also shown to be grass, as all the seeds had their lowest score against this background.

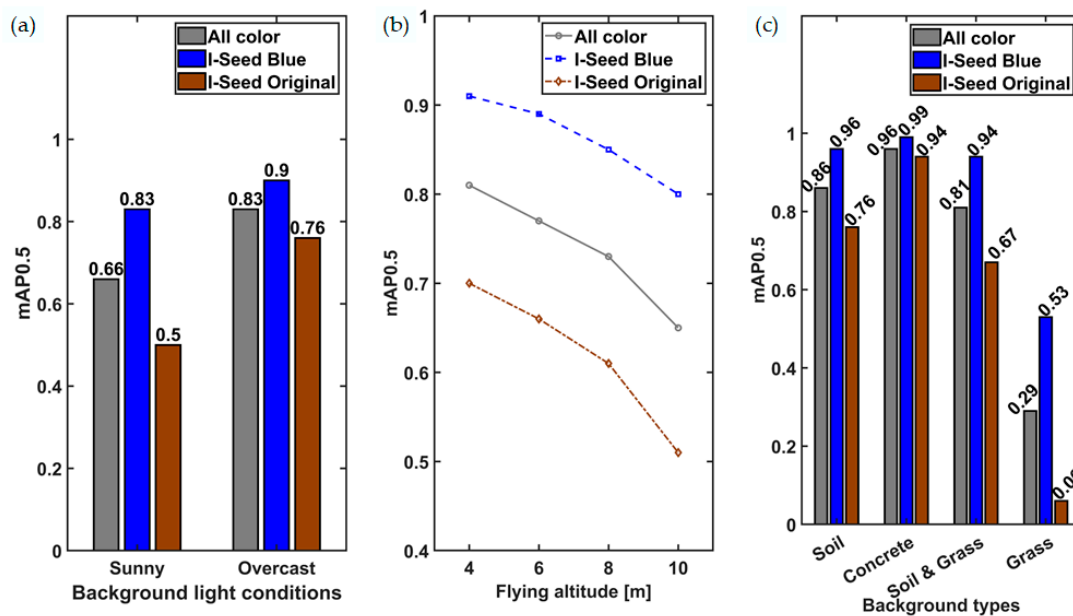


Figure 5. Influence of different parameters on mAP0.5 score: (a) daylight condition, (b) flying altitude, and (c) background type.

3.4. Parameter Combination

To thoroughly test the variants of YOLOv5n on our I-Seed dataset, 32 unique conditions in total were tested using two metrics, namely mAP(0.5:0.95) and mAP0.5, as is shown in Figure 6. mAP0.5 represents the scenario in which the IoU threshold is fixed at 0.5, whereas mAP(0.5:0.95) is the Average Precision at IoU values of 0.5 to 0.95 with a 0.05 interval. In the heat map shown in Figure 6, the blocks in shades of green indicate a higher score (i.e., better performance), the blocks in pale yellow represent an average score, and the red shades indicate the worst performance. It can be seen that combinations that involve a grass background have a much lower score than combinations with other backgrounds, where the majority of the blocks are shades of red, indicating a fairly low score. The metric mAP(0.5:0.95) is important to determine the accuracy of the model's bounding box predictions. When the mAP(0.5:0.95) is compared to mAP0.5, it ranges between 0.4 and 0.7. This suggests that the IoU of prediction has a value between 0.6 and 0.8. I-Seed blue has a higher IoU than the original I-Seed. This is true for the majority of parameter combinations, except in overcast conditions, at flight altitudes of 6 m and

8 m, and under all backgrounds except grass. For these six parameter combinations, the mAP(0.5:0.95) for the original color class performs somewhat better than for the blue class.

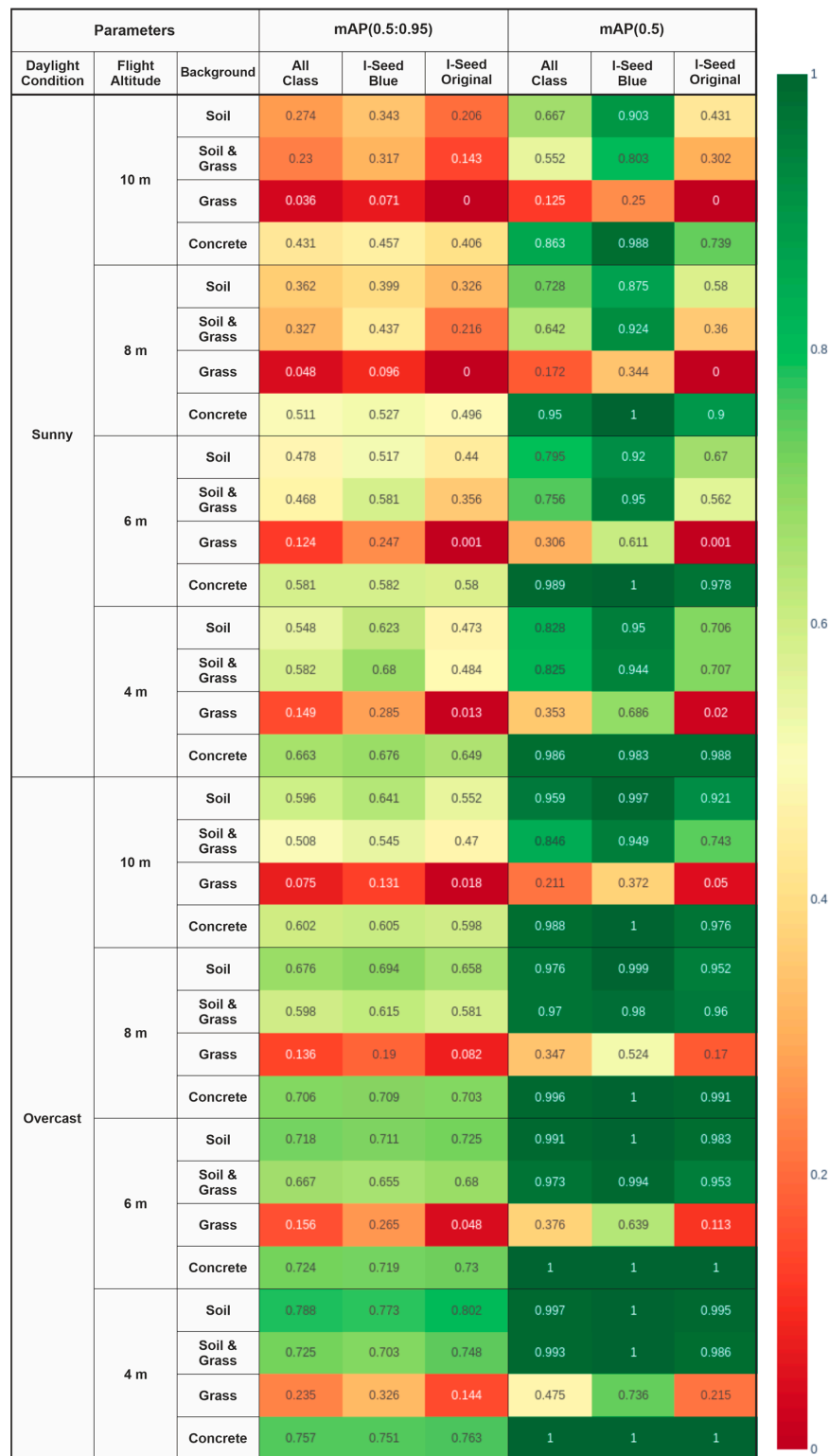


Figure 6. Detection performance for different I-Seed types under 32 unique conditions of varying environmental and acquisition parameters. The blocks in shades of green indicate a higher score for mAP(0.5:0.95) and mAP0.5 (i.e., better performance), the blocks in pale yellow represent an average score, and the red shades indicate the worst performance.

4. Discussion

The results of this study show that small object I-Seeds can be detected from UAV-acquired RGB images using the one-stage detector YOLOv5. The influence of three environmental parameters on detection accuracy was investigated for seed-like small object detection from aerial imagery. All these parameters affect the image quality, which determines the quality of the data labelling and training process.

The size of a physical object in an aerial image is proportional to the distance from the camera, i.e., the greater the flying altitude, the smaller an object in the photograph. Using the boundary tape in the FSD images as a reference for the ground truth's dimensions (Figure 4), the spatial resolution of pictures taken at 4 m, 6 m, 8 m, and 10 m is estimated to be 0.75 mm/px, 1.25 mm/px, 1.75 mm/px, and 2.25 mm/px, respectively. The more compact an object's dimensions, the fewer details the object detector can retrieve. For example, 42% of the images in the TD represent background only, whereas more than 50% of the images in FSD at higher flying altitudes (8 m and 10 m) constitute the background. The increased amount of background image induces higher rates of false positives in the prediction findings, resulting in a 6% and 20% decrease in detection performance for I-Seed blue and I-Seed original, respectively, for YOLOv5n + SAHI (Figures 3b and 4e).

In addition, the drone flies at a speed of 1 m/s with an exposure time varying between 1/50 s (overcast) and 1/200 s (sunny) on an automated mission. During the time that the image sensor of the camera is exposed to light, the drone moves 0.5 cm to 2 cm. Given the size of the seeds (approx. 3 cm), the image blur accounts for 16% to 67% of the seed size. Blurry images decrease the visual quality of photographs and interfere with the extraction of target features [41]. One of the ways to circumvent image blur is by switching the capture mode from equal distance interval to Hover and Capture mode. As is shown in Figure 7a,c, image blur becomes severe at the edges of the objects, while the images captured in Hover and Capture mode show distinct edges (Figure 7b,d). The model was trained on the images taken by automated flight to account for image blur and consequently, did not show any noticeable mAP0.5 improvement for FSD inference on non-blurry images. However, Hover and Capture mode is not realistic for automated flights, and therefore an alternate way to avoid image blur is to decrease the exposure time while increasing the ISO without compromising the optimum depth of focus. When using such an approach, the noise around the object edge is expected to increase.

An aerial image is composed of a target object and the background from a bird's-eye view. For detecting physically small objects such as I-Seeds, the overlap of the object's physical properties, e.g., dimension, shape, and color, with the image background determines the detection accuracy of the deep learning model. Since the color blue is not present in the tested backgrounds, I-Seed blue is easier to locate in the image (Figure 8), resulting in 30% more annotations during the data labelling process (Table A2). For the same reason, I-Seed blue has significantly better mAP0.5 scores than the original color I-Seed, as can be observed for different model variants varying only by 1% (Figure 3b), as well as for YOLOv5n + SAHI performing 28% better when detecting blue as opposed to original color I-Seeds (Figure 4e). The similarity in the color and shape aspects of original I-Seeds (brown in color) and background objects (e.g., soil and dry litter) induces lower contrast and smaller shape difference with the background, making original I-Seeds challenging for object detection. For example, numerous tiles containing original I-Seeds resulted in a low confidence level during the data labelling process, and therefore they were discarded. Consequently, higher variations in detection performance for different model variants and a lower mAP0.5 value can be observed for original I-Seeds in contrast to blue I-Seeds. As the complexity of the background increases (e.g., vegetated surfaces), the landing orientation and occultation of the I-Seeds by ground objects also increases. Different landing orientations other than normal to the surface reduce the object size in the image, and occultation by ground objects distorts the shape of the seeds in the image. As a result, the detection performance degrades by 46% and 88% for blue and original I-Seeds, respectively, for the grass background with a grass height of 5 cm (Figure 5c). Even though the model achieves

a mAP_{0.5} value of 0.736 for I-Seed blue on a grass background, its performance is regarded insufficient due to the presence of too many false positives on the grass background. This can be avoided by increasing the confidence level of the detection's upper limit and the match threshold for match metrics at the cost of the increased possibility of false negatives. Since detecting the I-Seeds is more important (TP and FP) than missing a few of them (FN), a lower confidence level and threshold is preferred.

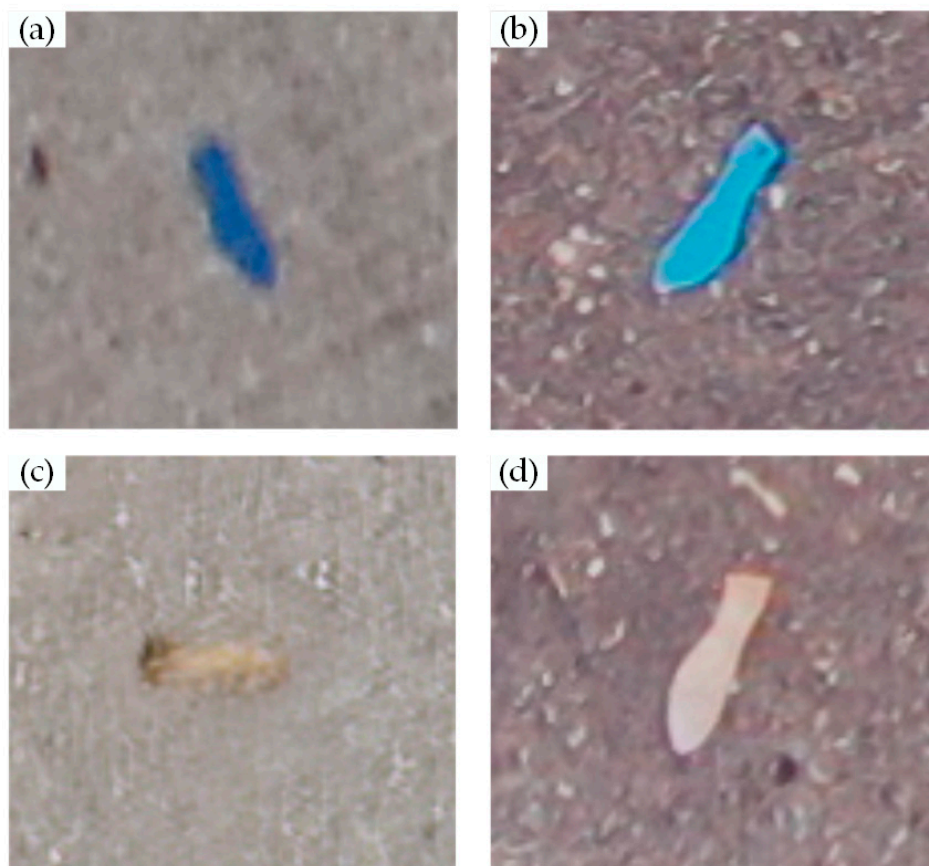


Figure 7. (a,c) represent images captured in equal distance mode. (b,d) represent images captured using Hover and Capture. All the images were captured at a flying altitude of 4 m.

The third parameter of investigation was daylight condition. Illumination plays a crucial role in defining sufficient contrast of the specific features in an image. Object detection algorithms do not identify the objects physically. Rather, they detect the objects in an image captured by the imaging system at a given illumination condition. The variation of illumination during aerial image acquisition has always been a major limiting factor for optical remote sensing studies [42]. Although YOLOv5 has been shown to be successful for small object detection in this work, daylight condition is found to play a significant role in determining the accuracy and precision of the detected image. The investigated daylight conditions, namely overcast and sunny, indicate that a higher accuracy was attained in overcast conditions (Figure 5a). This is because (a) the diffuse lighting in overcast conditions is higher (Table 2), resulting in well-lit soft lighting conditions for aerial photography, and (b) the quality of images captured under bright sunny conditions is compromised due to overexposure, which reduces the contrast between the background and I-Seed Original to a greater extent than I-Seed Blue (Figure 9). Taking the f-number (N_f), exposure time (t_{exp}), and exposure bias (B_{exp}) of the images from the image metadata, the exposure value (EV) can be calculated as $EV = \log_2 \left(\frac{N_{exp}^2}{t_{exp}} \right) + B_{exp}$ to compare different camera exposures under different illumination conditions [43]. Except for a small difference in FOV and flight

operational mode (automated and manual) between the two UAV platforms used, *EV* is analogous for both systems at ISO 100, showing a higher *EV* value in sunny conditions than in overcast conditions (Figure 9b,c). The effect of the background reflectivity on *EV* shows a significant difference in sunny and overcast conditions, as can be observed in Figure 9b. On the one hand, the contrast between the I-Seeds and the background is reduced for backgrounds with little or no ground objects (concrete and soil). On the other hand, strong shadowing effects are observed for heterogenous backgrounds such as grass. Rather than using the automatic exposure bracketing features of the camera, alternative acquisition approaches can be adopted to mitigate overexposure; for example, (a) underexposing the camera, e.g., by two stops, especially when the background is bright, (b) using physical filters such as neutral-density filters [44], tunable graduated filters [45], or digital filters [46], or (c) including image processing techniques such as automatic exposure algorithms [47].

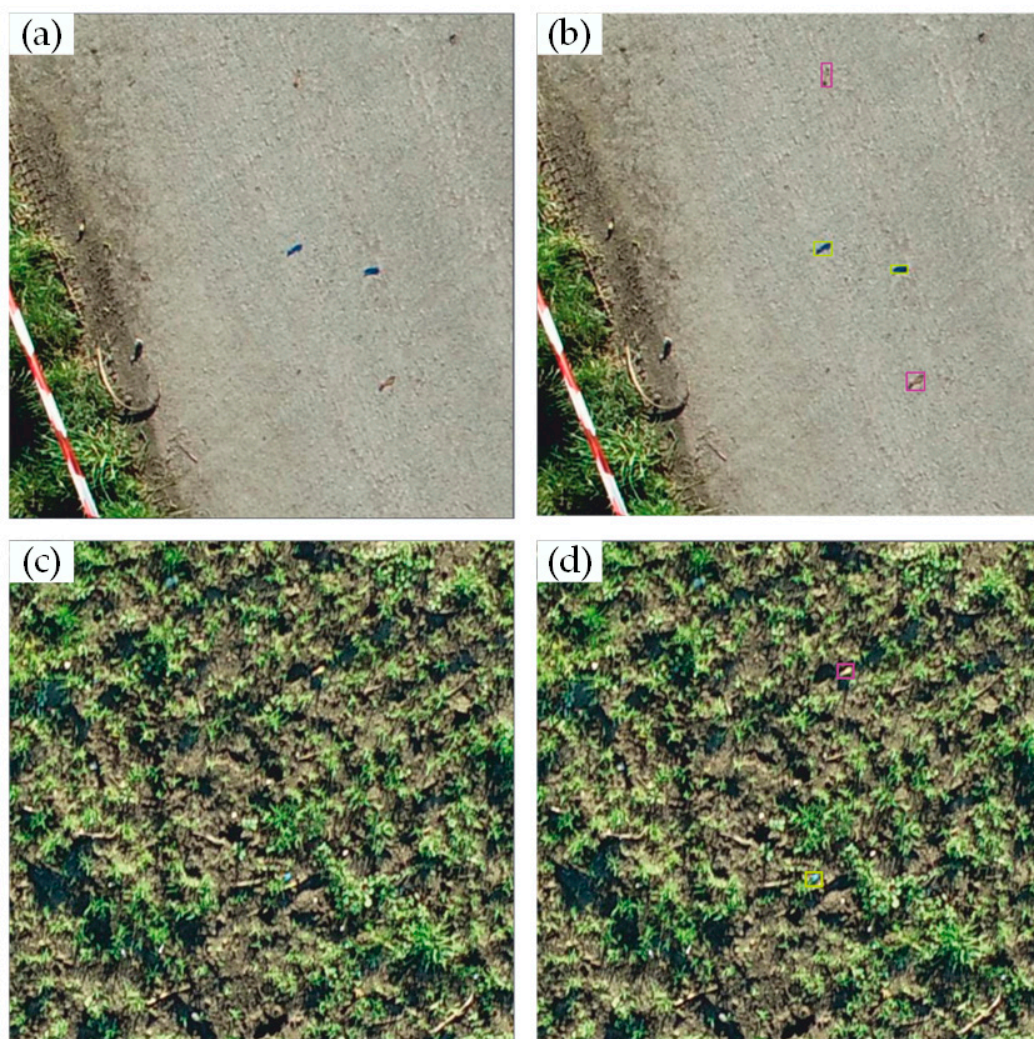


Figure 8. (a) Image captured on a concrete background, (b) annotated image captured on a concrete background, (c) image captured on a grass background, and (d) annotated image captured on a grass background. All the images were captured at a flying altitude of 4 m.

The majority of the literature on object detection from aerial perspective recognizes environmental effects on image quality and subsequent impact on object detection; however, the model for all the environmental factors tends to be generalized [48,49]. Liu et al. analyzed the effect of motion and fog blur on the accuracy of military object detection by augmenting distortions of various degree on high resolution images to simulate fog and motion blur [50]. They found a decrease of 26% and 44% in mAP0.5 value for severe

cases of fog and motion blur, respectively. In our case, the model was trained on motion-blurred images, and we did not observe any significant change in accuracy in case of blur-free images. Zhao et al. assessed the accuracy of YOLO-Highway model to detect highway center markings in different environmental conditions, namely partial occlusion and damage, as well as weak light condition [51]. They found a decrease of AP by 41% and 12% for partial occlusion and weak light condition, respectively. Tang et al. evaluated the effect of illumination (sunlight and shading) and occlusion (slight and severe) on *Camellia oleifera* fruit detection using binocular stereo vision [52]. They found no significant change in accuracy metrics for changing illumination, but observed around 5% decrease in case of severe occlusion. In our case, we observed significant effect of daylight condition in combination with background on the detection accuracy of the model (see Figure 6). We found a 20% increase in accuracy for overcast condition compared to sunny condition, and 67% decrease for grass background compared to soil and concrete, resulting primarily from occlusion (see Figure 5).

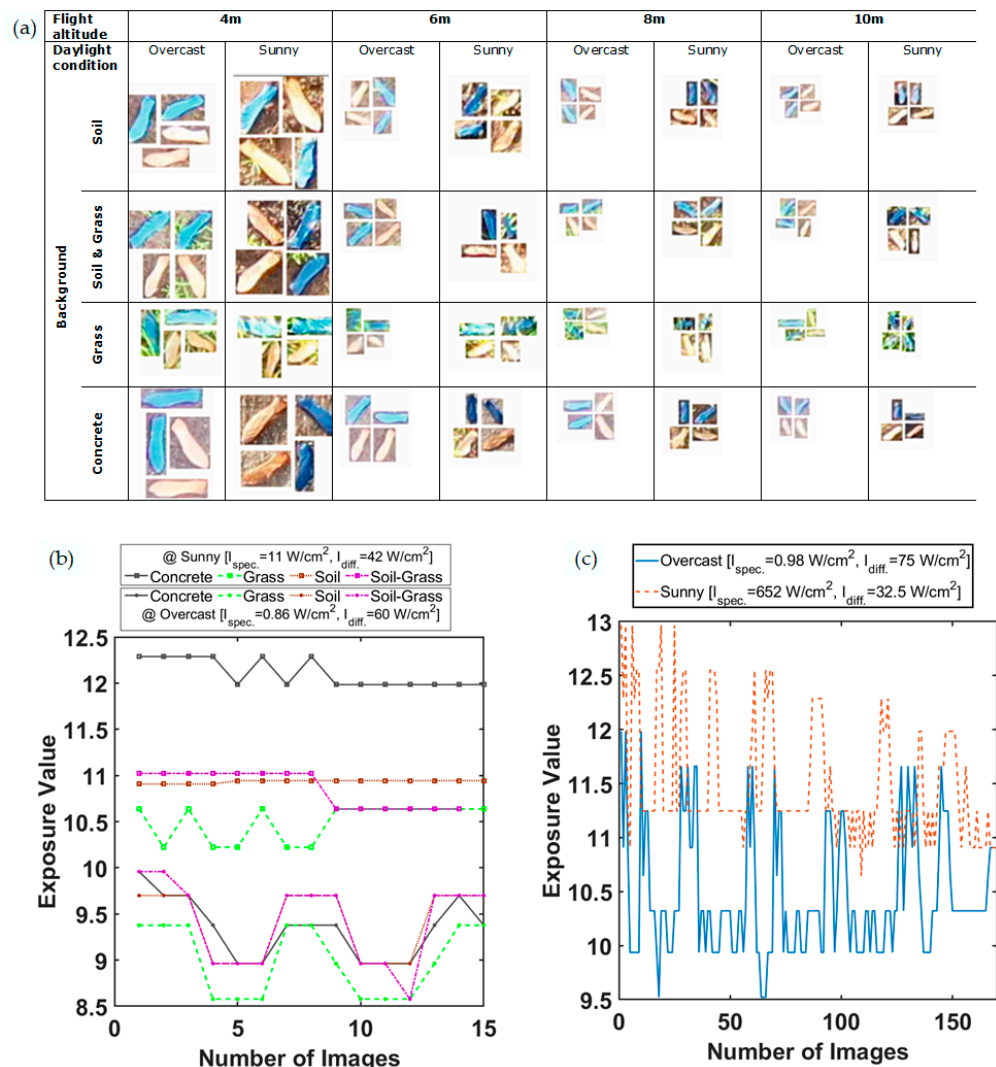


Figure 9. (a) I-Seed images with different assessment parameters. Exposure value calculated from image metadata for (b) manual flight (DS5 and DS6) and (c) automated flight (DS2 and DS3).

Within the context of this study, 32 different combinations of the three parameters were investigated, and the best mAP0.5 and mAP(0.5:0.95) result was found for the combination of overcast lighting condition, a flight altitude of 4 m, and a concrete background (see Figure 6). Of the three parameters, the major determining factor for the accuracy of seed-like

small object detection is background type, followed by flying altitude and lighting condition, respectively. As the complexity of the background increases, the detection performance degrades due to the increased number of false negatives, limiting the likelihood of detecting I-Seeds from UAV RGB imagery. Although a flight altitude of 4 m produces the best results, utilizing a low altitude in aerial view takes longer to complete image acquisition than using a higher flight altitude with the same field of view of the camera. To achieve acceptable results at a greater altitude, several post-processing steps can be employed that take into account the processing unit required for additional operations, such as applying a super-resolution method [53]. For datasets, improvements can be achieved with a higher resolution camera, by experimenting with the camera settings, or by utilizing lens filters to mitigate the effect of overexposure in bright environments, as well as by using training datasets with a more diverse focus on background photos to reduce false positives. For picture preparation, the addition of an image normalization method can be experimented with that uses RAW data as an input to reduce fluctuations in lighting conditions in the image to be used as input into the model. To make the algorithm lighter than the original, improvements on YOLOv5n used in this work can be realized via architectural adjustments, for example, (a) by lowering the width and depth multiple values while maintaining accuracy, and (b) by deleting the detection grid designed for large object identification since no object exceeds $96\text{px} \times 96\text{px}$ in size. There is always a trade-off between speed and accuracy [48,54]; however, several adjustments may also be incorporated to assess such constraints, for example, (a) optimizing the loss function for grass background, (b) setting an extra high-resolution feature map exclusively at the head of the algorithm, and (c) generating additional anchors for added details in feature maps. These aforementioned improvements will be tested in future experiments. This also includes the evaluation of recent developments in the YOLO family of algorithms (such as YOLOv6, v7, and v8) where an earlier study by Zhou et al. showed the added value of YOLOv7 for object detection in complex occluded environments [54].

5. Conclusions

This study aimed to develop a small object detection model for seed-like objects such as I-Seeds. The one-stage detection-based YOLO method was chosen to maintain the balance between accuracy and speed whilst preventing false positives on diverse and complex backgrounds. YOLOv5 and its four variants were evaluated for varying environmental conditions (daylight, flying height, and different backgrounds) using two different UAV platforms. Pre-trained models, trained on the COCO dataset, were employed, and transfer learning strategies were used to improve the model's performance. It was found that detection performance for I-Seed blue class is significantly higher than its original (brown) color class, demonstrating sufficient generalizability of the model and the absence of overfitting in the training set. The results indicated that a higher model detection accuracy was attained in overcast situations due to the clarity of the photographs. The flight altitude was found to be a major determinant of the results recorded from the detection process, clearly indicating an inverse relationship between altitude and model performance. Lastly, four different backgrounds, soil, grass, concrete, and soil and grass, were tested to analyze the performance of YOLOv5n. The amount of ground objects largely determines the model performance when detecting I-Seeds, followed by the contrast between the seed and the background. Consequently, the model performed best with a concrete background and worst with a grass background. When tested on the full-size dataset (FSD), only a slight difference was seen in the performance when compared to the tiled dataset. The FSD was used to evaluate the accuracy of the I-Seed YOLOv5n combined with slice-aided hyper inference (SAHI) as a simulation of real-time detection delivering similar results compared to the standard I-Seed YOLOv5n. This work provides a baseline for seed-like object detection from aerial imagery using a one-stage detector such that it can be implemented for the real-time detection of I-Seeds in the field during UAV flight. In future studies, the developed model will be implemented on small form-factor computational unit onboard

UAV to assess the real-time detection performance in comparison with offline processing. Other one-stage detectors, such as SSD, will also be assessed and compared with YOLOv5 in terms of detection accuracy, processing time, and computational load.

Author Contributions: Conceptualization, L.K. and Y.A.B.; methodology, L.K. and Y.A.B.; software, Y.A.B.; validation, L.K., Y.A.B. and H.M.; formal analysis, Y.A.B.; investigation, Y.A.B., L.K. and H.M.; resources, Y.A.B.; data curation, Y.A.B., H.M. and L.K.; writing—original draft preparation, H.M.; writing—review and editing, L.K., H.B. and H.M.; visualization, H.M. and Y.A.B.; supervision, L.K. and H.M.; project administration, L.K. and H.B.; funding acquisition, L.K. and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement no 101017940. The content of this publication is the sole responsibility of the authors. The Commission cannot be held responsible for any use that may be made of the information it contains.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to being in the process of uploading in a publicly accessible repository and generating the corresponding DOI. The codes can be accessed at: <https://github.com/hasib1229/I-Seed.git> (accessed on 2 March 2023).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Details on Data Processing

The size of the sub-image after segmentation is $512\text{px} \times 512\text{px}$ and the size of overlapping area is 10%. These parameters will have impact on total inferences/total time required to run inferences on a Full-Sized Dataset (FSD), and we have chosen the aforementioned parameters via trial and error for optimal performance.

The trained network was deployed on a machine with the following specifications: Ubuntu Linux 20.4, AMD Ryzen 5 3600, 32GB DDR4 RAM, and NVIDIA RTX3060Ti.

The training process is stopped after 300 epochs as mentioned in Sec. 3, which was inspired from the YOLOv5 algorithm trained on the COCO dataset. The algorithm starts overfitting if training is continued for more than 300 epochs.

Table A1. Data labelling of datasets.

	Image Size (Pixel)	No. of Images	Null Examples	Annotations			Average Annotations per Image
				I-Seed Blue	I-Seed Original	Total	
Tiled Dataset	512×512	12,300	5230	10,399	7361	17,760	1.4
Full Size Dataset	5472×3648	192	0	3264	3262	6526	34.0

Table A2. Set hyper parameters of YOLOv5.

Hyper Parameter	Value	Hyper Parameter	Value
Initial learning rate	0.01	focal loss gamma	0.0
OneCycleLR learning rate	0.1	HSE hue aug	0.015
momentum	0.937	HSE saturation aug	0.7
weight decay	0.0005	HSE value aug	0.4
warmup epochs	3.0	Rotation	0.0
warmup momentum	0.8	Translation	0.1
warmup bias lr	0.1	scale	0.5

Table A2. Cont.

Hyper Parameter	Value	Hyper Parameter	Value
box loss gain	0.05	shear	0.0
class loss gain	0.5	perspective	0.0
class BCELoss	1.0	flip up/down	0.0
object loss gain	1.0	flip left/right	0.5
object BCELoss	1.0	mosaic	1.0
IoU threshold	0.2	mixup	0.0
anchor multiple threshold	4.0	copy-paste	0.0

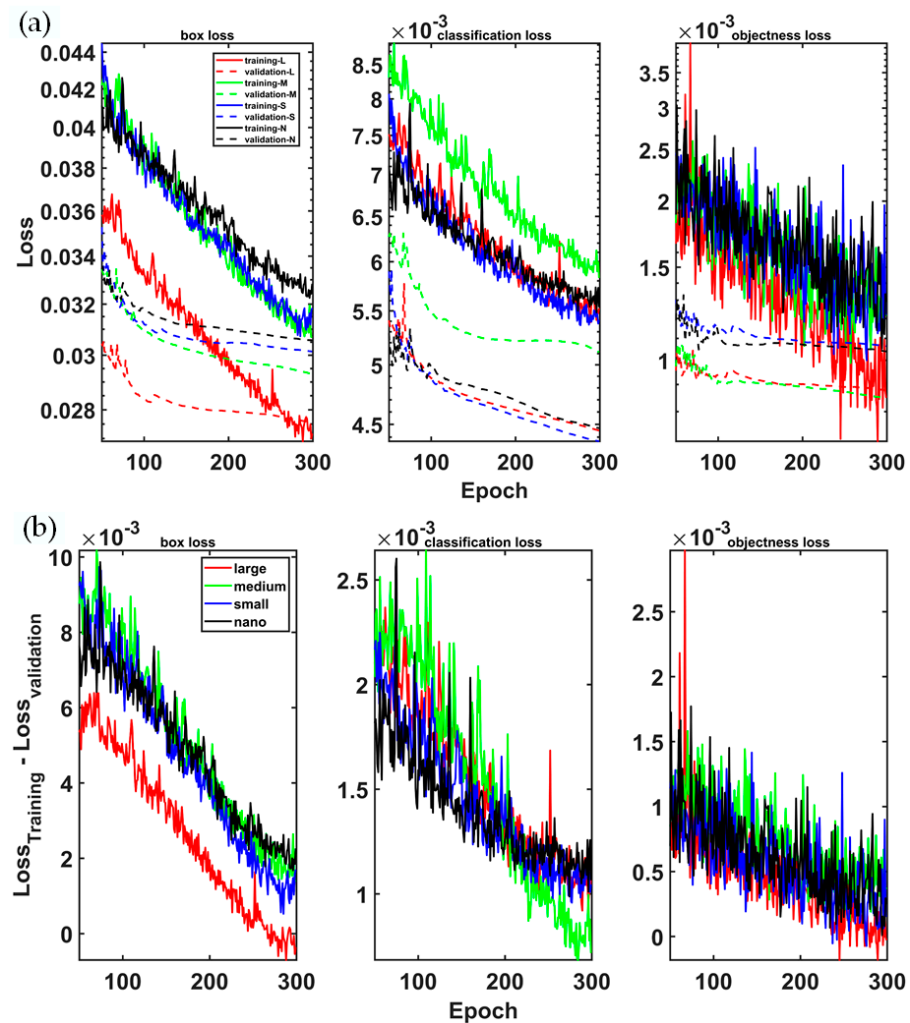


Figure A1. (a) Losses for YOLOv5 model variants as function of epoch. (b) Difference between training and validation losses.

References

- Colomina, I.; Molina, P. Unmanned Aerial Systems for Photogrammetry and Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [[CrossRef](#)]
- Sona, G.; Passoni, D.; Pinto, L.; Pagliari, D.; Masseroni, D.; Ortuani, B.; Facchi, A. UAV Multispectral Survey to Map Soil and Crop for Precision Farming Applications. In Proceedings of the Remote Sensing and Spatial Information Sciences Congress: International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences Congress, Prague, Czech Republic, 19 July 2016; International Society for Photogrammetry and Remote Sensing (ISPRS): Hannover, Germany, 2016; Volume 41, pp. 1023–1029.
- Cheng, G.; Han, J. A Survey on Object Detection in Optical Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [[CrossRef](#)]

4. Blaschke, T. Object Based Image Analysis for Remote Sensing. *ISPRS J. Photogramm. Remote Sens.* **2010**, *65*, 2–16. [[CrossRef](#)]
5. Nguyen, N.-D.; Do, T.; Ngo, T.D.; Le, D.-D. An Evaluation of Deep Learning Methods for Small Object Detection. *J. Electr. Comput. Eng.* **2020**, *2020*, 3189691. [[CrossRef](#)]
6. Ammar, A.; Koubaa, A.; Benjdira, B. Deep-Learning-Based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images. *Agronomy* **2021**, *11*, 1458. [[CrossRef](#)]
7. Pearse, G.D.; Tan, A.Y.; Watt, M.S.; Franz, M.O.; Dash, J.P. Detecting and Mapping Tree Seedlings in UAV Imagery Using Convolutional Neural Networks and Field-Verified Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *168*, 156–169. [[CrossRef](#)]
8. Zhang, H.; Wang, G.; Lei, Z.; Hwang, J.-N. Eye in the Sky: Drone-Based Object Tracking and 3d Localization. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 899–907.
9. Zhao, X.; Pu, F.; Wang, Z.; Chen, H.; Xu, Z. Detection, Tracking, and Geolocation of Moving Vehicle from Uav Using Monocular Camera. *IEEE Access* **2019**, *7*, 101160–101170. [[CrossRef](#)]
10. Kellenberger, B.; Marcos, D.; Tuia, D. Detecting Mammals in UAV Images: Best Practices to Address a Substantially Imbalanced Dataset with Deep Learning. *Remote Sens. Environ.* **2018**, *216*, 139–153. [[CrossRef](#)]
11. Hong, S.-J.; Han, Y.; Kim, S.-Y.; Lee, A.-Y.; Kim, G. Application of Deep-Learning Methods to Bird Detection Using Unmanned Aerial Vehicle Imagery. *Sensors* **2019**, *19*, 1651. [[CrossRef](#)]
12. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [[CrossRef](#)]
13. Wu, X.; Hong, D.; Chanussot, J. Convolutional Neural Networks for Multimodal Remote Sensing Data Classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–10. [[CrossRef](#)]
14. Liu, R.; Cheng, Z.; Zhang, L.; Li, J. Remote Sensing Image Change Detection Based on Information Transmission and Attention Mechanism. *IEEE Access* **2019**, *7*, 156349–156359. [[CrossRef](#)]
15. Hong, D.; Gao, L.; Yao, J.; Zhang, B.; Plaza, A.; Chanussot, J. Graph Convolutional Networks for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5966–5978. [[CrossRef](#)]
16. Hong, D.; Gao, L.; Yokoya, N.; Yao, J.; Chanussot, J.; Du, Q.; Zhang, B. More Diverse Means Better: Multimodal Deep Learning Meets Remote-Sensing Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4340–4354. [[CrossRef](#)]
17. Mazzolai, B.; Kraus, T.; Pirrone, N.; Kooistra, L.; De Simone, A.; Cottin, A.; Margheri, L. Towards New Frontiers for Distributed Environmental Monitoring Based on an Ecosystem of Plant Seed-like Soft Robots. In Proceedings of the Conference on Information Technology for Social Good, Rome, Italy, 9–11 September 2021; pp. 221–224.
18. Tong, K.; Wu, Y.; Zhou, F. Recent Advances in Small Object Detection Based on Deep Learning: A Review. *Image Vis. Comput.* **2020**, *97*, 103910. [[CrossRef](#)]
19. Liu, M.; Wang, X.; Zhou, A.; Fu, X.; Ma, Y.; Piao, C. Uav-Yolo: Small Object Detection on Unmanned Aerial Vehicle Perspective. *Sensors* **2020**, *20*, 2238. [[CrossRef](#)] [[PubMed](#)]
20. Chen, C.; Zhong, J.; Tan, Y. Multiple-Oriented and Small Object Detection with Convolutional Neural Networks for Aerial Image. *Remote Sens.* **2019**, *11*, 2176. [[CrossRef](#)]
21. Zhao, J.; Zhang, X.; Yan, J.; Qiu, X.; Yao, X.; Tian, Y.; Zhu, Y.; Cao, W. A Wheat Spike Detection Method in UAV Images Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 3095. [[CrossRef](#)]
22. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
23. Rui, C.; Youwei, G.; Huafei, Z.; Hongyu, J. A Comprehensive Approach for UAV Small Object Detection with Simulation-Based Transfer Learning and Adaptive Fusion. *arXiv* **2021**, arXiv:2109.01800.
24. Kos, A.; Majek, K.; Belter, D. Where to Look for Tiny Objects? ROI Prediction for Tiny Object Detection in High Resolution Images. In Proceedings of the 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 11–13 December 2022; pp. 721–726.
25. Courtrai, L.; Pham, M.-T.; Lefèvre, S. Small Object Detection in Remote Sensing Images Based on Super-Resolution with Auxiliary Generative Adversarial Networks. *Remote Sens.* **2020**, *12*, 3152. [[CrossRef](#)]
26. Pham, M.-T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-Stage Detector of Small Objects under Various Backgrounds in Remote Sensing Images. *Remote Sens.* **2020**, *12*, 2501. [[CrossRef](#)]
27. Zhao, K.; Ren, X. Small Aircraft Detection in Remote Sensing Images Based on YOLOv3. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Guangzhou, China, 12–14 January 2019; IOP Publishing: Bristol, UK, 2019; Volume 533, p. 012056.
28. Nina, W.; Condori, W.; Machaca, V.; Villegas, J.; Castro, E. Small Ship Detection on Optical Satellite Imagery with YOLO and YOLT. In Proceedings of the Advances in Information and Communication: Proceedings of the 2020 Future of Information and Communication Conference (FICC), San Francisco, CA, USA, 5–6 March 2020; Springer: Berlin/Heidelberg, Germany, 2020; Volume 2, pp. 664–677.
29. Mustafa, H.; Khan, H.A.; Bartholomeus, H.; Kooistra, L. Design of an Active Laser-Induced Fluorescence Observation System from Unmanned Aerial Vehicles for Artificial Seed-like Structures. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXIV, Berlin, Germany, 5–8 September 2022; SPIE: Bellingham, WA, USA, 2022; Volume 12262, p. 1226205.

30. Ali, S.; Siddique, A.; Ateş, H.F.; Güntürk, B.K. Improved YOLOv4 for Aerial Object Detection. In Proceedings of the 2021 29th Signal Processing and Communications Applications Conference (SIU), Istanbul, Turkey, 9–11 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–4.
31. Pang, J.; Li, C.; Shi, J.; Xu, Z.; Feng, H. R2-CNN: Fast Tiny Object Detection in Large-Scale Remote Sensing Images. *arXiv* **2019**, arXiv:1902.06042.
32. Jiang, B.; Qu, R.; Li, Y.; Li, C. Object Detection in UAV Imagery Based on Deep Learning: Review. *Acta Aeronaut. Astronaut. Sin.* **2021**, *42*, 524519.
33. Zhang, J.; Xie, T.; Yang, C.; Song, H.; Jiang, Z.; Zhou, G.; Zhang, D.; Feng, H.; Xie, J. Segmenting Purple Rapeseed Leaves in the Field from UAV RGB Imagery Using Deep Learning as an Auxiliary Means for Nitrogen Stress Detection. *Remote Sens.* **2020**, *12*, 1403. [[CrossRef](#)]
34. Du, J. Understanding of Object Detection Based on CNN Family and YOLO. In Proceedings of the 2nd International Conference on Machine Vision and Information Technology (CMVIT 2018), Hong Kong, China, 23–25 February 2018; IOP Publishing: Bristol, UK, 2018; Volume 1004, p. 012029.
35. Roboflow. Available online: <https://roboflow.com/annotate> (accessed on 15 November 2021).
36. Yolov5. Available online: <https://github.com/ultralytics/yolov5/releases/v6.0> (accessed on 1 December 2021).
37. Malta, A.; Mendes, M.; Farinha, T. Augmented Reality Maintenance Assistant Using Yolov5. *Appl. Sci.* **2021**, *11*, 4758. [[CrossRef](#)]
38. Akyon, F.C.; Altinuc, S.O.; Temizel, A. Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. *arXiv* **2022**, arXiv:2202.06934.
39. Jubayer, F.; Soeb, J.A.; Mojumder, A.N.; Paul, M.K.; Barua, P.; Kayshar, S.; Akter, S.S.; Rahman, M.; Islam, A. Detection of Mold on the Food Surface Using YOLOv5. *Curr. Res. Food Sci.* **2021**, *4*, 724–728. [[CrossRef](#)] [[PubMed](#)]
40. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22*, 464. [[CrossRef](#)] [[PubMed](#)]
41. Liu, X.; Shi, J.; Yu, X.; Li, X. Local Motion Blur Detection by Wigner Distribution Function. *Optik* **2022**, *251*, 168375. [[CrossRef](#)]
42. Arroyo-Mora, J.P.; Kalacska, M.; Løke, T.; Schläpfer, D.; Coops, N.C.; Lucanus, O.; Leblanc, G. Assessing the Impact of Illumination on UAV Pushbroom Hyperspectral Imagery Collected under Various Cloud Cover Conditions. *Remote Sens. Environ.* **2021**, *258*, 112396. [[CrossRef](#)]
43. Ray, S.F. *Camera Exposure Determination. The Manual of Photography: Photographic and Digital Imaging*; Taylor & Francis: Abingdon, UK, 2000; Volume 2.
44. Doumit, J. *The Effect of Neutral Density Filters on Drones Orthomosaics Classifications for Land-Use Mapping*; OSF: Peoria, IL, USA, 2020.
45. Hein, A.; Kortz, C.; Oesterschulze, E. Tunable Graduated Filters Based on Electrochromic Materials for Spatial Image Control. *Sci. Rep.* **2019**, *9*, 15822. [[CrossRef](#)] [[PubMed](#)]
46. Schöberl, M.; Oberdörster, A.; Fölsel, S.; Bloss, H.; Kaup, A. Digital Neutral Density Filter for Moving Picture Cameras. In Proceedings of the Computational Imaging VIII, San Jose, CA, USA, 17–21 January 2010; SPIE: Bellingham, WA, USA, 2010; Volume 7533, pp. 170–179.
47. Bernacki, J. Automatic Exposure Algorithms for Digital Photography. *Multimed. Tools Appl.* **2020**, *79*, 12751–12776. [[CrossRef](#)]
48. Jung, H.-K.; Choi, G.-S. Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions. *Appl. Sci.* **2022**, *12*, 7255. [[CrossRef](#)]
49. Dadrass Javan, F.; Samadzadegan, F.; Gholamshahi, M.; Ashatari Mahini, F. A Modified YOLOv4 Deep Learning Network for Vision-Based UAV Recognition. *Drones* **2022**, *6*, 160. [[CrossRef](#)]
50. Liu, H.; Yu, Y.; Liu, S.; Wang, W. A Military Object Detection Model of UAV Reconnaissance Image and Feature Visualization. *Appl. Sci.* **2022**, *12*, 12236. [[CrossRef](#)]
51. Zhao, Z.; Han, J.; Song, L. YOLO-Highway: An Improved Highway Center Marking Detection Model for Unmanned Aerial Vehicle Autonomous Flight. *Math. Probl. Eng.* **2021**, *2021*, e1205153. [[CrossRef](#)]
52. Tang, Y.; Zhou, H.; Wang, H.; Zhang, Y. Fruit Detection and Positioning Technology for a *Camellia oleifera* C. Abel Orchard Based on Improved YOLOv4-Tiny Model and Binocular Stereo Vision. *Expert Syst. Appl.* **2023**, *211*, 118573. [[CrossRef](#)]
53. Yue, L.; Shen, H.; Li, J.; Yuan, Q.; Zhang, H.; Zhang, L. Image Super-Resolution: The Techniques, Applications, and Future. *Signal Process.* **2016**, *128*, 389–408. [[CrossRef](#)]
54. Benjumea, A.; Teeti, I.; Cuzzolin, F.; Bradley, A. YOLO-Z: Improving Small Object Detection in YOLOv5 for Autonomous Vehicles. *arXiv* **2021**, arXiv:2112.11798.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.