

Weed species classification with UAV imagery and standard CNN models: Assessing the frontiers of training and inference phases

G.A. Mesías-Ruiz^{a,c}, I. Borra-Serrano^a, J.M. Peña^a, A.I. de Castro^b, C. Fernández-Quintanilla^a, J. Dorado^{a,*}

^a Institute of Agricultural Sciences, Spanish National Research Council (ICA-CSIC), C/Serrano 115b, Madrid, 28006, Madrid, Spain

^b Environment and Agronomy Department, National Agricultural and Food Research and Technology Institute (INIA-CSIC), Crta. Coruña Km. 7.5, Madrid, 28008, Madrid, Spain

^c School of Agricultural, Food and Biosystems Engineering (ETSIAAB), Polytechnic University of Madrid (UPM), Av. Puerta de Hierro, 2, Madrid, 28040, Madrid, Spain

ARTICLE INFO

Keywords:

Site-specific weed management (SSWM)
Deep learning
VGG16
ResNet152
Inception-ResNet-v2
Faster R-CNN

ABSTRACT

Accurate weed species identification is crucial for effective site-specific weed management (SSWM), enabling targeted and timely control measures for each weed in crop field. This study advanced the current approach to species-level weed identification during the early growth stage by integrating unmanned aerial vehicles (UAVs) imagery with standard convolutional neural networks (CNNs) models such as VGG16, Resnet152 and Inception-Resnet-v2. For this, a robust dataset was created with 33,467 labels of weeds (*Atriplex patula*, *Chenopodium album*, *Convolvulus arvensis*, *Cyperus rotundus*, *Lolium rigidum*, *Portulaca oleracea*, *Salsola kali*, *Solanum nigrum*) and crops (maize, tomato), which was subjected to different training, validation and test scenarios. Model inputs were adjusted in order to align them with the information represented by the UAV images. Initially, models were developed in balanced scenarios, gradually increasing label numbers to assess their performance. Inception-ResNet-v2 achieved over 90% accuracy with 400 labels, while ResNet152 and VGG16 required 600 and 800 labels, respectively, for similar accuracy. In a more complex and realistic scenarios with unbalanced datasets, Inception-ResNet-v2 outperformed, likely due to its deeper architecture and enhanced capability to capture intricate features and patterns within UAV images. The study emphasized the importance of the minority-to-majority species ratio in unbalanced datasets, which affects minority species classification. To prevent misclassification, it is crucial to determine the right number of labels for CNN model training and validation. Weed maps were generated after species classification using the Faster R-CNN algorithm as an object detector. This advancement in methodology facilitates the precise and efficient implementation of SSWM techniques.

1. Introduction

Site-specific weed management (SSWM) is an agronomic strategy that focuses on the precise location, identification and targeted management of weed populations in specific areas, rather than treating the entire field uniformly (Fernández-Quintanilla et al., 2018; Lati et al., 2021). Accurate and quick identification of weed species in an early-stage is one of the key components of SSWM, since different species require distinct control measures (Sa et al., 2018; Fernandez-Quintanilla et al., 2022). Weed species identification is often a challenging and time-consuming process that demands skilled personnel. Since, many weed species possess similar physical

characteristics, incorrect identification and treatment decisions may be made, leading to a decrease in control efficacy.

The digitization of agriculture is a contemporary approach that employs technology to enhance efficiency and productivity. This approach integrates various technologies, such as sensors, unmanned aerial vehicles (UAVs), Artificial Intelligence (AI), and big data analysis, to enhance crop yields while minimizing the environmental impact of crop production through a reduction in pesticide use. UAVs have great potential in agriculture owing to their capacity to rapidly and effectively gather large amounts of highly detailed imagery over crops and fields, thus improving spatial and temporal data that facilitates crop scouting, particularly the detection of weeds, pests and diseases (Rejeb et al.,

* Corresponding author.

E-mail addresses: gmesias@ica.csic.es (G.A. Mesías-Ruiz), irene.borra@ica.csic.es (I. Borra-Serrano), jmpena@ica.csic.es (J.M. Peña), ana.decastro@csic.es (A.I. de Castro), cesar@ica.csic.es (C. Fernández-Quintanilla), jose.dorado@csic.es (J. Dorado).

<https://doi.org/10.1016/j.cropro.2024.106721>

Received 11 September 2023; Received in revised form 9 April 2024; Accepted 1 May 2024

Available online 11 May 2024

0261-2194/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2022). Deep Learning (DL), a subfield of Machine Learning (ML), has demonstrated particular success in image processing, improving the accuracy of classification tasks and reducing variability in regression problems (Mesías-Ruiz et al., 2023). Currently, DL techniques such as convolutional neural networks (CNNs) outperform traditional ML techniques in object recognition (Ioffe and Szegedy, 2015). CNNs are a type of algorithm inspired by the neural mechanisms underlying the visual system, making them well-suited for computer vision and image recognition tasks. CNNs have been particularly successful in these areas because they are specifically designed to handle high-dimensional input data, such as images, and are able to automatically learn relevant features from the data.

DL algorithms have been recently applied to UAV-based image analysis with some restrictions regarding the number of classes, i.e. aiming to generate a segmented database containing only general classes such as crop, soil, shadows, and two groups of grass and broadleaf weeds (dos Santos Ferreira et al., 2017; Torres-Sánchez et al., 2021). Indeed, dos Santos Ferreira et al. (2017) utilized the CaffeNet architecture, a derivative of AlexNet, to classify weeds among crops, achieving an average classification accuracy of 99.1% and 99.5% for balanced and unbalanced datasets, respectively, using a dataset of 15,336 segmented crop labels, including 3,249 soil samples, 7,376 soybean samples, 3,520 grass samples and 1,191 broadleaf weed samples captured at a height of 4 m. Bah et al. (2018) used a supervised dataset of UAV-captured images at 20 m height, consisting of 30,105 weed images (thistles and young shoots) and 46,736 crop images (beans and spinach), employing the ResNet18 architecture for discriminatory analysis, resulting in a notable area under the curve value of 0.957. Huang et al. (2020) focused on weed identification, classification and mapping in two rice fields mainly infested by *Leptochloa chinensis*, *Cyperus iria*, *Digitaria sanguinalis* and *Echinochloa crus-galli*, using a UAV dataset captured at 10 m height, employing four pre-trained CNNs (AlexNet, VGG-16, GoogLeNet, and ResNet-101) as classifiers. The study demonstrated accuracy performances of 86.8%, 88.4%, 87.2%, and 86.6%, respectively, for discriminating between weeds and rice plants. Khan et al. (2021) devised an optimized semi-supervised framework using generative adversarial networks to classify weeds and crops, utilizing a dataset of 2,800 weed images and 2,600 pea and strawberry crop images captured at a flight altitude of 2 m, achieving classification accuracy nearing 90%.

Some advancements have been achieved in the classification of weed species. For instance, Chen et al. (2022) employed transfer learning with 27 DL models on a dataset of 5,187 images representing 15 distinct weed species at various growth stages, with the ResNet101 model demonstrating superior performance in weed species identification by achieving the highest *F1-score* of 99.1%. Precise identification of weed species is fundamental to improve control techniques, as evidenced in the work of Valente et al. (2022), where they emphasize the importance of accurately mapping of *Rumex obtusifolius* for implementing effective control strategies. By utilizing images captured at a 10 m altitude, they achieved an *F1-score* of 78.4% using the pre-trained MobileNet model. Additionally, the integration of diverse CNN architectures has enabled significant advances in solving various computer vision tasks. An example is the study by Shahi et al. (2023), where segmentation models such as SegNet, DeepLabV3+ and UNet, were combined with classification architectures such as VGG16, ResNet50, DenseNet121, EfficientNetB0 and MobileNetV2. The most noteworthy outcome was an *F1-score* of 88.2%, achieved through the combination of EfficientNetB0 and UNet. The study successfully automated the semantic segmentation of weeds like *Sorghum halepense*, *Convolvulus arvensis* and *Portulaca oleracea* within a cotton crop, utilizing a dataset (Krestenitis et al., 2022) acquired by flights at 5 m height.

After reviewing the literature and evaluating the performance of various models in image classification tasks, the present study focused on three standard CNN architectures, selected based on their performance and wide adoption in the scientific community: VGG16 (Simonyan and Zisserman, 2014), ResNet152 (He et al., 2016), and

Inception-ResNet-v2 (Szegedy et al. (2017)). These architectures offer a diverse set of features and depths, enabling a comprehensive evaluation of their performance.

The main objective of this research was to investigate the constraints and limitations associated with the training and testing phases when developing a methodology for classifying weed species in early-stage field conditions using CNN architectures and UAV imagery. Indeed, the development of an automated and accurate method to detect and classify weed species at an early growth stage can support SSWM systems by enabling specific crop protection measures, reducing crop management costs, and minimizing herbicide-related environmental impact. In particular, the study addressed four specific objectives: 1) Determine the minimum number of labels, i.e. optimal dataset size, required to achieve an acceptable CNN model performance with an accuracy, precision and recall value above 90%. 2) Evaluate the performance, efficiency and suitability of three CNN models in classifying weed species using a balanced dataset. 3) Examine the impact of an unbalanced dataset, specifically focusing on scenarios where a) there is a higher number of crop labels, and b) there are varying numbers of labels for both crop and weed species, on the accuracy of weed species classification and the overall performance of the CNN models. And 4) Implement an object detector integrated with the most effective classification model to assess its performance in accurately locating and identifying weed species for mapping purposes within a field. This research contributes to the progress of CNN models for UAV image classification and detection of weed species, offering valuable insights for future studies in this field.

2. Materials and methods

Fig. 1 provides an overview of the methodology to achieve the proposed objectives. Firstly, UAV flight plans were designed to allow automated flights, ensuring a consistent approach to image capture. Secondly, the preprocessing stage involved the creation of orthomosaics, image splitting and species labeling. Thirdly, three CNN-based models were trained using both the training and validation datasets to perform the classification task. Subsequently, the generalization of knowledge to new data was assessed using widely-used classification metrics. Finally, the best classification model was integrated with an object detection architecture to automatically identify weed species and locate them in the orthomosaics.

2.1. Acquisition of UAV images

Two different locations were used in this study. On the one side, UAV images of an irrigated maize crop were acquired at the CSIC experimental farm La Poveda, Arganda del Rey, Madrid, Spain (40°18'59.25"N 3°29'21.53"W), at an early stage of crop growth (4 leaves unfolded, BBCH 14 (Meier, 2018)). The area has a Mediterranean Continental climate characterized by cold winters, hot summers, and an average yearly rainfall of around 400 mm. The soil type is sandy loam, with a plow layer containing 1.6% organic matter. An area of around 0.75 ha was sampled, which consisted of conventionally tilled maize. The maize was grown with rows spaced 0.75 m apart and a density of 85,000 plants per hectare. On the other side, UAV images an irrigated tomato crop were acquired from a commercial farm located in Santa Amalia, Badajoz, Spain (38°59'39.0"N 6°03'31.7"W), also at an early stage of the crop growth (first flower bud visible, BBCH 501 (Meier, 2018)). Soil in this farm is predominantly sandy and has low organic matter content. Approximately 1.2 ha were surveyed, comprising conventionally tilled tomato plants. The tomato crop was cultivated with 1.5-m row spacing and a plant density of 25,000 per hectare.

A low-cost RGB camera, model Sony ILCE-6300L (Sony Group Corporation, Tokyo, Japan), was mounted on the quadcopter UAV, model md4-1000 (microdrones GmbH, Siegen, Germany) for image acquisition. The camera featured an APS-C type Exmor[®] CMOS sensor (23.5 ×

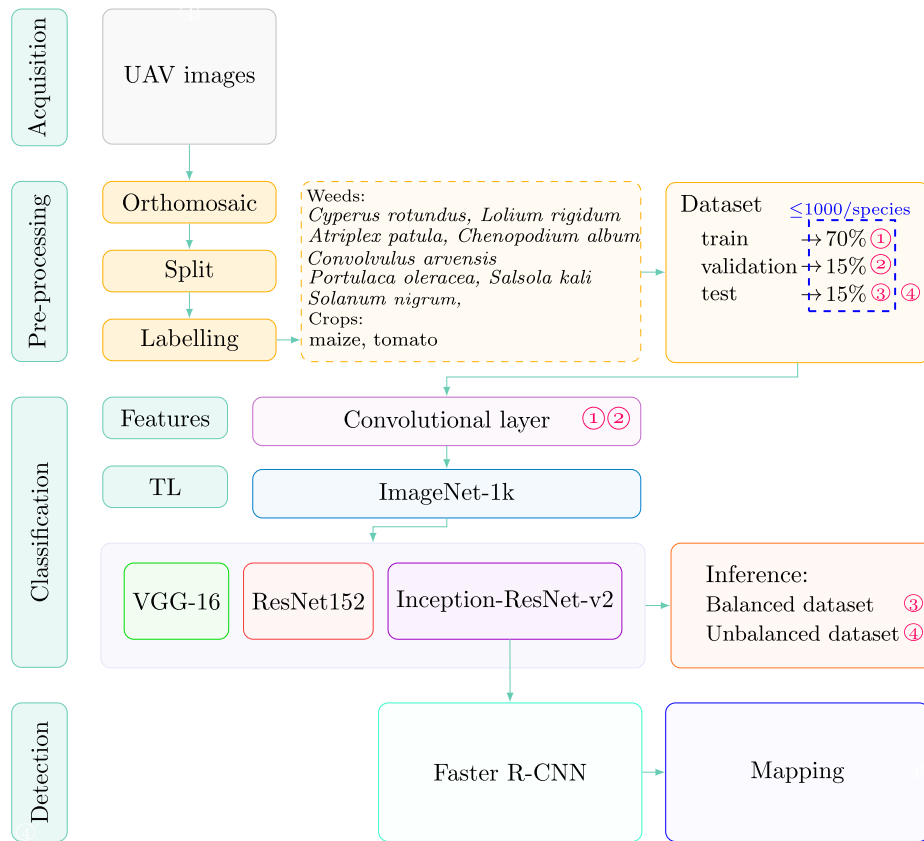


Fig. 1. Outline of the proposed methodology for the classification of weed species in early growth stage of maize and tomato crops.

15.6 mm) with 24.2 effective megapixels. The UAV flight plan parameters were set using the mdCockpit application (microdrones GmbH, Siegen, Germany). The configuration specified a constant horizontal speed of 2 m/s at a fixed altitude of 11 m above ground level, resulting in a ground sampling distance of 0.17 cm per pixel. The acquired images had an overlap ratio of 70% laterally and 70% frontally, with each image measuring $6,000 \times 3,376$ pixels.

2.2. Preprocessing

2.2.1. Orthomosaic

Using the acquired imagery together with geospatial data extracted from strategically placed control points in the experimental fields (seven and six for the maize and tomato, respectively), we systematically generated geomatics products with Agisoft PhotoScan software (Agisoft LLC, St. Petersburg, Russia). These products enabled the creation of orthomosaics, incorporating the information extracted from the RGB channels. Consequently, the generation of an orthomosaic involved stitching together multiple overlapping UAV images to produce a single high-resolution image that covers the entire area of interest.

2.2.2. Split

The orthophotos were subdivided into smaller sections to enhance processing efficiency and reduce computational complexity in subsequent analysis. This subdivision was prompted by the substantial size of the orthomosaics, which size were $61,000 \times 41,175$ pixels (5.3 GB) for maize and $72,304 \times 34,574$ pixels (6.3 GB) for tomato, making them unwieldy for direct processing. This procedure was executed by implementing an algorithm that automatically partitioned the orthophoto into split images of $1,000 \times 1,000$ pixels. The algorithm was developed in Python, using the necessary libraries, such as rasterio for geospatial data management and PIL (Pillow) for image manipulation. After accessing

the orthomosaics, the geospatial coordinates of each upper left corner were obtained. Iteration was performed on the partitions defined on the x and y axes. Subsequently, an image was generated and saved as a JPG file. This partitioning process facilitated the recognition and labelling of distinct species.

2.2.3. Labelling

Experts labeled the weed species by manually drawing bounding boxes and annotating several visible objects each split image. This procedure was carried out with the free software graphical tool labelling (Tzutalin, 2015). The following weed species were labeled to compile the dataset: 1) *Solanum nigrum* (4,175 labels); 2) *Cyperus rotundus* (3, 500); 3) *Portulaca oleracea* (1,900); 4) *Chenopodium album* (1,200); 5) *Convolvulus arvensis* (1,200); 6) *Salsola kali* (1,200); 7) *Atriplex patula* (1, 017); and 8) *Lolium rigidum* (1,014). In addition, 12,364 maize plants and 6,622 tomato plants were labeled. The annotations made on the split images were saved with XML extension in PASCAL VOC format. Dataset generation included the following steps: 1) reading the XML file coordinates that generate a square delimiting the weed; 2) extraction of the weeds on the split images; 3) assignment of a distinct identifier to designate its corresponding split area; 4) classifying the weed species according to their labels; and 5) storing them in separate directories based on their respective labels. The final result is a collection of individual images corresponding to each label, i.e. each plant of every species. This generated dataset is available at (Mesías-Ruiz et al., 2024).

2.2.4. Dataset

DL models are highly sensitive to quality and consistency of the source dataset, which should be diverse, representative, and in sufficient quantity to ensure robust and accurate performance. Since there is no prior research on the optimal size of training, validation, and test subsets for developing a DL model, we addressed this by dividing the dataset

into multiple subsets to test the models across various scenarios. Initially, we relied on the number of labels for the least frequent weed species in the source dataset, namely *Lolium rigidum* and *Atriplex patula*, each with slightly more than 1,000 labels. Several balanced datasets of different sizes were created: 100, 200, 400, 600, 800 and 1,000 labels per weed species. This strategy was implemented to avoid bias towards the majority species. A constant distribution of 70% for training, 15% for validation and 15% for test was maintained for the implementation of each selected classification model.

Additionally, two tests (Tests I and II) were conducted using unbalanced datasets (Table 1) to evaluate the generalization ability of the models in scenarios that better reflect real-world conditions. These tests were performed on the models developed using a balanced dataset, where the maximum number of labels per weed species (i.e. 700 for training and 150 for validation) was utilized. Test I was designed to simulate a scenario where crop labels dominate the dataset. To conduct this analysis, the relationship between the number of crop and weed labels was analyzed by adjusting the number of crop labels at various ratios: 1:1, 1:2, 1:10, 1:20, and 1:30. This enabled us to achieve a range of crop labels from 150 to 4,500. Alternatively, Test II aimed to assess the models' ability to generalize on an imbalanced dataset with varying numbers of labels for each crop and weed species. Based on the results of Tests I and II, the optimal correlation between species imbalance was identified to achieve classification metrics that align with our objective. This involved evaluating the frontiers at which a decrease in model performance occurred as a result of the number of species labels. In this study, separate datasets were analyzed for the maize-weed and tomato-weed cropping systems.

2.3. Classification

2.3.1. Models

The three models used for classification task, namely VGG16, ResNet152, and Inception-ResNet-v2, have unique architectures characterized by varying numbers of layers and connectivity patterns. This study explored the key features of these models, evaluating their strengths and limitations, and the factors that differentiate them from other CNN architectures.

The VGG16 model is known for its architectural simplicity and performance improvements due to the implementation of the VGG block, which increases the depth of CNNs. Specifically, the VGG16 model comprised of 16 convolution layers with a 3×3 kernel convolution, five Max pooling layers with a 2×2 size, and three fully connected layers after the last Max pooling layer. ReLu activation was used in all layers, and the final layer employed a softmax classifier.

The ResNet152 model is an architecture that introduces residual connections to facilitate the training of very deep networks. By adding layers to the network, its complexity and expressiveness are increased. The introduction of the residual block in ResNet152 addresses a common issue encountered in other architectures, where shallower

convolutional networks tend to perform better than their deeper counterparts. This is due to problems such as vanishing gradients and the curse of dimensionality, where increasing the number of layers may not always improve performance. The residual block overcomes this challenge by enabling information to skip certain layers and directly pass to the next layer through a residual connection, thereby improving the performance of the network with deeper layers (He et al., 2016).

The Inception-ResNet-v2 model was created by experimenting with different convolution kernel sizes ranging from 1×1 to 11×11 in order to achieve optimal performance. Through this process, it was discovered that utilizing a combination of kernels of varying sizes was most effective, leading to the development of the Inception block. Compared to other blocks, the Inception block is more computationally efficient due to its reduced parameter count, while also generating more expressive features by combining features produced by kernels of differing sizes (Szegedy et al., 2015).

2.3.2. Training, validation and test

Training, validation and test of the CNN models were submitted to the Keras-Tensorflow DL framework (Abadi et al., 2016). The Ubuntu 20.04.4 LTS operating system and Python 3.9 were used on a workstation with an Intel® Core (TM) i7-12700K CPU, a NVIDIA GeForce RTX 3070 Ti Graphic Processing Unit (GPU) with 8 GB of memory. To optimize the classifiers parameters, the expected input of the models was modified by adjusting it to the average size of our label set. For the VGG16 and ResNet152 models an image size of 64×64 pixels was used, while for Inception-ResNet-v2 the image size was 75×75 pixels. After loading the models, the transfer learning technique (Pan and Yang, 2010) was employed with the pre-trained ImageNet weights, replacing the last fully connected layer by a new fully connected layer with 9 neurons. The hyper-parameters for all three models were identical, with a batch size of 32, 30 epochs, adaptive moment estimation (Adam) like as optimizer with a learning rate of 10^{-5} , and categorical cross-entropy serving as the loss function. Leveraging the advantages of using transfer learning in image pre-training, we hypothesized that having a sufficient number of images for each class (e.g., up to 1,000 images) within a balanced dataset could obviate the need for data augmentation (Shawky et al., 2020; Khalifa et al., 2021) in our study. Such an approach would facilitate our research objective of evaluating and comparing the performance of the CNN-based models while preserving the inherent quality of the original dataset derived from our UAV imagery. However, exploring data augmentation techniques is a potential area for future research. The optimal training accuracies for all case studies were achieved between 22 and 27 epochs.

2.3.3. Evaluation

The evaluation of the classification was carried out through a confusion matrix to better understand the performance of the models. The matrix includes four predictions categorized as follows: T_P = true positive, T_N = true negative, F_P = false positive, and F_N = false negative. The criteria of "type I error" (false positive) and "type II error" (false negative) are used to perform a detailed analysis on the performance of the models. Additionally, four different metrics commonly used in ML were derived from the four predictions of the confusion matrix.

- 1) **Accuracy:** determines the degree of approximation of the measurements of a quantity to the real value, it can lead to errors in the quality of the model when the class imbalance is high.

$$Accuracy = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \quad (1)$$

- 2) **Precision:** represents the positive prediction value, which measures the quality of the model.

Table 1

Distribution of labels of the unbalanced datasets: Test I with crop as the pre-dominant species, and Test II with a varying number of labels for each species.

Species	Unbalanced dataset	
	Test I	Test II
<i>Cyperus rotundus</i>	150	2,500
<i>Lolium rigidum</i>	150	14
<i>Atriplex patula</i>	150	17
<i>Chenopodium album</i>	150	200
<i>Convolvulus arvensis</i>	150	200
<i>Portulaca oleracea</i>	150	900
<i>Salsola kali</i>	150	200
<i>Solanum nigrum</i>	150	3,175
Maize	[150; 4,500]	11,364
Tomato	[150; 4,500]	5,622

$$\text{Precision} = \frac{T_p}{T_p + F_p} \quad (2)$$

3) *Recall*: measures the ability of the model to identify true positives.

$$\text{Recall} = \frac{T_p}{T_p + F_N} \quad (3)$$

4) *F1-score*: Calculates the harmonic mean between precision and recall, providing a balance between the two metrics. It is a general score that represents the model overall performance, indicating its ability to retrieve relevant results. *F1-score* is especially useful for unbalanced classes.

$$\text{F1-score} = \frac{2T_p}{2T_p + F_p + F_N} \quad (4)$$

2.4. Object detection

The detection and localization of weed species in the two fields was carried out by applying the Faster R-CNN algorithm as an object detector. This algorithm is composed of two fundamental modules: a fully convolutional network in charge of proposing regions of interest (RPN) within the image, and the Fast R-CNN detector that executes the detection of objects within these proposed regions (Ren et al., 2016). The implementation of the Faster R-CNN detector was carried out using the TensorFlow object detection API. The Adam optimizer was used, with an image input size of 640×640 pixels, and a minimum confidence threshold of 0.4 was set to validate the detections.

Model performance evaluation was performed individually for maize and tomato crops, using the images acquired in subsection 2.2.2, which were adjusted to the model input size. After obtaining the predictions and locations of the objects of interest, the local coordinates of the bounding boxes were extracted and aligned on the orthophoto. These translated coordinates were exported to a CSV file, which simplified the generation of weed maps and their visualization on the orthomosaic.

For the evaluation of maize, 30 frames of $1m^2$ with their corresponding georeferencing, randomly distributed in the field, were located on the orthophoto. These frames were used to carry out the field verification of the weed species detected. For the tomato crop, the model results were compared with approximately 200 plants of each weed species (i.e. *Cyperus rotundus*, *Portulaca oleracea* and *Solanum nigrum*), previously labeled on the orthophoto by experts. The percentage

detected and correctly classified was calculated. These approaches allowed a rigorous evaluation of the discriminative capacity of DL architectures applied to agricultural fields.

3. Results

3.1. Inference of the CNN models for classification as a function of dataset size

As expected, performance of the studied CNN models improved as the number of labels increased, with the highest accuracy obtained with 1,000 labels for each weed species in both crops (Fig. 2). Inception-ResNet-v2 achieving the highest mean accuracy of 98.6% for the both crops types, while ResNet152 and VGG16 had mean accuracies of 95.6% and 92.6%, respectively. These models achieved high accuracy with minimal variance (less than $\pm 0.13\%$) in error means. An acceptable threshold over 80% accuracy was achieved with only 200 labels, although the three CNN models exceeded 90% accuracy with 800 labels per weed species. Inception-ResNet-v2 consistently outperformed the other models, achieving over 90% accuracy with just 400 labels, while ResNet152 and VGG16 required 600 and 800 labels, respectively, to reach similar accuracy.

Fig. 2 illustrates that the relationship between the increase in labels and the improvement in model accuracy was not linear. When the number of labels was increased from 100 to 200, there was a larger relative variation in accuracy compared to increasing labels from 400 and above. In the maize crop, the VGG16, ResNet152, and Inception-ResNet-v2 models showed accuracy improvements of 8.6%, 10.8%, and 13.7% respectively, when the number of labels increased from 100 to 200. Similarly, in the tomato crop, the corresponding accuracy gains were 6.9%, 9.5%, and 13.7%. However, the Inception-ResNet-v2 model approached an asymptote, as the relative variation between 800 and 1,000 labels did not exceed 0.4% for both maize and tomato crops.

3.2. Inference of the CNN models with balanced datasets

Fig. 3 shows the analysis of the performance of the three models for each weed species using a balanced dataset with 150 labels of each species. In maize crop, the precision consistently resulted in a mean above 86%. Overall, Inception-ResNet-v2 achieved the highest precision values for overall species, whereas VGG16 had the lowest values. The highest mean precision was achieved for *Portulaca oleracea* (99.9% with Inception-ResNet-v2; and 95.7% with VGG16) and for *Cyperus rotundus*

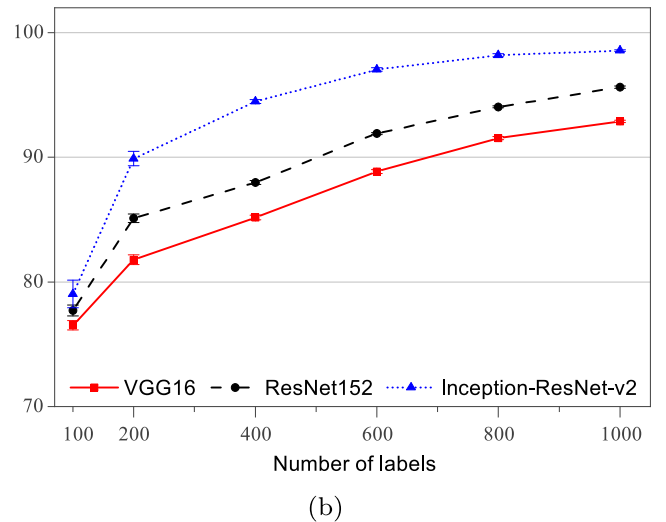
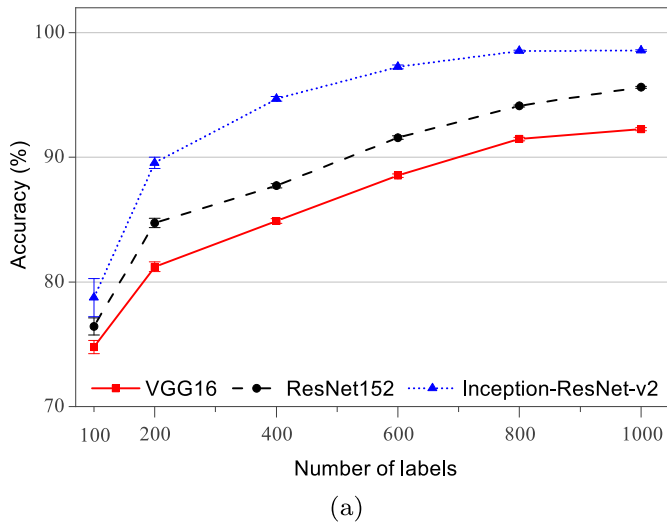


Fig. 2. Relationship between the number of labels and model accuracy in the balanced data set for the VGG16, ResNet152 and Inception-ResNet-v2 models in two different crops: (a) Maize crop, and (b) Tomato crop.

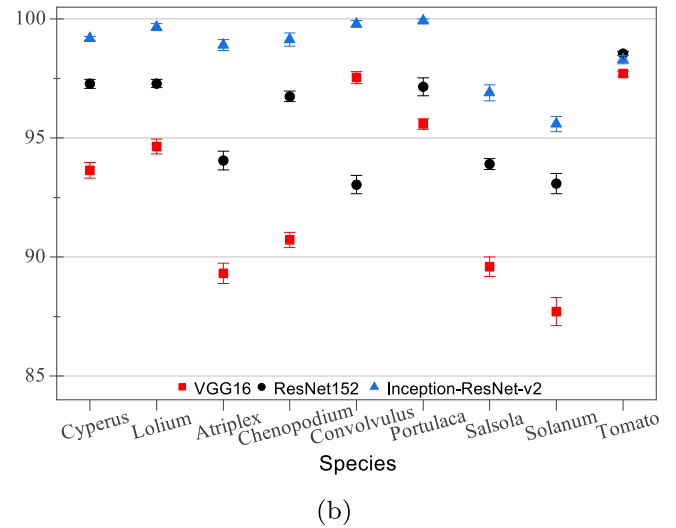
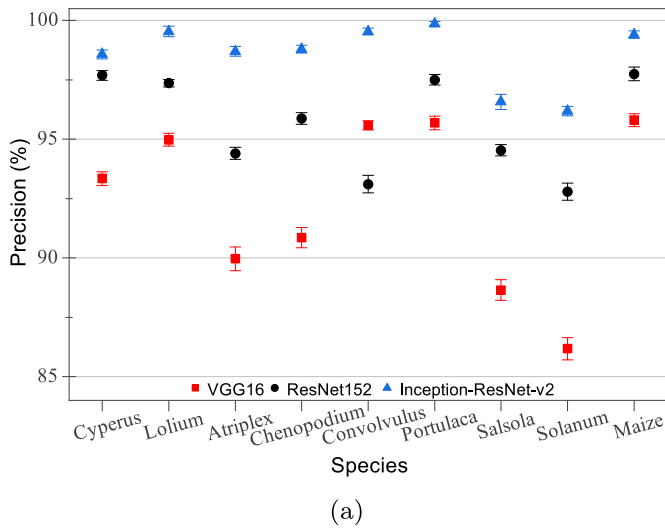


Fig. 3. Mean precision (bars indicating standard error among ten executions for each model) using the maximum number of labels (1000 for weed species) in the balanced dataset for VGG16, ResNet152 and Inception-ResNet-v2 models as affected by weed species in two different crops: (a) Maize crop, and (b) Tomato crop.

(97.7%) with ResNet152, while the lowest was observed for *Solanum nigrum* (86.2% with VGG16; 92.8% with ResNet152; and 96.2% with Inception-ResNet-v2).

In the case of tomato, the performance analysis of the three models

across different weed species consistently yielded a mean precision above 87%. Similar to maize crop, Inception-ResNet-v2 generally achieved the highest precision values, while VGG16 had the lowest values. The highest mean precision was attained for *Portulaca oleracea* (99.9%) with

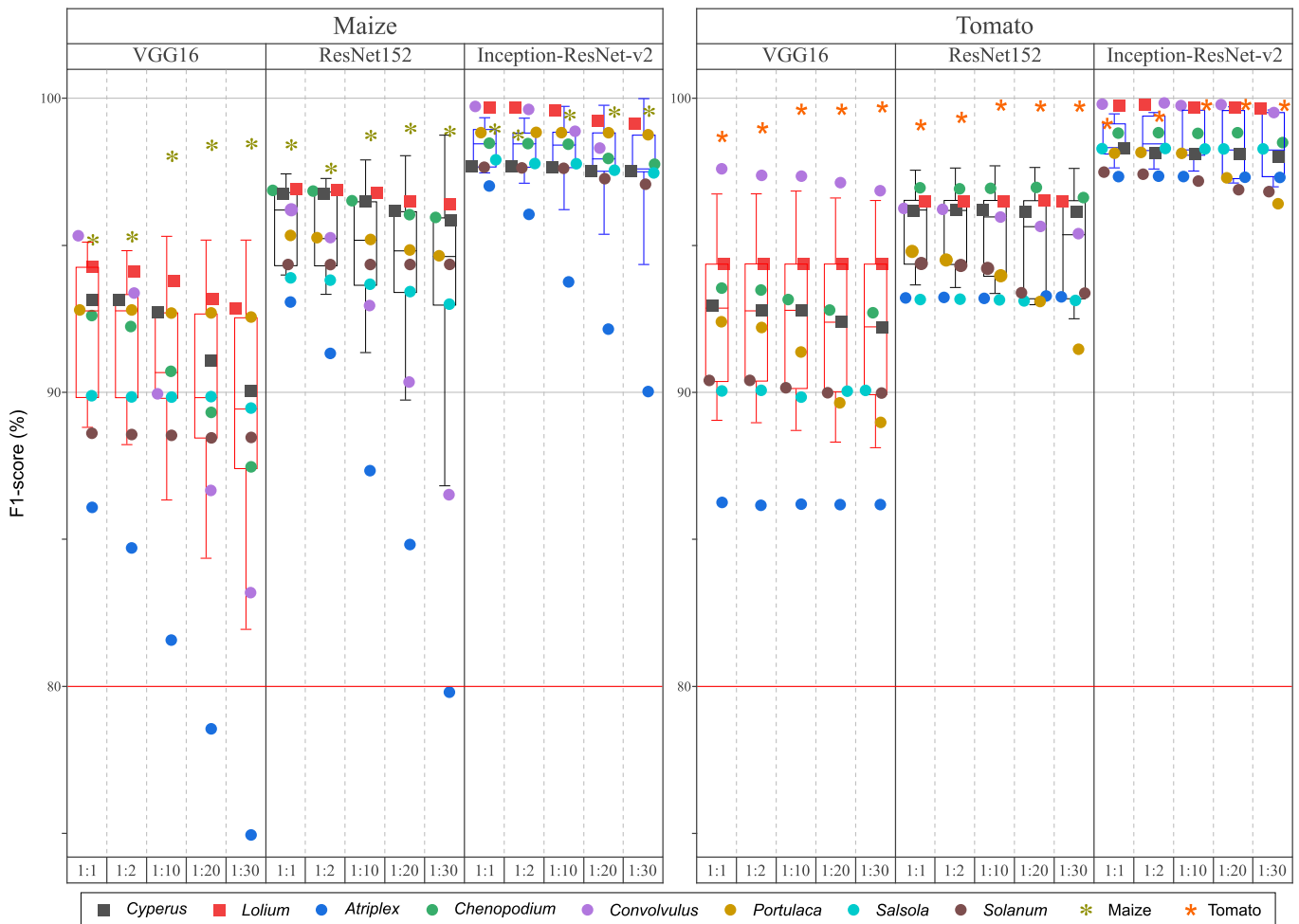


Fig. 4. Box (mean $F1$ -score) and whisker (± 1 SD) diagram showing Test I results across the range of weed-to-crop label ratios (from 1:1 to 1:30), for maize and tomato, using ten executions for the VGG16, ResNet152 and Inception-ResNet-v2 models.

Inception-ResNet-v2, *Convolvulus arvensis* (97.5%) with VGG16, and *Cyperus rotundus* and *Lolium rigidum* (97.3%) with ResNet152. The lowest mean precision was observed for *Solanum nigrum* (87.7%) with VGG16 and (95.7%) with Inception-ResNet-v2, as well as for *Convolvulus arvensis* (93.0%) with ResNet152.

3.3. Inference of CNN models with unbalanced dataset: crop as predominant class (Test I)

Test I was applied to analyze the performance of the different classification models in real-life scenarios, i.e. in which the source dataset is composed by an unbalanced number of labels with the crop as predominant species. Fig. 4 illustrates the evolution of the *F1-score* metric by weed species and the number of maize labels increasing in different ratios. In general, most weed species (e.g. *Cyperus rotundus*, *Lolium rigidum*, *Portulaca oleracea*, *Salsola kali* and *Solanum nigrum*) consistently displayed mean *F1-score* values above 86%, irrespective of the weed-to-crop label ratio. Nevertheless, certain weed species showed a declining *F1-score* value with an increasing number of crop labels. *Atriplex patula* consistently demonstrated a decrease in *F1-score* value across all three models, dropping below 80% in both VGG16 and ResNet152 models when a large number of maize labels were present. The decline was less pronounced but still noticeable in *Convolvulus arvensis* for these two models. In line with the declining trend observed in some weed species, the analysis of the confusion matrices in maize crop (Fig. 5) revealed a remarkable number of false positives in *Atriplex patula* ($F_p = 0.8\%$ and 0.9% for VGG16 and ResNet152, respectively) and *Convolvulus arvensis* ($F_p = 0.9\%$ and 0.6% for VGG16 and ResNet152, respectively). In addition, the percentage of false negatives generated between *Atriplex patula* and *Solanum nigrum* may have contributed to decrease the *F1-score*, with $F_N = 0.2\%$ and 0.1% for VGG16 and ResNet152, respectively.

Test I showed improved results in the three models for the tomato crop in comparison to the maize crop (Fig. 4), with consistent *F1-score* values observed across all weed species, regardless of the weed-to-crop label ratio. Remarkably, the Inception-ResNet-v2 and ResNet152 models consistently achieved *F1-score* values above 91% in the tomato crop. In contrast, the VGG16 model produced slightly lower values (though still above 86%). Among the weed species using the VGG16 model, only *Atriplex patula* displayed the lowest *F1-score* value. However, this species maintained a stable *F1-score* value above 86.2% across the entire range of weed-to-crop label ratios.

3.4. Inference of CNN models with unbalanced dataset: unequal number of labels for each species (Tets II)

Based on the *F1-score* values (Table 2), the results of Test II indicated

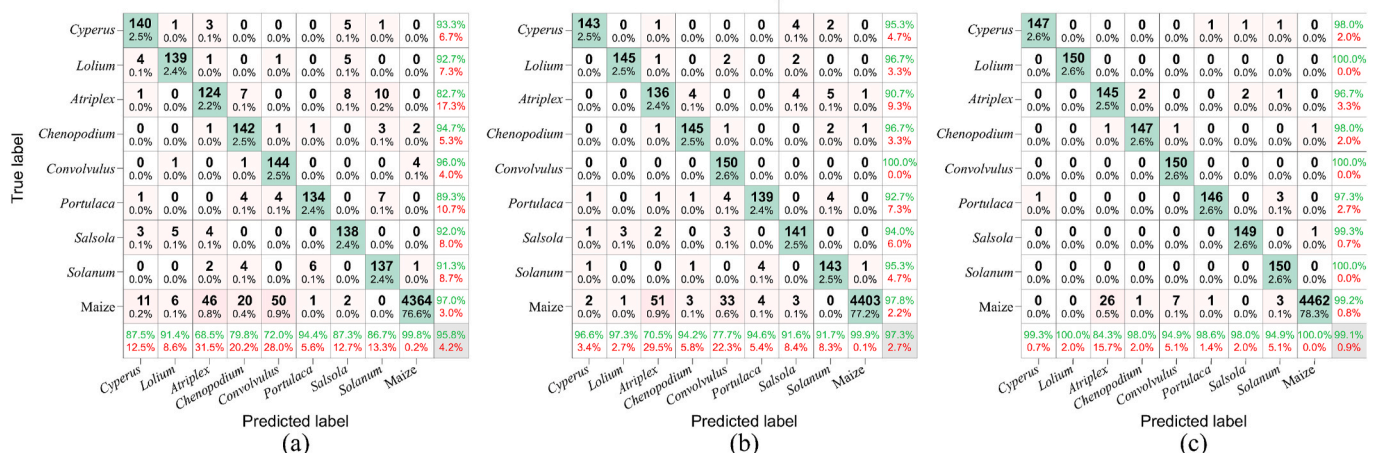


Fig. 5. Confusion matrices of CNN models using Test I on maize crop and individual weed species. (a) VGG16, (b) ResNet152, (c) Inception-ResNet-v2.

Table 2

F1-score metrics for three CNN models using Test II with an unbalanced dataset of different weed species and crops.

Species	Maize crop (%)				Tomato crop (%)			
	VGG16	ResNet152	Inception- ResNet-v2		VGG16	ResNet152	Inception- ResNet-v2	
<i>Cyperus rotundus</i>	95.17 ± 0.09	97.30 ± 0.05	98.43 ± 0.03		95.74 ± 0.06	97.53 ± 0.04	98.37 ± 0.04	
<i>Lolium rigidum</i>	25.46 ± 0.72	55.01 ± 1.19	58.53 ± 1.56		27.94 ± 0.47	64.36 ± 1.23	61.25 ± 1.74	
<i>Atriplex patula</i>	11.21 ± 0.28	13.57 ± 0.34	27.26 ± 1.06		20.80 ± 0.44	27.73 ± 0.53	54.61 ± 1.53	
<i>Chenopodium album</i>	61.22 ± 0.44	85.18 ± 0.26	94.28 ± 0.47		66.43 ± 0.51	85.98 ± 0.29	94.62 ± 0.43	
<i>Convolvulus arvensis</i>	74.56 ± 0.61	79.21 ± 0.58	93.18 ± 0.52		94.79 ± 0.27	95.26 ± 0.19	98.94 ± 0.22	
<i>Portulaca oleracea</i>	90.14 ± 0.14	93.30 ± 0.12	95.96 ± 0.14		88.77 ± 0.13	91.89 ± 0.13	95.25 ± 0.07	
<i>Salsola kali</i>	82.49 ± 0.46	83.45 ± 0.33	95.78 ± 0.23		83.70 ± 0.33	85.29 ± 0.25	95.54 ± 0.26	
<i>Solanum nigrum</i>	92.95 ± 0.10	96.09 ± 0.05	97.51 ± 0.10		92.85 ± 0.10	95.90 ± 0.05	97.67 ± 0.05	
Maize	98.44 ± 0.04	98.84 ± 0.04	99.50 ± 0.03					
Tomato					99.41 ± 0.01	99.49 ± 0.01	99.72 ± 0.01	

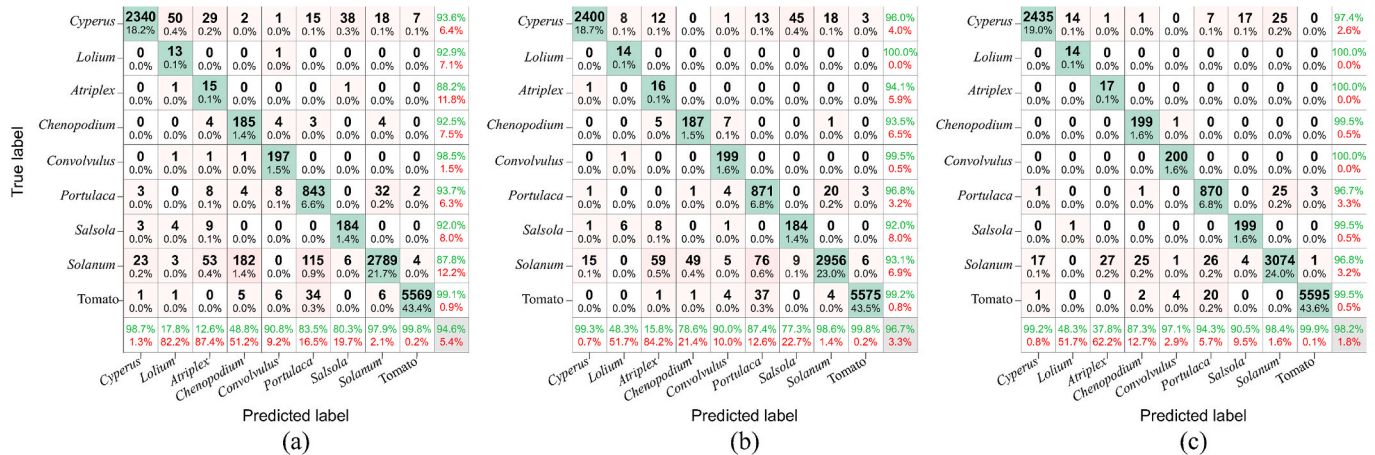


Fig. 6. Confusion matrices of CNN models using Test II on tomato crop and individual weed species. (a) VGG16, (b) ResNet152, (c) Inception-ResNet-v2.

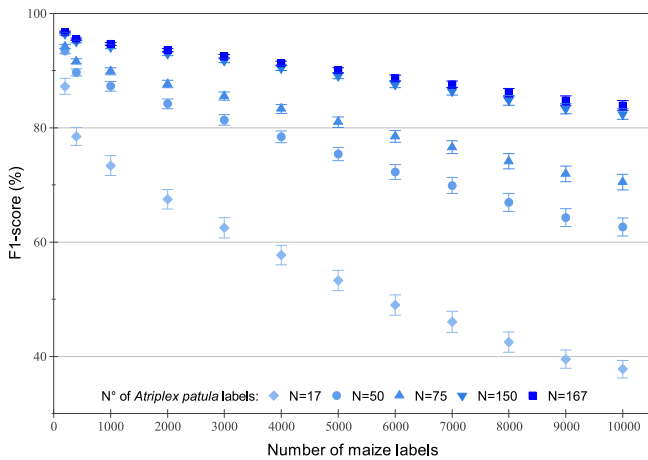


Fig. 7. Assessment of the optimal ratio of minority to majority species using *Atriplex patula* labels and maize labels (from $n = 17$ to $n = 150$) and maize labels (e.g. from 200 to 10,000) to achieve an *F1-score* above 80% with the Inception-ResNet-v2 model.

3.5. Detection and mapping of weed species

After integrating the RPN and other Faster R-CNN modules with the output of the Inception-ResNet-v2 feature extractor, the inference process was performed on the dataset of real images corresponding to each field (subsection 2.2.2). The inference times resulted in 120.7 s for the maize field and 141.6 s for the tomato field. During this period, the algorithm detected 70,134 individual plants in the maize field (including *Atriplex patula* 676, *Chenopodium album* 8,373, *Convolvulus arvensis*

8,881, *Lolium rigidum* 71, *Salsola kali* 2,293, and maize 49,840), and 14,377 individual plants in the tomato field (*Cyperus rotundus* 9,248, *Portulaca oleracea* 3,164, and *Solanum nigrum* 1,965). Upon acquiring the predictions and weed locations in each individual image, these co-ordinates were subsequently transferred to the orthophoto coordinates, enabling the generation of weed maps and their visualization in the orthomosaics (Fig. 8).

Table 3 shows the evaluation of the detection and classification results for various species provided by the architecture integrating Faster R-CNN and Inception-ResNet-v2, contrasting with ground-truth data in both maize and tomato crops. The ground-truth data consisted of a total of 970 labels, distributed between 356 and 614 for maize and tomato crops, respectively. The model successfully detected and classified 794 labels (237 for maize and 557 for tomato crops), with only eight instances of misclassification. Furthermore, *Cyperus rotundus* and *Portulaca oleracea* demonstrate outstanding performance, with *F1-score* values above 99%. Conversely, species like *Chenopodium album* and *Solanum nigrum* exhibit high accuracy but lower recall, indicating proper classification when detected, albeit with some instances missed. Notably, *Salsola kali* displays comparatively lower overall performance, with an *F1-score* of 36.4%.

Fig. 9 illustrates cases of evaluation of the model's discrimination performance. This analysis involved comparing the output images generated by our Faster R-CNN Inception-ResNet-v2 model with real images of maize and tomato fields.

4. Discussion

The contribution of this study allows the detection, classification and location of weeds at the species level in an early growth stage, which enables tailored management and control strategies per species. This is

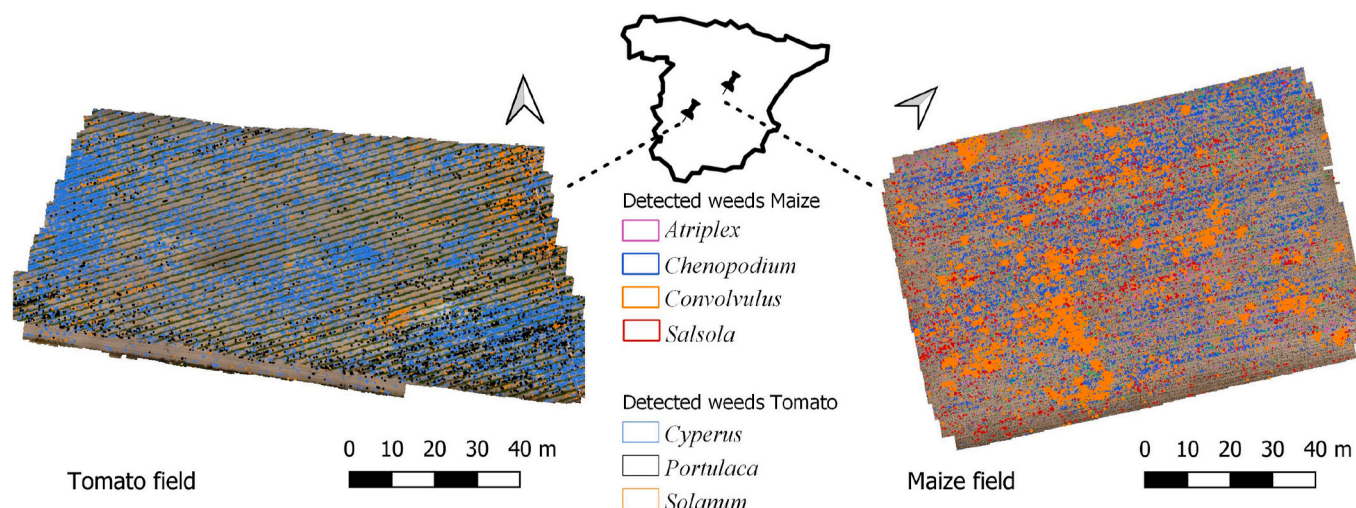


Fig. 8. Weed maps created using the integrated architecture of Faster R-CNN and Inception-ResNet-v2 for the tomato and maize fields.

particularly important because significant morphological and physiological variations exist within each weed species and they required different treatments. Identifying weed species at an early-stage using DL in UAV imagery poses a significant challenge (Rai et al., 2023). The development of an automated method for detecting and classifying weed species entails constraints to be solved and limitations, notably during training and testing phases. In this sense, we assessed the minimum dataset size required to attain 90% accuracy. The results revealed that increasing the number of labels in the training and validation phase enhances the accuracy of CNN-based models in the classification task.

In our study, we found that using a UAV to capture images at an altitude of 11 m resulted in orthophotos with a pixel size of 0.17 cm, which is a higher altitude and lower spatial resolution compared to similar research. For instance, (dos Santos Ferreira et al., 2017; Shahi et al., 2023) conducted weed classification research using UAV imagery at altitudes of 4 and 5 m, respectively, achieving relatively higher spatial resolutions of 0.11 and 0.14 cm/pixel. Consequently, extracting useful features from our images poses a greater challenge. To mitigate this issue, one approach could involve training a more complex model capable of learning from smaller images. However, such a solution entails longer training time, increased consumption of computing resources, and hence increased energy expenditure. The CNN models selected for this study, exhibited varying levels of performance. Specifically, in balanced datasets, the minimum dataset size required to train CNN models and achieve accuracies exceeding 90% ranged from just 400 labels for Inception-ResNet-v2 to 800 labels for VGG16, with intermediate label requirements for the ResNet152 model.

Due to the robustness of our dataset, we were able to train the CNN models with equal numbers of labels for each species, thereby preventing any bias towards the dominant class (Liu et al., 2022). The CNN models, VGG16, ResNet152 and Inception-ResNet-v2, achieved noteworthy classification accuracies of 92.6%, 95.6% and 98.6%, respectively, when evaluated using models created with the maximum number of balanced labels per species. These results demonstrate the ability of CNN models to generalize knowledge by adapting to new data from natural environments at early growth stages, showing promising potential for implementing SSWM systems at an early stage of crop growth.

Previous research has effectively integrated remote sensing methods with artificial neural networks to differentiate weed groups, such as monocotyledons versus dicotyledons (Torres-Sánchez et al., 2021). Employing a similar methodology, dos Santos Ferreira et al. (2017) utilized CaffeNet model on their segmented database, assessing its performance against traditional ML classifiers (e.g. Support Vector Machines, AdaBoost, and Random Forests). The evaluation involved feature

extraction techniques such as shape, color and texture. Their results showed an accuracy of over 98% in classifying all classes using the CNN model, thus concluding that the CNN model did not rely on the choice of feature extractor, unlike other models. Similar to prior studies (Hasan et al., 2021), our research aimed to take a step further by demonstrating the capability of CNN architectures in discriminating between multiple weed species in crop fields. We utilized a UAV imagery-based dataset, which proved robust to training across various CNN architectures through transfer learning techniques. The results were remarkable in accurately classifying all the weed species present in the two crop fields under investigation.

Although we acknowledge the limitations of our evaluation dataset in representing the full extent of field conditions, the analysis performed with Test II considering an unbalanced dataset provided results that closely approximate real-life scenarios. Under these more realistic conditions, the Inception-ResNet-v2 model produced the best results among the applied metrics, likely due to its greater depth and ability to capture more complex features and patterns in the UAV imagery. Indeed, the Inception-ResNet-v2 model achieved a performance (mean *F1-score*) ranging from 27.3% to 99.7%, while the ResNet152 model exhibited a range of 13.6%–99.5%, and the VGG16 model showed a range of 11.2%–99.4%. It is worth noting that the low values in the three CNNs corresponded to the results for the minority weed species, *Atriplex patula*. This is likely attributed to its confusion with another weed, *Solanum nigrum*, as indicated by the high percentages of false negatives observed in the confusion matrices, irrespective of the crop. When calculating the mean value of the whole species, Inception-ResNet-v2 model showed a performance (mean *F1-score*) above 86%, ResNet152 nearly 80% and VGG16 nearly 74%. Based on these results, the Inception-ResNet-v2 architecture proved to be the most effective model of weedy species, regardless of whether the dataset was balanced or unbalanced. In previous studies on individual species classification using UAV imagery, dos Santos Ferreira et al. (2019) reported a mean accuracy of 83.4% for four species using VGG16, while Chew et al. (2020) achieved *F1-score* ranging from 49% to 96% for six species. Considering the lower number of species classified in previous works, some authors have suggested that the performance of CNN-based models tends to deteriorate when training on multiple species (Dyrmann et al., 2016; Olsen et al., 2019), making our work interesting as our dataset includes ten species.

A significant contribution of this study was the finding that the accuracy of weed species classification relies on identifying the ratio between minority and majority species, as it can have a substantial impact on the efficacy of SSWM systems. Neglecting this ratio could result in

Table 3
Confusion matrix and metrics for the evaluation of classification and detection by species, derived from the comparison of model results and ground-truth results in maize and tomato crops.

Species	Cyperus rotundus	Lolium rigidum	Atriplex patula	Chenopodium album	Convolvulus arvensis	Portulaca oleracea	Salsola kali	Solanum nigrum	Maize	Tomato	Support	Precision (%)	Recall (%)	F1-score (%)	Detection (%)
Cyperus rotundus	199										203	100.0	98.0	99.0	98.0
Lolium rigidum [†]		-									-	-	-	-	-
Atriplex patula			2								2	66.7	100.0	80.0	100.0
Chenopodium album				52	2						125	96.3	41.6	58.1	43.2
Convolvulus arvensis			1	2	33		1				49	94.3	67.3	78.6	75.5
Portulaca oleracea						204					205	99.0	99.5	99.3	99.5
Salsola kali							2				8	66.7	25.0	36.4	25.0
Solanum nigrum						2		154			206	100.0	74.8	85.6	75.7
Maize									140		172	100.0	81.4	89.7	81.4
Tomato *										-	-	-	-	-	-
Total											970				

[†] *Lolium rigidum* was not found in the true ground frames.
The tomato category was excluded from evaluation due to its nature as a continuous row crop.

errors in decision-making and unintended consequences that could affect productivity weed control, hence crop productivity. Based on our results, in order to ensure an *F1-score* greater than 80% for each species using the models VGG16, ResNet152 and Inception-ResNet-v2 on an unbalanced dataset, a ratio of 1:12, 1:24, and 1:107 between the majority and minority species, respectively, should be maintained. Often, control methods are tailored to target a specific species. However, if the relationship between the target species and the other species for which a CNN-based identification system can accurately identify is not considered, there is a possibility of misapplying control methods to the crop.

This article introduces inference time data, 120.7 s for maize and 141.6 s for tomato. This information not only provides insights into the model's computational efficiency (which clearly depends on the hardware used and the complexity of the orthomosaics), but also offers a general insight of how quickly the model can analyze and process large datasets. It is important to note that the efficiency demonstrated by the model is not comparable to that of human experts who could manually undertake the same task. Regarding individual plant detection, the findings demonstrate notable efficacy. For maize, the algorithm successfully identified a total of 70,134 individual plants, providing a detailed distribution between different weed species and the maize crop itself. Likewise, for tomato, the model detected 14,377 individual plants underscoring its capability to discriminate both the primary crop and the weed population within the field. Indeed, the evaluation of prediction accuracy by Faster R-CNN Inception-ResNet-v2 model against ground-truth data revealed generally satisfactory species detection rates, albeit with a few exceptions, such as *Chenopodium album* in maize and *Solanum nigrum* in tomato fields. The lower detection rate for these species could possibly be attributed to the small size of their seedlings during the early stages, typically less than 5 cm². Addressing this challenge might necessitate higher resolution UAV imagery, enabling more accurate identification and classification.

In summary, by employing 850 labels for each species in the training and validation phases of the Inception-ResNet-v2 classifier, along with the Faster R-CNN object detector, our approach has successfully identified all weed species within the maize and tomato experiments. The precise identification of individual weed classes and crops has enabled the generation of detailed and customized maps, thereby facilitating accurate and site-specific weed management strategies.

5. Conclusions

This research represents a significant advance in early-growth stage weed species classification using UAV imagery and standard CNN models, such as VGG16, ResNet152 and Inception-ResNet-v2. We introduce a methodological innovation addressing the challenge of training a CNN model with images of varying sizes, deviating from its original design specifications and causing performance decline. To counteract this, we adapt the model inputs to align with the information conveyed by low spatial resolution images, providing an effective solution to mitigate model performance degradation in scenarios with image size variability. This approach offers valuable insights for the practical application of CNN in diverse environmental settings. In addition, we assessed the constraints and limitations related to the development of a weed species classification method during both the training and testing phases. Our findings reveal that, in balanced datasets, the minimum dataset size necessary to train CNN models and attain accuracies exceeding 90% varied from 400 to 800 labels, depending on the particular model used. Inception-ResNet-v2 consistently outperformed other models with accuracy values of 96% in both maize and tomato crops, likely due to its deep architecture and greater ability to capture intricate features and patterns within UAV images.

In complex and realistic scenarios with unbalanced datasets, Inception-ResNet-v2 also showed the best performance. The study revealed the critical role of the minority-to-majority species ratio on the classification accuracy, particularly affecting less represented species,

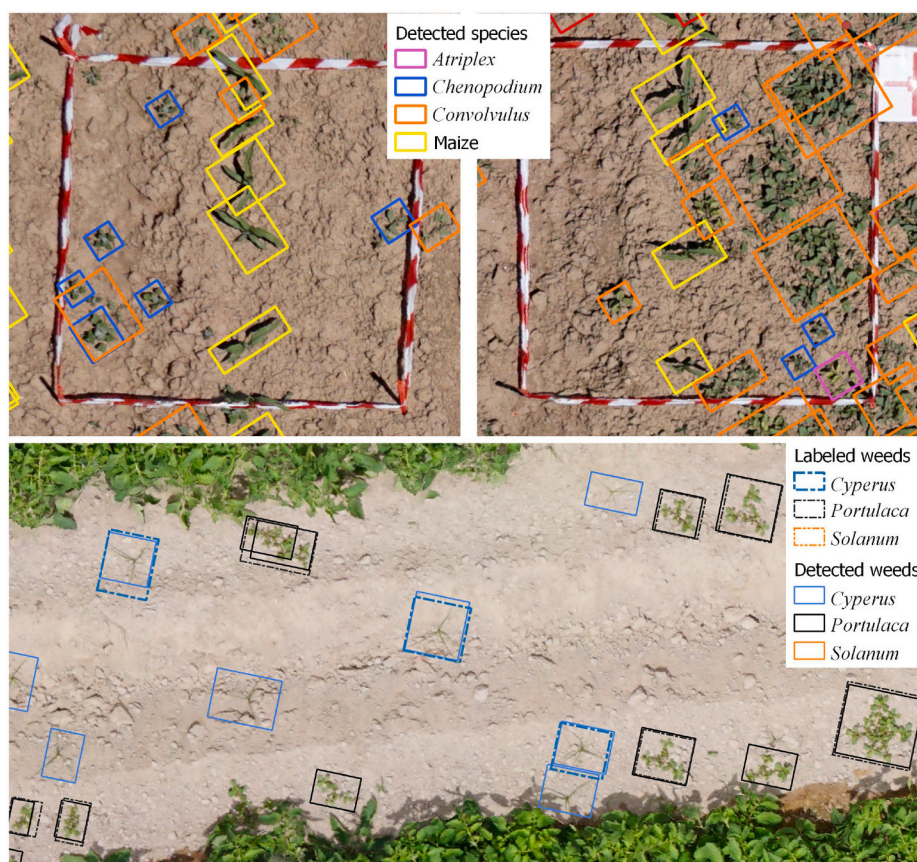


Fig. 9. Detection and identification of crop and weed species in maize (upper images), as well as weed species in tomato (lower image), using the integrated architecture of Faster R-CNN with the Inception-ResNet-v2 classifier on split orthophoto images.

like *Atriplex patula* and *Lolium rigidum* weeds. Unequal label distribution penalizes minority species, decreasing accuracy, while favoring more abundant species, such as crops. Therefore, determining the appropriate number of labels for CNN model training and validation is crucial for the precise and effective implementation of SSWM techniques. Our results suggest maintaining ratios of 1:12, 1:24, and 1:107 between the majority and minority species for VGG16, ResNet152 and Inception-ResNet-v2 models, respectively, to ensure *F1-score* exceeding 80% per species.

The integration of Faster R-CNN algorithms with Inception-ResNet-v2 shows promise for detecting and classifying various weed species in maize and tomato crops. While these architectures present opportunities for optimization in agriculture, it is worth recognizing their limitations. Factors such as image resolution, plant overlap and the presence of other structures or elements in the field can affect detection accuracy. Generalization to different environmental conditions and crop types requires further evaluation. Despite the challenges, the results obtained support the potential usefulness of these architectures in agriculture and provide valuable information for crop management. Further research is recommended to address limitations and refine the applicability of the technology in diverse agricultural settings.

Funding

This work was supported by the Spanish Research State Agency (AEI) through the Project PID2020-113229RB-C41/AEI/10.13039/501100011033. The lead author, G.A. Mesías-Ruiz has been a beneficiary of a FPI fellowship by the Spanish Ministry of Education and Professional Training (PRE2018-083227). The research of I. Borra-Serrano was financed by the grant FJC2021-047687-1 funded by MCIN/AEI/10.13039/501100011033 and European Union NextGenerationEU/PRTR.

CRediT authorship contribution statement

G.A. Mesías-Ruiz: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **I. Borra-Serrano:** Writing – review & editing, Software, Methodology, Investigation, Formal analysis, Conceptualization. **J.M. Peña:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization. **A.I. de Castro:** Writing – review & editing. **C. Fernández-Quintanilla:** Writing – review & editing. **J. Dorado:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We acknowledge the support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). The authors thank all those who have helped in the realization of this dataset. The authors thank the technicians at the Tech4Agr0 research group for their support during field sampling and image labelling.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. Tensorflow: large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv doi:<https://arxiv.org/abs/1603.04467>.
- Bah, M.D., Hafiane, A., Canals, R., 2018. Deep learning with unsupervised data labeling for weed detection in line crops in UAV images. *Rem. Sens.* 10, 1690. <https://doi.org/10.3390/rs10111690>.
- Chen, D., Lu, Y., Li, Z., Young, S., 2022. Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems. *Comput. Electron. Agric.* 198, 107091 <https://doi.org/10.1016/j.compag.2022.107091>.
- Chew, R., Rineer, J., Beach, R., O'Neil, M., Ujeneza, N., Lapidus, D., Miano, T., Hegarty-Craver, M., Polly, J., Temple, D.S., 2020. Deep neural networks and transfer learning for food crop identification in UAV images. *Drones* 4, 7. <https://doi.org/10.3390/drones4010007>.
- dos Santos Ferreira, A., Freitas, D.M., da Silva, G.G., Pistori, H., Folhes, M.T., 2019. Unsupervised deep learning and semi-automatic data labeling in weed discrimination. *Comput. Electron. Agric.* 165, 104963 <https://doi.org/10.1016/j.compag.2019.104963>.
- dos Santos Ferreira, A., Matte Freitas, D., Gonçalves da Silva, G., Pistori, H., Theophilo Folhes, M., 2017. Weed detection in soybean crops using convnets. *Comput. Electron. Agric.* 143, 314–324. <https://doi.org/10.1016/j.compag.2017.10.027>.
- Dyrmann, M., Karstoft, H., Midtby, H.S., 2016. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. <https://doi.org/10.1016/j.biosystemseng.2016.08.024>.
- Fernandez-Quintanilla, C., Dorado, J., Andújar, D., Peña-Barragán, J.M., 2022. Advanced detection technologies for weed scouting. In: Pereditor, K. (Ed.), *Advances in Integrated Weed Management*. Burleigh Dodds Science Publishing Limited 2021, Cambridge, pp. 205–227. <https://doi.org/10.19103/AS.2021.0098.09>.
- Fernández-Quintanilla, C., Peña, J.M., Andújar, D., Dorado, J., Ribeiro, A., López-Granados, F., 2018. Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? *Weed Res.* 58, 259–272. <https://doi.org/10.1111/wre.12307>.
- Hasan, A.S.M.M., Sohel, F., Diepeveen, D., Laga, H., Jones, M.G., 2021. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* 184, 106067 <https://doi.org/10.1016/j.compag.2021.106067>.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.* 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- Huang, H., Lan, Y., Yang, A., Zhang, Y., Wen, S., Deng, J., 2020. Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery. *Int. J. Rem. Sens.* 41, 3446–3479. <https://doi.org/10.1080/01431161.2019.1706112>.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Bach, F., Blei, D. (Eds.), *32nd Proc. Int. Conf. Mach. Learning. PMLR, Lille, France*, pp. 448–456. URL: <https://proceedings.mlr.press/v37/ioffe15.html>.
- Khalifa, N.E., Loey, M., Mirjalili, S., 2021. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artif. Intell. Rev.* <https://doi.org/10.1007/s10462-021-10066-4>.
- Khan, S., Tufail, M., Khan, M.T., Khan, Z.A., Iqbal, J., Alam, M., 2021. A novel semi-supervised framework for UAV based crop/weed classification. *PLoS One* 16, e0251008. <https://doi.org/10.1371/journal.pone.0251008>.
- Krestenitis, M., Raptis, E.K., Kapoutsis, A.C., Ioannidis, K., Kosmatopoulos, E.B., Vrochidis, S., Kompatsiaris, I., 2022. CoFly-WeedDB: a UAV image dataset for weed detection and species identification. *Data Brief* 45, 108575. <https://doi.org/10.1016/j.dib.2022.108575>.
- Lati, R.N., Rasmussen, J., Andujar, D., Dorado, J., Berge, T.W., Wellhausen, C., Pflanz, M., Nordmeyer, H., Schirrmann, M., Eizenberg, H., Neve, P., Jørgensen, R.N., Christensen, S., 2021. Site-specific weed management—constraints and opportunities for the weed research community: insights from a workshop. *Weed Res.* 61, 147–153. <https://doi.org/10.1111/wre.12469>.
- Liu, D., Zhao, J., Wu, J., Yang, G., Lv, F., 2022. Multi-category classification with label noise by robust binary loss. *Neurocomputing* 482, 14–26. <https://doi.org/10.1016/j.neucom.2022.01.031>.
- Meier, U., 2018. Growth Stages of Mono- and Dicotyledonous Plants. *BBCH Monograph*. <https://doi.org/10.5073/20180906-074619>.
- Mesías-Ruiz, G.A., Borra-Serrano, I., Peña, J., de Castro, A.I., Fernández-Quintanilla, C., Dorado, J., 2024. Unmanned Aerial Vehicle Imagery for Early Stage Weed Classification and Detection in Maize and Tomato Crops. *DIGITAL.CSIC*. <https://doi.org/10.20350/digitalCSIC/16131>.
- Mesías-Ruiz, G.A., Pérez-Ortiz, M., Dorado, J., de Castro, A.I., Peña, J.M., 2023. Boosting precision crop protection towards agriculture 5.0 via machine learning and emerging technologies: a contextual review. *Front. Plant Sci.* 14 <https://doi.org/10.3389/fpls.2023.1143326>.
- Olsen, A., Kononov, D.A., Philippa, B., Ridd, P., Wood, J.C., Johns, J., Banks, W., Girgenti, B., Kenny, O., Whinney, J., Calvert, B., Azghadi, M.R., White, R.D., 2019. DeepWeeds: a multiclass weed species image dataset for deep learning. *Sci. Rep.* 9, 2058. <https://doi.org/10.1038/s41598-018-38343-3>.
- Pan, S.J., Yang, Q., 2010. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>.
- Rai, N., Zhang, Y., Ram, B.G., Schumacher, L., Yellavajjala, R.K., Bajwa, S., Sun, X., 2023. Applications of deep learning in precision weed management: a review. *Comput. Electron. Agric.* 206, 107698 <https://doi.org/10.1016/j.compag.2023.107698>.
- Rejeb, A., Abdollahi, A., Rejeb, K., Treiblmaier, H., 2022. Drones in agriculture: a review and bibliometric analysis. *Comput. Electron. Agric.* 198, 107017 <https://doi.org/10.1016/j.compag.2022.107017>.
- Ren, S., He, K., Girshick, R., Sun, J., 2016. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebis, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., 2018. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Rem. Sens.* 10 <https://doi.org/10.3390/rs10091423>.
- Shahi, T.B., Dahal, S., Sitaula, C., Neupane, A., Guo, W., 2023. Deep learning-based weed detection using uav images: a comparative study. *Drones* 7, 624. <https://doi.org/10.3390/drones7100624>.
- Shawky, O.A., Hagag, A., El-Dahshan, E.S.A., Ismail, M.A., 2020. Remote sensing image scene classification using cnn-mlp with data augmentation. *Optik* 221, 165356. <https://doi.org/10.1016/j.ijleo.2020.165356>.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint. <https://arxiv.org/abs/1409.1556>.
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. 31st AAAI conf. Artif. Intell. AAAI 2017, 4278–4284doi. <https://doi.org/10.1609/aaai.v31i1.11231>.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>.
- Torres-Sánchez, J., Mesas-Carrascosa, F.J., Jiménez-Brenes, F.M., de Castro, A.I., López-Granados, F., 2021. Early detection of broad-leaved and grass weeds in wide row crops using artificial neural networks and UAV imagery. *Agronomy* 11. <https://doi.org/10.3390/agronomy11040749>.
- Tzutalin, 2015. LabelImg. URL: <https://github.com/tzutalin/labelImg>.
- Valente, J., Hiremath, S., Ariza-Sentís, M., Doldersum, M., Kooistra, L., 2022. Mapping of Rumex obtusifolius in nature conservation areas using very high resolution UAV imagery and deep learning. *Int. J. Appl. Earth Obs. Geoinf.* 112, 102864 <https://doi.org/10.1016/j.jag.2022.102864>.