






Semantic Segmentation of Weeds and Crops in Multispectral Images by Using a Convolutional Neural Networks Based on U-Net

Miguel Ángel Chicchón Apaza¹(✉) ,
Héctor Manuel Bedón Monzón¹(✉) , and Ramon Alcarria² 

¹ Exponential Technology Group (GITX-ULIMA),
Institute of Scientific Research (IDIC), University of Lima, Lima, Peru
chicchon.am@gmail.com, hbedon@ulima.edu.pe

² Telematics Engineering Department, Technical University of Madrid (UPM),
Madrid, Spain

Abstract. A first step in the process of automating weed removal in precision agriculture is the semantic segmentation of crops, weeds and soil. Deep learning techniques based on convolutional neural networks are successfully applied today and one of the most popular network architectures in semantic segmentation problems is U-Net. In this article, the variants in the U-Net architecture were evaluated based on the aggregation of residual and recurring blocks to improve their performance. For training and testing, a set of data available on the Internet was used, consisting of 60 multispectral images with unbalanced pixels, so techniques were applied to increase and balance the data. Experimental results show a slight increase in quality metrics compared to the classic U-Net architecture.

Keywords: Deep learning · Convolutional neural network · U-Net · Semantic segmentation · Precision agriculture

1 Introduction

The sustainability of agriculture is one of Sustainable Development Objectives (SDO) of the United Nations. To achieve this objective, new smart farming methods are required to increase or maintain crop yields minimizing environmental impact. Precision agriculture techniques achieve this goal through the spatial study of key indicators of crop health and the application of treatments such as herbicides, pesticides and fertilizers, only in relevant areas [1].

Conventional weed control systems apply uniformly with the same dose of herbicide in the entire field. In contrast, the new perception-controlled elimination systems offer the potential to perform a treatment for each plant, for example, by selective spraying or mechanical weed control. However, this process requires a plant classification system that can analyze the image data recorded in the field in real time and label

The original version of this chapter was revised: the author's name has been changed as "Ramon Alcarria". The correction to this chapter is available at https://doi.org/10.1007/978-3-030-42520-3_52

individual plants as crops or weeds [2]. Field images acquired with these new systems can provide abundant information, however, their natural environment with different plants that grow together in a messy scene present many challenges [3]. Among the challenges are vegetation segmentation (vegetation in the first layer and ground in the background), segmentation of individual plants, segmentation of crops and weeds, and phenotyping of individual plants. The first three challenges are addressed directly by machine learning. The fourth challenge includes the growth stage, the position of the plant stem, the amount of biomass, the leaf count, the leaf area, among others. In addition, the crop/weed coverage index, crop spacing, crop plant counts and other derived measurements are of special interest to farmers.

This article focuses on the design, implementation and evaluation of deep learning algorithms based on the U-Net convolutional network architecture for crop and weed segmentation in multispectral images used in precision agriculture. The main contribution is the evaluation of modifications on the U-Net network in order to make it more optimal for the recognition of weeds and crops. For this reason, three variants of the U-Net convolutional network architecture are presented and its performance is evaluated using metrics such as the Jaccard index or Intersection over Union (IoU) and recall. The rest of this article is structured as follows: Section 1 presents a synthesis of the contribution of the main articles focused on this problem. Section 2 describes the methodology to follow. Experimentation and results are presented in Sect. 3 and finally in Sect. 4 discussions and future work.

1.1 Related Work

Image filters are used using the computational vision approach [4]. Søggaard [5] uses active shape models to classify weed types.

For weed discrimination, models were used for real-time detection [6] using the Haar wavelet transform (HWT) for image decomposition and the k-nearest neighbors (KNN) method obtaining 94% of precision improving the used baselines. Random Forest and support vector machines (SVM) are used for detection [7]. Also, semi-supervised approaches were used [8].

In recent years several studies have been carried out for the application of deep learning in agriculture, among them we have the works [9, 11] of the techniques used for deep learning. Convolutional neural networks (CNN) are studied in [2, 11]. In other investigations, unsupervised models of labeling is used first, then apply CNN based on ResNet18 [12]. Another approach uses a CNN with sliding windows [13], where from the calculation of a relationship between weed detection rate (WD) and crop waste (CW), it was discovered that the size of the sliding window of [80 80] results in an effective detection of weeds with 63.28% and a minor cause of crop damage with 13.33%. Lottes [14] uses fully convolutional networks (FCN) with an encoder-decoder structure achieving a level of completeness of 92.4% for weeds and 96.1% for cultivation. In other investigations, 86.2% accuracy is achieved for 22 types of weeds with crops [15] and 94% accuracy at pixel level [16].

1.2 U-Net

In biomedical image segmentation context, it is assumed that thousands of training data are required for successful training of a deep learning network. Ronneberger [17] presented the U-Net model based in CNN with a training strategy that focus primarily on data augmentation and contraction-expansion to use the available data more efficiently (see Fig. 1). The network can be trained from few images and its performance is remarkable. U-Net was also used in other applications such as radiofrequency [18]. The use of U-Net for this problem is explain in [10], where it is compared with other neural networks. There are other alternatives such as SegNet [19] applied to weed detection [1] or WeedMap that has been used in precision agriculture [20].

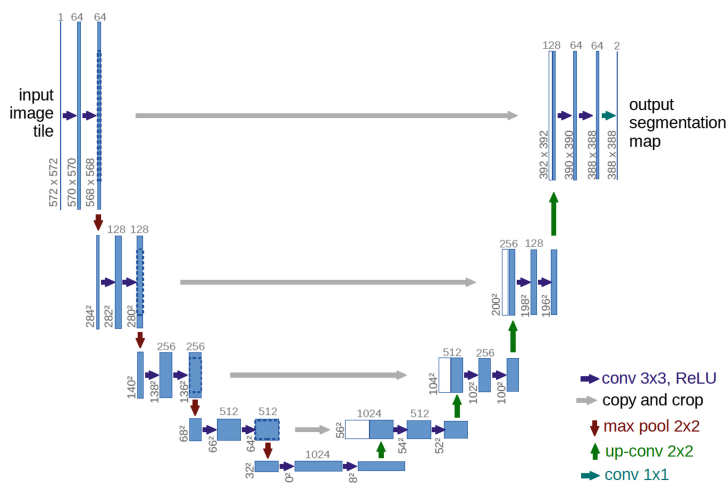


Fig. 1. Architecture U-Net [17].

1.3 Dataset

For the labeling data process, it is necessary Human intervention, which can be a very tedious task, initiatives [21] are proposed for the automatic generation of data sets based on a series of key features. Other several investigations use their own set of data taken on drones or cameras [12, 15]. Huag [3] proposes a data set of 60 images called CWFID (Crop Weed Field Image Dataset) which is complemented in [1]. This data set is used in investigations [7, 10] and in this article.

2 Methods

The objective of this article is to answer the questions: Is the U-Net convolutional network architecture effective for the segmentation of weeds and crops? Is it possible to improve the effectiveness of the U-Net convolutional network architecture by adding

residual and recurrent layers for weed and crop segmentation? To answer these questions, the methodology described below is followed (see Fig. 2):

- Acquisition of the data set containing masks of weeds, soils and crops.
- Pre-processing through data augmentation explained above.
- Separation of test sets, validation and tests.
- Reduced tests (less steps) of the model using the hyper-parameters chosen in order to choose the best values.
- Training with the chosen hyper-parameters, using the set of tests and validation.
- Obtaining the metrics defined by validating the model with the set of tests.

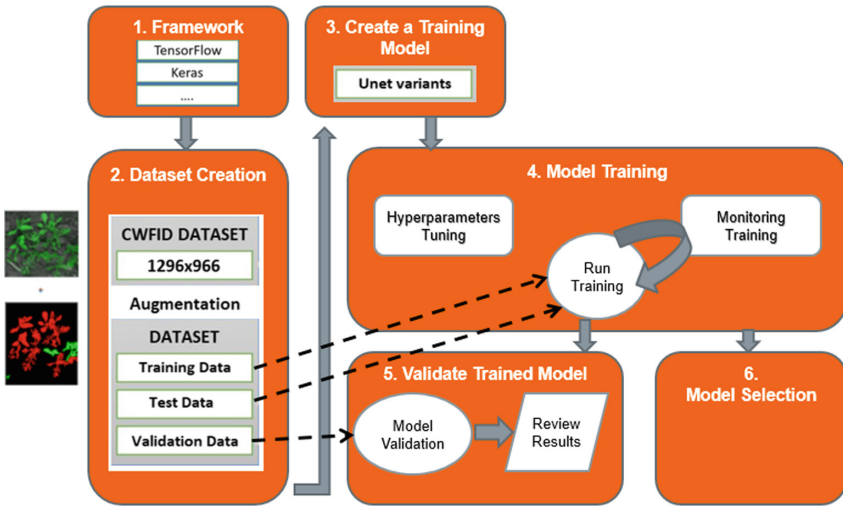


Fig. 2. Proposed process.

2.1 Pre-processing

Generally, these data sets contain very few images, so augmentation was performed with the following strategies:

- Reflection of images horizontally and vertically.
- Sliding images.
- Noise by altering the RGB channels.
- Elastic deformation.
- Gaussian noise.
- Cropping.

Additionally, the size of the images was reduced in order to have sufficient computational capacity to perform the tests.

2.2 Quality Metrics

The Jaccard index or Intersection Over Union (IOU) was used, since it is a metric widely used in object detection and allows measuring the degree of similarity between the predicted image and the mask image.

Another metric used is recall due to the interest in controlling the proportion of real positives correctly identified. In the case of the problem, it is of interest to keep the number of crops identified as weeds (negative faults) as low as possible [7].

Additionally, precision and F1 score were used as complementary metrics in order to make comparisons with the baseline.

2.3 Proposed Model

The models evaluated are variants of the U-Net convolutional network architecture, which is one of the most popular architectures in segmentation applications.

First, a recurrent convolutional neural network based on U-Net was evaluated, since the accumulation of characteristics with recurrent residual convolutional layers guarantees a better representation of the characteristics for segmentation tasks. Secondly, a residual convolutional neural network based on U-Net models was evaluated, because a residual unit helps the training of a deep architecture. Thirdly, a recurrent residual convolutional neuronal network was evaluated in order to use the advantages already mentioned. In Fig. 3 the U-Net base architecture is observed, where the blocks in red are convolutional units modified according to the variants shown in Fig. 4.

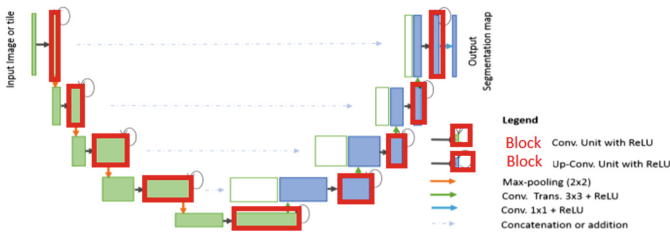


Fig. 3. Architecture variants U-Net [17]. (Color figure online)

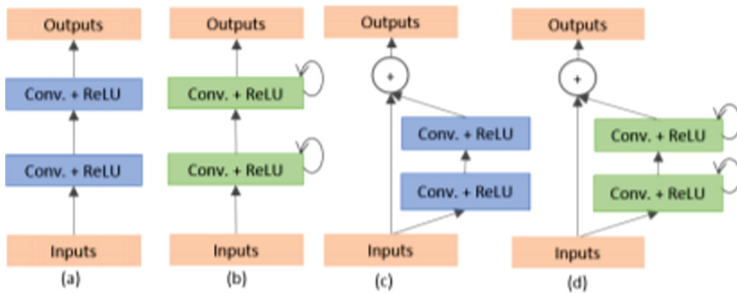


Fig. 4. Variants of convolutional units: (a) front convolutional units, (b) recurrent convolutional block (c) residual convolutional unit, and (d) recurrent residual convolutional units [22].

3 Results

3.1 Dataset

The data set used in this investigation is the Crop Weed Field Image Dataset (CWFID) [3], which consists of a set of 60 images of 1296×966 pixels, labeled with 3 classes (soil, weed, crop) that are shown in Fig. 5. Scaling was performed to reduce images to 246×256 pixels in order to improve computational capacity.

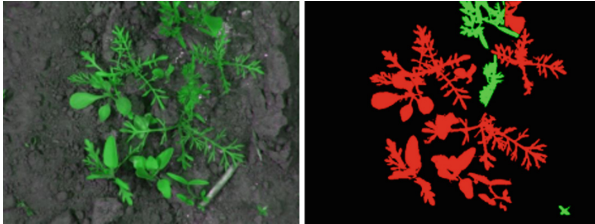


Fig. 5. Right: multispectral image. Left: labeled Image [3].

A data set with the following characteristics has been prepared: 40 images randomly chosen as training set. From the training set we will take the images number 11, 20, 41 and 52 to be aligned with the baseline. The set of tests will be the remaining 20 images.

The image is reduced to a resolution of 256 by 256 pixels.

For the augmentation the following strategies are carried out that will be applied only to the training and validation sets:

- Reflection of images horizontally, vertically and diagonally. With this we would have 3 additional images for each image. The Numpy library written in Python was used.

On all the images generated previously the following strategies were used:

- Sliding of the images: The sliding was done by random values and filling the remaining space with part of the image as shown in Fig. 6.
- Noise by altering the RGB channels. A color will be chosen randomly as shown in Fig. 6.
- Elastic deformation with random selection of alpha and sigma values as shown in Fig. 6.
- Gaussian noise in order to prevent overfitting. It will be added to each model and the best value will be validated by selecting hyper parameters.
- Crop in order to generate new images using fragments of it.

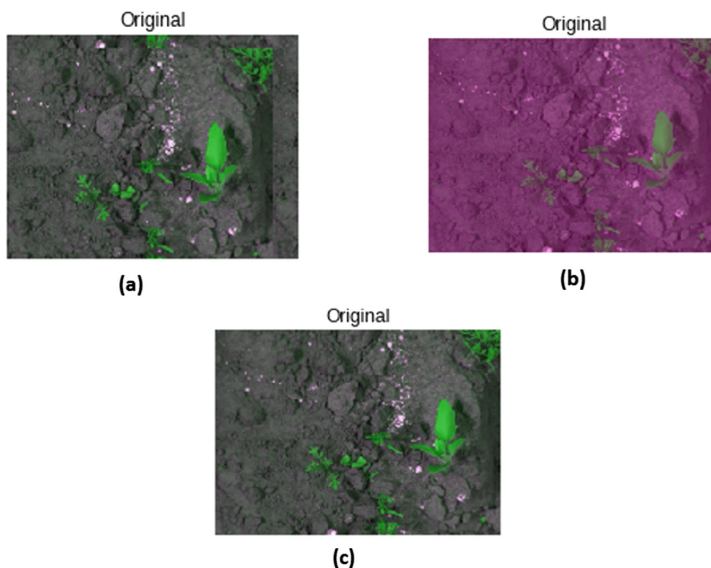


Fig. 6. (a) Sliding images. (b) Noise with channel alteration. (c) Elastic deformation. (Color figure online)

After making these modifications we have the following sets of data: 1560 images as a training set, 520 as a validation set and 20 as a test set.

3.2 Experimentation Environment and Baseline

Google Collaboratory has been used as a cloud platform. It allows us to carry out a collaborative and distributed work. It uses an Intel (R) Xeon (R) CPU @ 2.30 GHz processor with 12 GB of RAM. A Tesla P100 GPU with 16 GB of memory. Experiments were performed using Anaconda as a development environment and Python 3.6 as a programming language. The neural network models were developed using the Keras library on Tensorflow 2.0.

The strategies used were aligned to the research carried out by Cereda [10] which contains experimentation with U-Net and he uses the chosen metrics. The proposed models were developed from [22]. Cereda [13] conducted the experiment with 10 classifiers, which were evaluated with the indicators: Accuracy, Precision, Recall, F1 and Jaccard. This evaluation was performed at the pixel level in full size of the images extracted for the test set. The results of the neural network models are shown in Table 1. It can be seen in the results of Table 1 that the U-Net classifier has better performance levels in the majority of quality indicators used in the evaluation of the investigation.

Table 1. Results obtained in Cereda’s research with the data set.

Model	Accuracy	Precision	Recall	F1	Jaccard
U-Net	0.897	0.886	0.972	0.922	0.879
U-ReNet	0.863	0.873	0.942	0.898	0.879
U-ReNet2	0.873	0.895	0.915	0.897	0.879
ReSeg	0.854	0.860	0.946	0.894	0.879
ReConv	0.737	0.770	0.883	0.809	0.879
ReConcat	0.743	0.769	0.846	0.799	0.879

3.3 Model Training

The following hyper parameters have been used during training. For them, each model was run 10 times with data set 1 using the following hyper-parameters.

- Learning rate (lr): It controls how much the weights of our model are adjusted with respect to the gradient. Possible values assigned: 0.01, 0.005, 0.001.
- L2: Assigned possible values: none, 0.01, 0.001, 0.0001.
- Gaussian filter: it will help us control overfitting. Possible values assigned: 0.5, 0.05, 0.005.
- Dropout that will be added to each of the convolutional layers. Possible values assigned: none, 0.1, 0.2.
- Batch normalization that will be added to each of the convolutional layers.

After performing the tests, the following hyper parameters have been chosen for each model in Table 2.

Table 2. Better hyper parameters.

Model	Jaccard
U-Net	lr = 0.001, Gaussian filter = 0.005, batch normalization
ResU-Net	lr = 0.001, Gaussian filter = 0.005, batch normalization
RU-Net	lr = 0.001, Gaussian filter = 0.005, batch normalization
R2U-Net	lr = 0.001, Gaussian filter = 0.005, batch normalization

Next, the training was carried out using the selected hyper parameters. Each model was executed in 200 periods using as a loss function: categorical crossentropy, Adam as an optimizer and a batch size of 30 for U-Net and ResU-Net, and 10 for RU-Net and R2U-Net. Table 3 shows the most relevant configurations.

Table 3. Description and relevant settings.

Model	Total parameters	Trainable parameters	Non-trainable parameters
U-Net	7768099	7764131	3968
ResU-Net	8117827	8113859	3968
RU-Net	24281347	24269443	11904
R2U-Net	24631075	24619171	11904

The metrics obtained are shown in Table 4.

Table 4. Results obtained with the data set.

Model	Accuracy	Precision	Recall	F1	Jaccard
U-Net	0.8825	0.8939	0.9307	0.8351	0.9760
ResU-Net	0.8724	0.9397	0.9669	0.8505	0.9749
RU-Net	0.9027	0.9212	0.9675	0.8917	0.9783
R2U-Net	0.8660	0.9311	0.9649	0.8376	0.9740

The learning curves for the training and validation sets are shown in Fig. 7. The execution times are shown in Table 5.

Table 5. Runtime in milliseconds per image.

Model	Time
U-Net	12
ResU-Net	14
RU-Net	37
R2U-Net	40

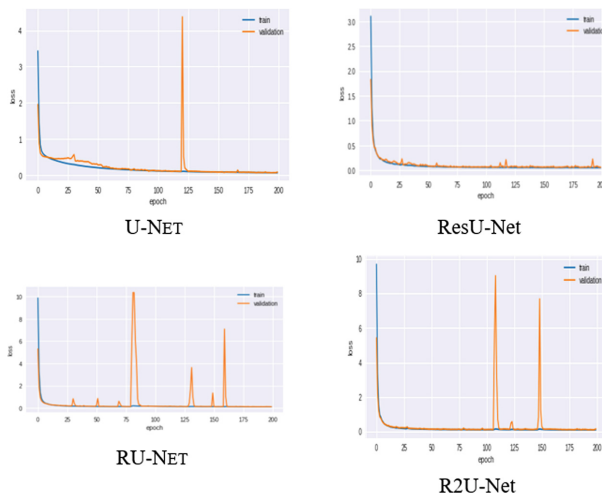


Fig. 7. Learning curve for data set.

4 Discussion

As part of this chapter, based on the results, some topics will be discussed to interpret the experiments performed and find opportunities for improvements.

In Table 6 shows the best values of each data set. The results obtained with respect to the baseline are better except in the recall metric and F-1. The RU-Net model obtained the best results in all metrics except Precision.

The research proposes three additional models that don't use the baseline, and U-Net is the only one presented in both. When comparing the results, the baseline has been exceeded in precision and Jaccard.

Table 6. Better values in the data set.

Model	Accuracy	Precision	Recall	F1	Jaccard
Base line	U-Net	U-ReNet2	U-Net	U-Net	All
Base line	0.897	0.925	0.972	0.922	0.879
Best model	RU-Net	ResU-Net	RU-Net	RU-Net	RU-Net
Best model	0.9027	0.9397	0.9675	0.8917	0.9783

The learning curve of the models used is shown in Fig. 7. It is important to note that all graphs have a similar shape. Note that some models show temporary fluctuations in the loss function in the validation set. This could be due to a possible noise in the data due to augmentation.

It is show clearly that the model differentiates well between the cultivation and vegetation. However, we see some problems to distinguish between cultivation and weeds. One of the main problems detected is when the weeds (red) and the crop (green) are overlapping or very close as is shown in Fig. 8.

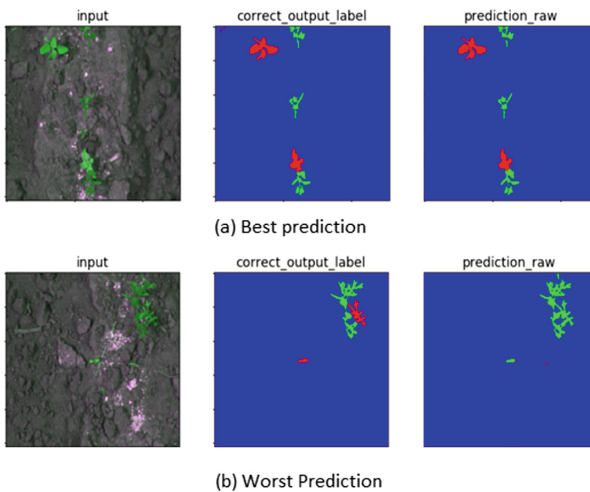


Fig. 8. Predictions with RU-Net. (Color figure online)

During the experiment, data augmentation has been used, as part of the improvement opportunity it is proposed to increase the amount of data augmentation. For example, in the baseline It is used up to 25,000 images with the almost 1000 used. Also, perform more tests by making changes in hyperparameters. Perform tests with a larger batch size, although this requires greater computational capacity. Additionally, Try other optimizers like RMSProp. Tests were performed using dropout layers and batch normalization layers where better results could be observed. It is necessary to improve the architecture using and/or proposing improvements to the layers such as attention mechanisms that allow efficient location of objects and an increase in performance in general.

4.1 Conclusions

From a practical point of view, this work should be expanded to be able to distinguish different types of weeds and to estimate the growth status of the crops. This implies extending the manual annotation to include this new data. For the weed detection problem, it is necessary to obtain a larger data set than the used for the present investigation.

The main objective of the present work was to carry out the experimentation of architectures of neural networks based on U-Net applied to the segmentation of crops and weeds having as base line an experimentation already carried out [10]. It is concluded from the results that using recurrent layers within the U-Net architecture allows to improve the effectiveness in the problem of crop and weed segmentation with multispectral images of the data set used. In contrast, the residual layers did not add any improvement.

From the evaluation analysis of the segmentation, it was observed that the same metric result can be obtained in different ways, therefore, it should be interesting to investigate which of the metrics is most suitable for resolving this type of problem. Finally, it is proposed to deepen research on topics such as data augmentation, the choice of hyper-parameters and assembly models in order to achieve better results. Additionally, Perform the experiment with other architectures and different data sets.

Acknowledgments. This research was supported by National Agriculture Innovation Program (PNIA) of Peru and the Institute of Scientific Research (IDIC) of the University of Lima.

References

1. Sa, I., et al.: WeedNet: dense semantic weed classification using multispectral images and MAV for smart farming. *IEEE Robot. Autom. Lett.* **3**(1), 588–595 (2017). <https://doi.org/10.1109/LRA.2017.2774979>
2. Milioto, A., Lottes, P., Stachniss, C.: Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp. 2229–2235. IEEE (2018). <https://doi.org/10.1109/icra.2018.8460962>

3. Haug, S., Ostermann, J.: A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8928, pp. 105–116. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16220-1_8
4. Molina-Villa, M.A., Solaque-Guzmán, L.E.: Machine vision system for weed detection using image filtering in vegetables crops. *Revista Facultad de Ingeniería Universidad de Antioquia* **80**, 124–130 (2016). <https://doi.org/10.17533/udea.redin.n80a13>
5. Søggaard, H.T.: Weed classification by active shape models. *Biosyst. Eng.* **91**(3), 271–281 (2005). <https://doi.org/10.1016/j.biosystemseng.2005.04.011>
6. Ahmad, I., Siddiqi, M.H., Fatima, I., Lee, S., Lee, Y.K.: Weed classification based on Haar wavelet transform via k-nearest neighbor (k-NN) for real-time automatic sprayer control system. In: Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, p. 17. ACM (2011). <https://doi.org/10.1145/1968613.1968634>
7. Voorhoeve, L.: *Machine Learning for Crop and Weed Classification* (2018)
8. Lottes, P., Stachniss, C.: Semi-supervised online visual crop and weed classification in precision farming exploiting plant arrangement. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5155–5161. IEEE (2017). <https://doi.org/10.1109/iros.2017.8206403>
9. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: a survey. *Comput. Electron. Agric.* **147**, 70–90 (2018). <https://doi.org/10.1016/j.compag.2018.02.016>
10. Cereda, S.: A comparison of different neural networks for agricultural image segmentation (2017)
11. Potena, C., Nardi, D., Pretto, A.: Fast and accurate crop and weed identification with summarized train sets for precision agriculture. In: Chen, W., Hosoda, K., Menegatti, E., Shimizu, M., Wang, H. (eds.) IAS 2016. AISC, vol. 531, pp. 105–121. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-48036-7_9
12. Bah, M.D., Hafiane, A., Canals, R.: Deep learning with unsupervised data labeling for weeds detection on UAV images (2018). arXiv preprint [arXiv:1805.12395](https://arxiv.org/abs/1805.12395)
13. Kantipudi, K., Lai, C., Min, C.-H., Chiang, R.C.: Weed detection among crops by convolutional neural networks with sliding windows. In: 14th International Conference on Precision Agriculture, Quebec (2018)
14. Lottes, P., Behley, J., Milioto, A., Stachniss, C.: Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. *IEEE Robot. Autom. Lett.* **3**(4), 2870–2877 (2018). <https://doi.org/10.1109/LRA.2018.2846289>
15. Dyrmann, M., Karstoft, H., Midtiby, H.S.: Plant species classification using deep convolutional neural network. *Biosyst. Eng.* **151**, 72–80 (2016). <https://doi.org/10.1016/j.biosystemseng.2016.08.024>
16. Dyrmann, M., Mortensen, A.K., Midtiby, H.S., Jørgensen, R.N.: Pixel-wise classification of weeds and crops in images by using a fully convolutional neural network. In: Proceedings of the International Conference on Agricultural Engineering, Aarhus, Denmark, pp. 26–29 (2016)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Akeret, J., Chang, C., Lucchi, A., Refregier, A.: Radio frequency interference mitigation using deep convolutional neural networks. *Astron. Comput.* **18**, 35–39 (2017)

19. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
20. Sa, I., et al.: Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sens.* **10**(9), 1423 (2018)
21. Di Cicco, M., Potena, C., Grisetti, G., Pretto, A.: Automatic model based dataset generation for fast and accurate crop and weeds detection. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5188–5195. IEEE (2017)
22. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation (2018). arXiv preprint [arXiv:1802.06955](https://arxiv.org/abs/1802.06955)