

Article

Supervised Image Classification by Scattering Transform with Application to Weed Detection in Culture Crops of High Density

Pejman Rasti , Ali Ahmad, Salma Samiei , Etienne Belin  and David Rousseau *

LARIS, UMR INRA IRHS, Université d'Angers, 62 avenue Notre Dame du Lac, 49000 Angers, France; pejman.rasti@univ-angers.fr (P.R.); ali.ahmad@insa-lyon.fr (A.A.); salma.samiei@univ-angers.fr (S.S.); etienne.belin@univ-angers.fr (E.B.)

* Correspondence: david.rousseau@univ-angers.fr

Received: 23 December 2018; Accepted: 23 January 2019; Published: 26 January 2019



Abstract: In this article, we assess the interest of the recently introduced multiscale scattering transform for texture classification applied for the first time in plant science. Scattering transform is shown to outperform monoscale approaches (gray-level co-occurrence matrix, local binary patterns) but also multiscale approaches (wavelet decomposition) which do not include combinatory steps. The regime in which scatter transform also outperforms a standard CNN architecture in terms of data-set size is evaluated (10^4 instances). An approach on how to optimally design the scatter transform based on energy contrast is provided. This is illustrated on the hard and open problem of weed detection in culture crops of high density from the top view in intensity images. An annotated synthetic data-set available under the form of a data challenge and a simulator are proposed for reproducible science. Scatter transform only trained on synthetic data shows an accuracy of 85% when tested on real data.

Keywords: weed detection; scatter transform; deep learning; machine-learning classification; annotation; synthetic data; local binary pattern

1. Introduction

Deep learning is currently tested world-wide in almost all application domains of computer vision as an alternative to purely handcrafted image analysis [1]. When inspecting the convolutional coefficients in the first layers of deep neural networks, these are very similar to Gabor wavelets. While promoting a universal framework, deep neural networks seem to systematically converge toward tools that humans have been studying for decades. This empirical fact is used by computer scientists in the so-called transfer learning where the first layers of an already trained network are re-used [2]. This has also triggered interest by mathematicians to revisit the use of wavelets to produce universal machine-learning architectures. This interdisciplinary cross-talk resulted in the proposal of the so-called scatter transform [3], which is roughly a cascade of wavelet decomposition followed by non-linear and pooling operators. If this deep architecture bares some similarity with the standard deep learning, it does not include the time-consuming feed-forward propagation algorithm. However, it proved its comparable efficiency to deep learning while offering a very rational way of choosing the parameters of the network compared to the rather empirical current art of tuning neural networks.

Despite its intrinsic interest to address multiple scales problems compared to deep learning, scatter transform since its introduction in 2013 has been applied only on a relatively small variety of pattern recognition computer vision problems notably including iris recognition [4], rainfall classification in radar images [5], cell-scale characterization [6,7], or face recognition, [8]. Also, in these applications

scatter transform has shown its efficiency, but it was not systematically compared with other techniques in a comprehensible way. We propose to extend the scope of investigation of the applicability of scatter transform algorithm to plant science with a problem of weed detection in a background of culture crops of high density. This plant science problem is important for field robotics where the mechanical extraction of weed is a current challenge to be addressed to avoid the use of phytochemical products. From a methodological point of view, this classification problem here will also serve as a use case to assess the potential of the scatter transform when compared with other single scale and multiple scales techniques.

A large variety of platforms, sensors, and data process already exist to monitor weeds at various temporal and spatial scales. From remote sensing supported by satellites to cameras located on unmanned aerial vehicles (UAVs) or on ground-based platforms, many systems have been described and compared for the weed monitoring in arable culture crops [9–11]. Related to the observation scale of our use case, by focusing on the imaging scales of UAVs and ground-based platforms, some studies exploiting RGB data have addressed crop weed classification with a large variety of machine-learning approaches. The problem of segmentation of crop fields from typical weeds, performing vegetation detection, plant-tailored feature extraction, and classification to estimate the distribution of crops and weeds has recently been solved with convolutional neural networks in the field [12,13] and in real-time [14]. Earlier, Aitkenhead, M. et al. [15] evaluated weed detection in fields of crop seedlings using simple morphological shape characteristic extraction and self-organizing neural network. Bayesian classifier was used in [16] for plant and weed discrimination. Shape, texture features [12,17–19] or wavelet transform [20,21] coupled with various classifiers including support vector machine (SVM), relevance vector machine (RVM), fuzzy classifier, or random forests were also shown to provide successful pipelines to discriminate between plant and weeds.

The above list of reference is of course not exhaustive and new pipelines will continue to appear because of the large variety of crops shape and imaging platform. In this context, scatter transform constitutes a candidate of possible interest worth to be assessed on a plant–weed classification problem. Also, by comparison with the existing work on weed detection, the computer vision community has focused on the relatively low density of crops and weed where the soil constitutes a background to be classified in addition to crop and weed. In this paper, we consider the case of culture crops of high density, i.e., where the soil is not visible from the top view. In this case, the culture is the background and the object to be detected are weeds of wild type. The contrast in color between the background and the weed, in this case, is obviously here very low by comparison with lower density culture.

2. Material and Methods

We start by introducing the computer vision problem considered, the data-set, the expected scales included in these images and the algorithms tested for comparison with the multiscale scatter transform algorithm.

2.1. Images and Challenges

We consider the situation of a culture crops of a high density of plants (mache salad) with the undesired presence of some weeds. Images were acquired with the imaging system fixed on a robot as displayed in Figure 1. Acquisition trials, as visible in Figure 1, were done under plastic tunnels without additional light. Some sample images are given in Figure 2. Examples of weed detected in such images are shown in Figure 3 to illustrate the variability of shapes among these wild types of weeds. The computer vision task considered in this article consists in detecting the weeds from the top view as shown in the ten real-world images of Figure 2. This is challenging indeed since the intensity or color contrast between weed and crop is very weak. Also, due to the lighting conditions during acquisition, the global intensity may vary from one image to another. The contrast between weeds and plants rather stands in terms of texture since the shape of the plant considered is rather round while

the weeds included in the data-set Figure 3 are much more indented. Therefore, this computer vision problem is well adapted to test scatter transform which is a texture-based technique.

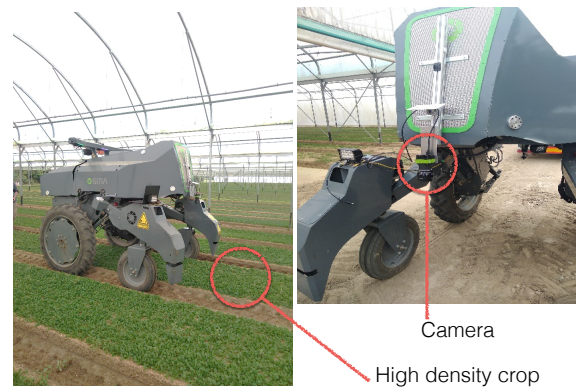


Figure 1. Global view of the imaging system fixed on a robot moving above mache salads of high density. RGB images are captured by a JAI manufactured camera of 20 M pixels with a spatial resolution of 5120×3840 pixels, mounted with a 35 mm objective. The typical distance of plants to camera is of 1 m.

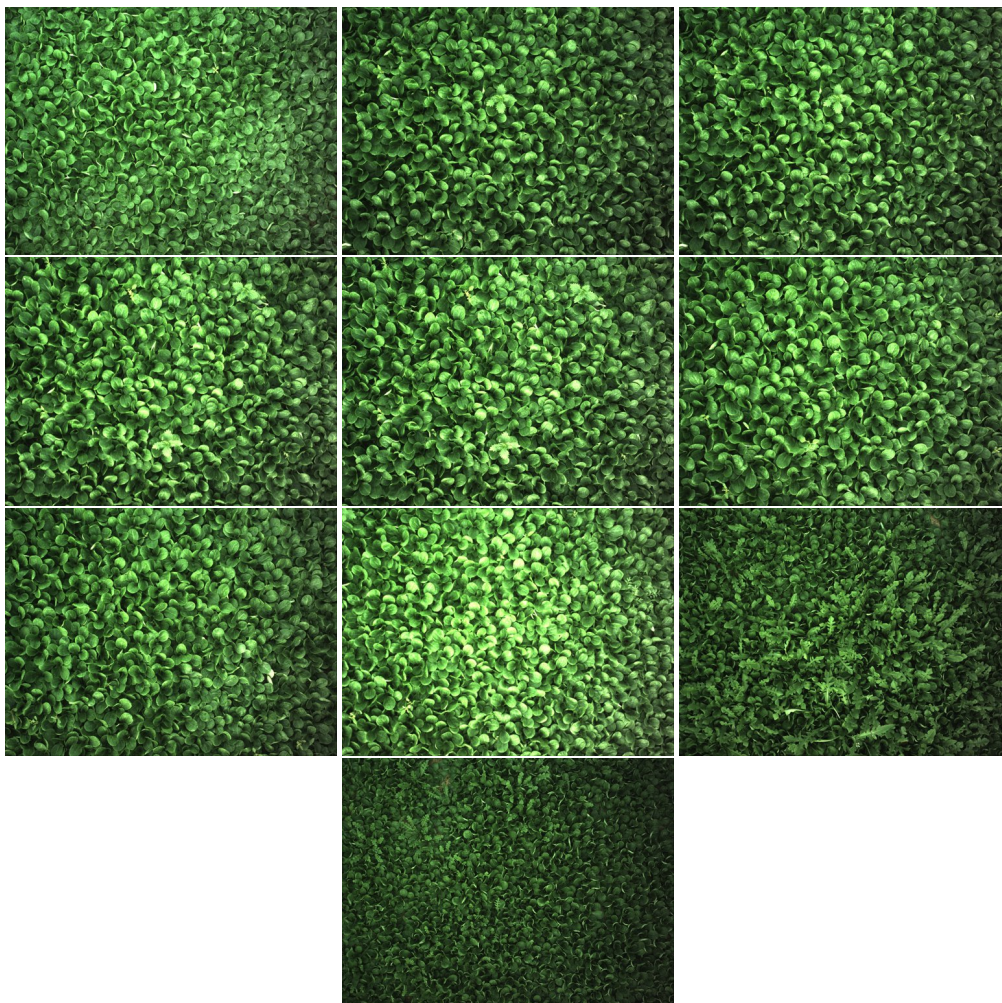


Figure 2. Set of 10 RGB images from top view for the detection of weed out of plant used as testing data-set in this study.

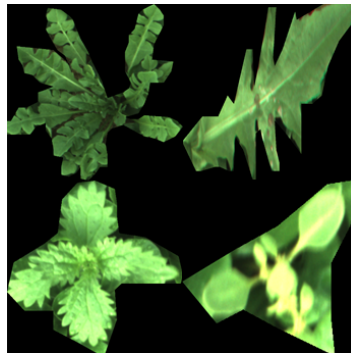


Figure 3. Illustration of different types of weeds used for the experiment.

A ground truth of the position of the weed in the ten images of Figure 2 was produced under the form of finely segmented weed and bounding box patches including these weeds. The total number of weeds being relatively low (21), we decided to generate a larger data-set with synthetic images. To simulate images similar to the real images acquired, we created a simulator which places weeds (among the 21 found in real images) from the annotated weed data-set in images of plants originally free from any weed along the pipeline shown in Figure 4.

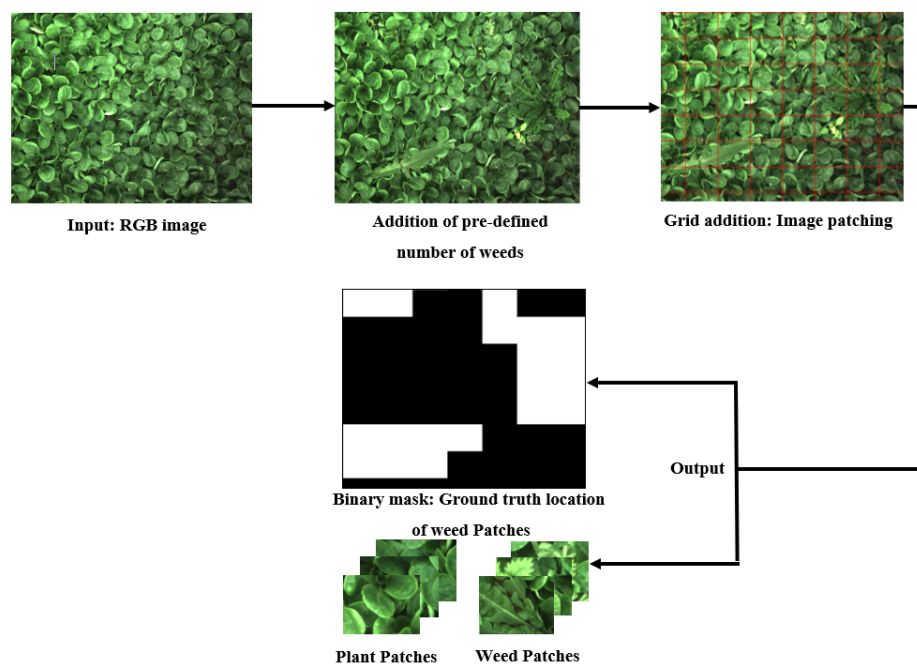


Figure 4. Simulation pipeline for the creation of images of plant with weed of Figure 3 similar to the one presented in Figure 2.

2.2. Scales

With a spatial resolution of 5120 by 3840 pixels included in the images of our data-set, and as illustrated in Figure 5, multiple anatomical structures of the dense weed/plant culture are accessible in our images. From tiny to coarse sizes, i.e., scales, this includes texture in the limb, the veins, and the leaf. There are possibly discriminant features between the two classes (weed/plant) to be found in these three scales either taken individually or combined with each other. To offer the possibility of a multiple scale analysis, together with a reasonably small computation time, classification is done at the scale of patches chosen as double size of the typical size of leaves, $2 \times \max\{S_w, S_p\}$, with rectangles

of 250 by 325 pixels where $S_w = 163$ pixels and $S_p = 157$ on average. With this constraint, we also keep for the patch the same ratio between height and width as in the original image for a periodic patch grid.

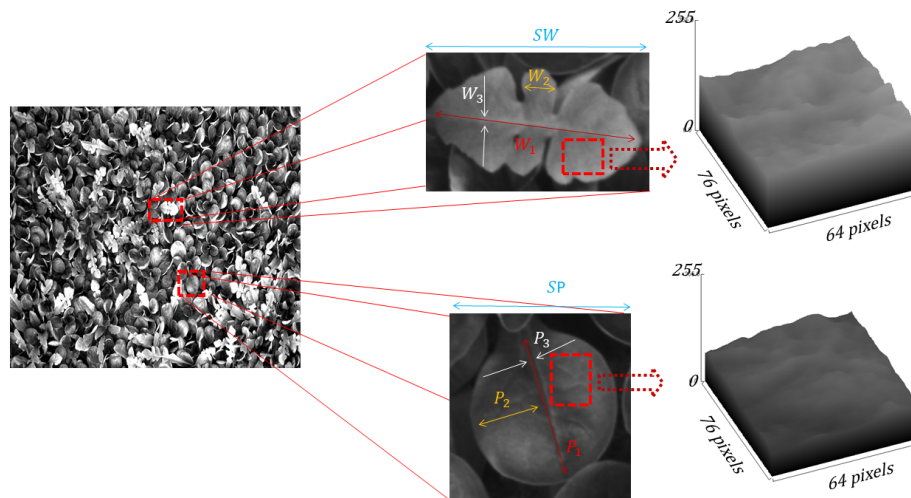


Figure 5. Anatomical scales where (W_i, P_i) presents the scales of weeds and plants respectively; (W_1, P_1) points toward the texture of the limb, (W_2, P_2) indicates the typical size of leaflet and (W_3, P_3) stands for the width of the veins. S_w and S_p show the size of a leaf of weed and plant, respectively. The classification of weed and plant is done at the scale of a patch taken as $2 \times \max(S_p, S_w)$ in agreement with a Shannon-like criteria.

2.3. Data-Set

With the simulator of Figure 4, we produced a total amount of 3292 patches containing weed and 3292 patches only with plants. The binary classification (weed/plant) is realized on these patches. This balanced data-set serves both for the training and the testing stages to assess the performance of different machine-learning tools. The data sets together with the simulator are proposed as supplementary material under the form of a free executable and a set of images (<https://uabox.univ-angers.fr/index.php/s/iuj0knyzOUgsUV9>).

2.4. Classifiers

In this section, we describe how we apply the scatter transform [3] on the weed detection problem introduced in the previous section. For comparison, we then propose a set of alternative techniques. This paper uses independent k-fold cross-validation to measure the performance of the scatter transform coupled to the classifier depicted in Figure 6 and compare other feature extractors coupled to the same classifier. The performances of these classifiers are measured by the metric of the accuracy of correct classification by

$$accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (1)$$

where TP indicates that the prediction is positive, and the actual value is positive. FP indicates that the prediction value is positive, but the actual value is negative. TN indicates that the prediction value is negative, and the actual value is negative. FN indicates that the prediction value is negative, but the actual value is positive.

2.4.1. Scatter Transform

A scattering transform defines a signal representation which is invariant to translations and potentially to other groups of transformations such as rotations or scaling. It is also stable to deformations and is thus well adapted to image and audio signal classification. A scattering transform is implemented with a convolutional network architecture, iterating over wavelet decompositions and complex modulus. Figure 6 shows a schematic view of a scatter transform network working as a feature extractor and coupled to a classifier after dimension reduction.

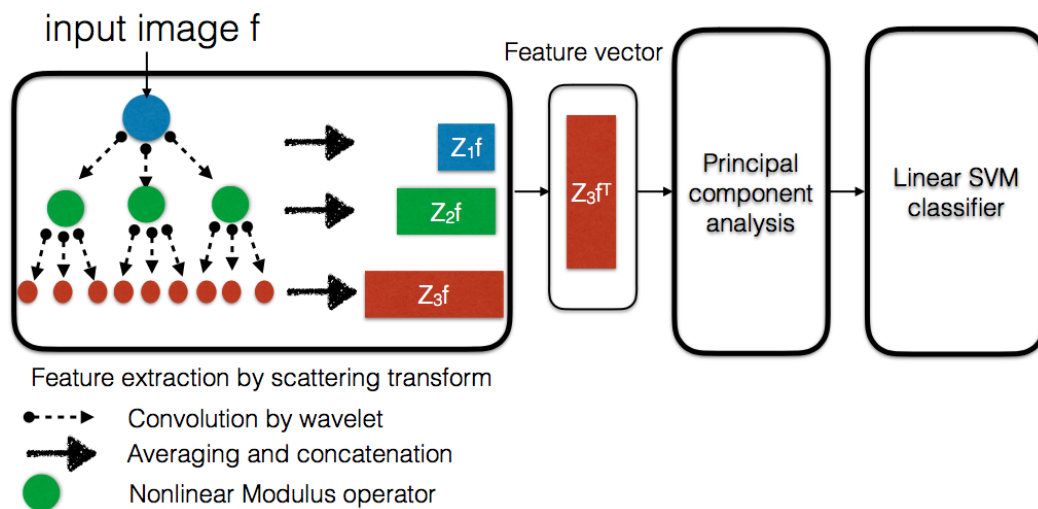


Figure 6. Schematic layout of the weed/plant classifier based on the scattering transform with three layers. The feature vector transmitted to the principal component analysis (PCA) step consists in the scatter vector $Z_m f$ of the last layer of Equation (2) after transposition.

The scatter vectors Z_m at the output of the first three layers $m = 1, 2, 3$ for an input image f are defined by

$$\begin{aligned} Z_1 f &= \{|f| \star \phi\} \\ Z_2 f &= \{\dots, |f \star \psi_{j,\theta}| \star \phi, \dots\} \\ Z_3 f &= \{\dots, |f \star \psi_{j,\theta}| \psi_{k,\phi} \star \phi, \dots\}, \end{aligned} \quad (2)$$

where the symbol \star denotes the spatial convolution, $|\cdot|$ stands for the L_1 norm, ϕ is an averaging operator, $\psi_{j,\theta}$ is a wavelet dilated by 2^j and rotated by θ . The range of scales $j = \{0, 1, \dots, J\}$ and the number of orientations $\theta = \{0, \pi/L, \dots, \pi(L-1)/L\}$ are fixed by integers J and L . The number of layers is between $m = 1$ to $m = M$. In our case, we considered as mother wavelet the Gabor filter with implementation provided under MATLAB in (<https://www.di.ens.fr/data/scattering/>) for scatter transform.

Scatter transform differs from a pure wavelet decomposition because of the non-linear modulus operator. With this nonlinearity, decomposition of the image is not done on a pure orthogonal basis (whether wavelet basis is orthogonal or not) and this opens the way of a possible benefit in the concatenation of several layers with a combination of wavelet decompositions at different scales. Interestingly, these specific properties of the scatter transform match the intrinsic multiscale textural nature of our weed detection problem which therefore constitutes an appropriate use case to assess the potential of the scatter transform in practice. A visualization of output images for various filter scale j at $m = 2$ for a given orientation is shown in Figure 7. It clearly appears in Figure 7 that the various scales (texture of the limb and veins at $j = 3$, border shape at $j = 4$ and global leaf shape at $j = 8$) presented in Section 2.2 can be captured with the different scaling factor applied on the wavelet. In our study, we empirically picked $L = 8$ orientations and investigated up to $J = 8$ scales since there are

no other anatomical items larger than the leaf itself. The number of layers tested was up to $M = 4$ as proposed in [3] since the energy after some layers although none zero is logically vanishing.

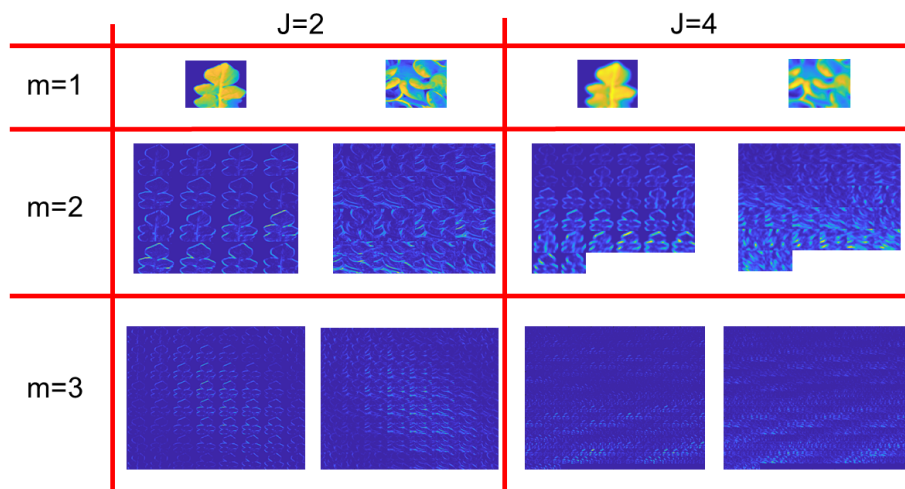


Figure 7. Output images for each class (weed on left and plant on right) and for each layer m of the scatter transform.

In the application of scatter transform to classification found in the literature so far, the optimization of the architecture was done a posteriori after supervised learning. This is rather time-consuming. We investigated the possibility to select a priori the best architecture by analyzing the distribution of relative energy E_m at the output of each layer as given by

$$E_m = \frac{\|Z_m f\|^2}{\|f\|^2} . \tag{3}$$

We computed these energies for the whole data-set as given in Table 1. As noticed in [3], the relative energy is progressively vanishing when the number of layers increases. This observation advocates for the use of a limited number of layers. However, these energies are computed on the whole population of patches including both plants and weeds and therefore it tells nothing about where to find the discriminant energy between each class throughout the feature space produced by the scatter transform. Tables 2 and 3 show the average relative energy for the weeds’ patches data-set, \bar{E}_{w_m} , and plants’ patches data-set, \bar{E}_{p_m} , for different layers m and various maximum scale J .

To show this discriminant energy between each class, various criterion could be proposed. We tested the percentage of energy similarity, Q_m , between the two classes defined by

$$Q_m = \frac{\operatorname{argmin}(\bar{E}_{w_m}, \bar{E}_{p_m})}{\operatorname{argmax}(\bar{E}_{w_m}, \bar{E}_{p_m})} \times 100. \tag{4}$$

According to this criterion, the best architecture of the scatter transform can be chosen at the point of η where the minimum Q_m between each class is found as a function of J by $\eta = \operatorname{argmin}_J(Q_m(J))$. The energy similarity $Q_m(J)$ are represented in Figure 8 and this clearly demonstrates that the contrast between classes is more pronounced on coefficient with small relative energy. This observation, not stressed in the original work of [3], indicates that it should be possible to draw benefit from the contribution of these small discriminative coefficients and thus this demonstrates the interest of the combinatory step of the scatter transform.

Table 1. Average percentage of energy of scattering coefficients E_m on frequency-decreasing paths of length m (scatter layers), with $L = 8$ orientations and various filter scale range, J , for the whole database of plants and weeds patches.

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$J = 1$	96.18	2.35	-	-	-
$J = 2$	91.81	4.61	0.28	-	-
$J = 3$	85.81	8.46	0.89	0.03	-
$J = 4$	85.81	13.15	1.97	0.17	0.006
$J = 5$	81.46	15.36	3	0.36	0.024
$J = 6$	79.04	16.81	3.44	0.53	0.048
$J = 7$	80.74	17.05	3.49	0.63	0.071

Table 2. Average percentage of energy of scattering coefficients E_m on frequency-decreasing paths of length m (scatter layers), depending upon the maximum scale J and $L = 8$ filter orientations for the weed class patches.

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$J = 1$	99.90	0.0985	-	-	-
$J = 2$	99.71	0.2798	0.0098	-	-
$J = 3$	99.07	0.8832	0.0443	0.0016	-
$J = 4$	97.55	2.2669	0.1663	0.0080	0.0003
$J = 5$	95.10	4.3892	0.4667	0.0343	0.0020
$J = 6$	92.07	6.8696	0.9522	0.0983	0.0076
$J = 7$	89.26	9.0102	1.5049	0.1979	0.0196

Table 3. Average percentage of energy of scattering coefficients on frequency-decreasing paths of length m (scatter layers), depending upon the maximum scale J and $L = 8$ filter orientations for the plant class patches.

	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 4$
$J = 1$	99.92	0.0711	-	-	-
$J = 2$	99.76	0.2339	0.0040	-	-
$J = 3$	99.17	0.7984	0.0281	0.0003	-
$J = 4$	97.75	2.0899	0.1380	0.0041	0.00003
$J = 5$	95.41	4.1411	0.4215	0.0254	0.0006
$J = 6$	92.34	6.6553	0.9078	0.0892	0.005
$J = 7$	89.37	8.9341	1.4817	0.1944	0.0171

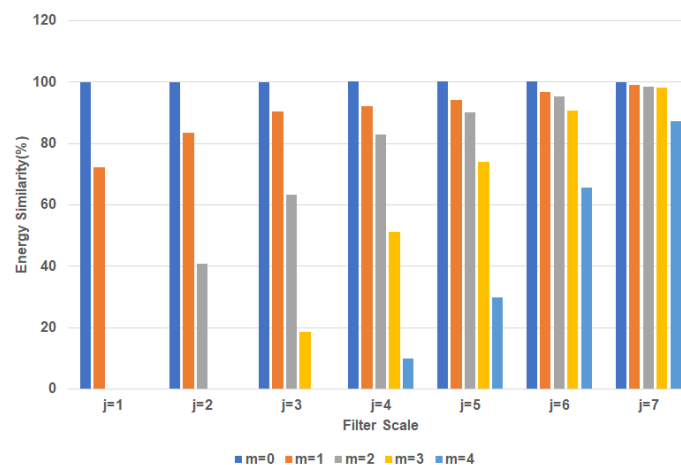


Figure 8. Energy similarity, $Q_m(J)$, between energy of weeds and plants data sets based on Tables 2 and 3.

Also, from the observation of Figure 8, our approach indicates that a priori the best discriminant energy between each class is to be expected with a scatter architecture corresponding to $M = 4$ and $J = 4$ which provides the minimum energy similarity, η , between the energy of images of the weeds' class and the plants' class.

2.4.2. Other Methods

To assess the possible interest of the scatter transform in our weed detection problem, we consider several alternative feature extractor algorithms. First, since the scatter transform by construction works on a feature space which includes multiple scales, it is expected to perform better than any state of the art monoscale method, i.e., working on a feature space tuned on a single size, when applied on a multiple scales problem (such as the one we have here with veins, limb, leaf). Second, since the scatter transform works on a combination of wavelet decomposition between scales it should perform slightly better than a pure wavelet decomposition chosen on the same wavelet basis but without the use of the non-linear operator nor the scales combination. Finally, because scatter transform shares some similarities with convolutional neural networks it should also be compared with the performance obtained with a deep learning algorithm. Based on this rationale, we propose the following alternative feature extractor for comparison with the feature extractor of the scatter transform where the same PCA followed by a linear SVM is used for the classification.

Local binary pattern: Under the original form of [22] and as used in this article, for a pixel positioned at (x, y) , local binary pattern (LBP) indicates a sequential set of the binary comparison of its value with the eight neighbors. In other words, the LBP value assigned to each neighbor is either 0 or 1, if its value is smaller or greater than the pixel placed at the center of the mask, respectively. The decimal form of the resulting 8-bit word representing the LBP code can be expressed as follows

$$LBP(x, y) = \sum_{n=0}^7 2^n s(i_n - i_{x,y}) \quad (5)$$

where $i_{x,y}$ corresponds to the gray value of the center pixel, and i_n denotes that of the n th neighboring one. Besides, the function $\zeta(x)$ is defined as follows

$$\zeta(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0. \end{cases} \quad (6)$$

The LBP operator remains unaffected by any monotonic gray scale transformation which preserves the pixel intensity order in a local neighborhood. It is worth noticing that all the bits of the LBP code hold the same significance level, where two successive bits value may have different implications. The process of Equation (5) is produced at the scale of the patch defined in the previous section. The $LBP(x, y)$ of each pixel inside this patch are concatenated to create a fingerprint of the local texture around the pixel at the center of the patch. Equations (5) and (6) are applied on all patches of an image.

Gray-Level Co-Occurrence Matrix: A statistical approach that can well describe second-order statistics of a texture image is provided by the so-called gray-level co-occurrence matrix (GLCM). GLCM was firstly introduced by Haralick et al. [23]. A GLCM is essentially a two-dimensional histogram in which the (i, j) th element is the frequency of event i co-occurring with event j . A co-occurrence matrix is specified by the relative frequencies $C(i, j, d, \theta)$ in which two pixels, separated by a distance d , occurs in a direction specified by the angle θ , one with gray-level i and the other with gray-level j . A co-occurrence matrix is therefore a function of distance d , angle θ and grayscales i and j .

In our study, as perceptible in images of Figure 2, the weed-plant structures are isotropic meaning that they show no specific predominant orientations. As a logical consequence, and as already stated in similar weed classification problem using GLCM [24–26], choosing multiple orientations θ would not improve the classification performance. We therefore arbitrarily chose a fixed $\theta = 0$ which enables to probe on average leaves positioned in all directions. For distance, d , it is taken at $d = 2$ pixels

which correspond to a displacement capable of probing the presence of edges, veins, and structures in the limb.

Gabor filter: Same Gabor filters as in the scatter transform were applied to the images to produce a feature space. By contrast with the scatter transform, no non-linearities are included in this process and only one layer of filters is applied. For a fair comparison in this experiment, scale range J and number of orientations L of the Gabor filter bank are chosen at the same value as in the scatter transform.

Deep learning: Representation learning, or deep learning, aims at jointly learning feature representations with the required prediction models. We chose the predominant approach in computer vision, namely deep convolutional neural networks [27]. The baseline approach resorts to standard supervised training of the prediction model (the neural network) on the target training data. No additional data sources are used. In particular, given a training set comprised of K pairs of images f_i and labels \hat{y}_i , we train the parameters θ of the network r using stochastic gradient descent to minimize empirical risk:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^K \mathcal{L}(\hat{y}_i, r(f_i, \theta)) \quad (7)$$

\mathcal{L} denotes the loss function, which is cross-entropy in our case. The minimization is carried out using the ADAM optimizer [28] with a learning rate of 0.001.

The architecture of network $r(\cdot, \cdot)$, shown in Figure 9, has been optimized on a hold-out set and is given as follows: five convolutional layers with filters of size 3×3 and respective numbers of filters 64, 64, 128, 128, 256 each followed by ReLU activations and 2×2 max pooling; a fully connected layer with 1024 units, ReLU activation and dropout (0.5) and a fully connected output layer for 2 classes (weeds, plants) and SoftMax activation. Given the current huge interest on deep learning many other architectures could be tested and possibly provide better results. As a disclaimer, we stress that the architecture proposed in Figure 9 is of course not expected to provide the best performance achievable with any neural network architecture. Here the tested CNN serves as a simple reference with a level of complexity of the architecture adapted to the size of the input image and training data sets.

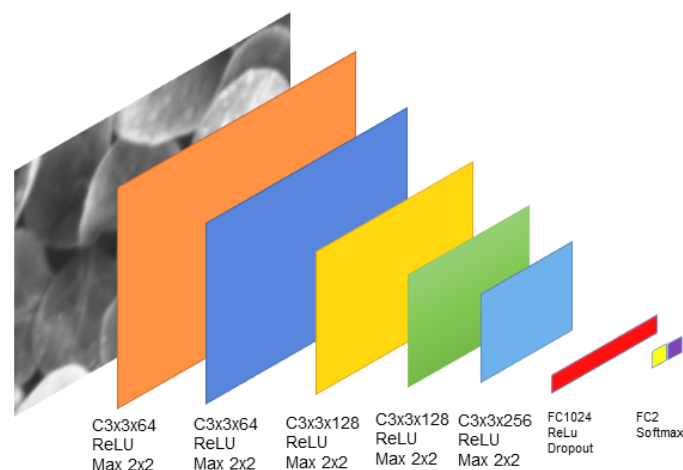


Figure 9. Architecture of the deep network optimized for the task on classification.

3. Result

In this section, we provide experimental results using the experimental protocol for the assessment of scatter transform (Section 2.4) as well as the different alternative feature extraction techniques chosen for comparison in Section 2.4.2.

The scatter transform produces a data vector containing the $Z_m f$ of Equation (2) whose dimension is reduced by a standard PCA and then applied to a linear kernel SVM. To compare the performance of different structures of scatter transform on the database, we used a different combination of filter scales, j , and the number of layers, m , to realize which structure is the best fit for our data. Table 4 shows the classification accuracy of these structures where 10-fold cross-validation approach is used for classification. The best weed/plant classification results with scatter transform are obtained for $J = 4$ and $m = 4$. This a posteriori exactly corresponds to the prediction done a priori from the energy-based approach presented in the method section.

Table 4. Percentage of correct classification for 10-fold cross-validation classification on simulation data with scatter transform for various values of m and J .

	$J = 1$	$J = 2$	$J = 3$	$J = 4$	$J = 5$	$J = 6$	$J = 7$	$J = 8$
$m = 1$	70.37%	77.89%	82.74%	86.17%	88.96%	91.94%	94.14%	95.05%
$m = 2$	—	91.95%	95.26%	95.54%	95.86%	95.82%	95.73%	95.55%
$m = 3$	—	—	95.41%	95.44%	95.21%	95.07%	95.03%	96.00%
$m = 4$	—	—	—	96.31%	96.02%	96.05%	96.16%	96.11%

We considered this optimal scatter transform structure with $J = 4$ and $m = 4$ and compared it with all alternative methods described in Section 2.4. Table 5 shows the recognition rates of weed detection on the data where a k-fold cross-validation approach of SVM classification with the different number of folds is used. Scatter transform appears to outperform all compared handcrafted methods. This demonstrates the interest of the multiscale and combinatory feature space produced by scatter transform. It is important to notice that to have a fair comparison of these alternative methods, we adapted the feature spaces of all algorithms to the same size. The minimum size of the whole feature space is selected, and feature space of other algorithms are reduced to that specific size. In our techniques, the minimum feature space belongs to the GLCM method which has a size of $N \times 19$ where N represents the number of samples. The PCA algorithm is adapted to our models to reduce the dimensions of the feature space generated by other techniques to the size of $N \times 19$.

As shown in Table 5 and Figure 10, when compared with CNN, like most handcrafted methods, scatter transform performs better for small data sets. The limit where CNN and scatter transform are found to perform equally is found to be 10^4 on the weed detection problem as given in Figure 10. This demonstrates the interest of the scatter transform in case of rather small data sets. It is, however, to be noticed that an intrinsic limitation of scatter transform is that it works only with patches to perform a classification while some architectures of convolutional neural network would also be capable of performing segmentation directly in the whole image (see for instance U-Net) [29].

Table 5. Percentage of correct classification by using k-fold Cross-validation on simulated data.

	5 Folds	6 Folds	7 Folds	8 Folds	9 Folds	10 Folds	Average std
Scatter Transform (0.6584×10^4 samples)	94.9%	95.2%	95.3%	95.7%	95.8%	95.8%	± 1.1
LBP (0.6584×10^4 samples)	85.5%	86.1%	86.3%	85.8%	86.9%	86.7%	± 0.4
GLCM (0.6584×10^4 samples)	87.4%	91.6%	90.9%	92.1%	92.4%	92.3%	± 0.7
Gabor Filter (0.6584×10^4 samples)	88.0%	88.2%	88.7%	88.6%	89.4%	89.3%	± 1.3
Deep Learning (0.6584×10^4 samples)	89.4%	89.9%	91.1%	91.5%	91.9%	92.1%	± 1.4
Deep Learning (2.8×10^4 samples)	97.6 %	97.9 %	97.9 %	98.2%	98.1%	98.3%	± 0.9

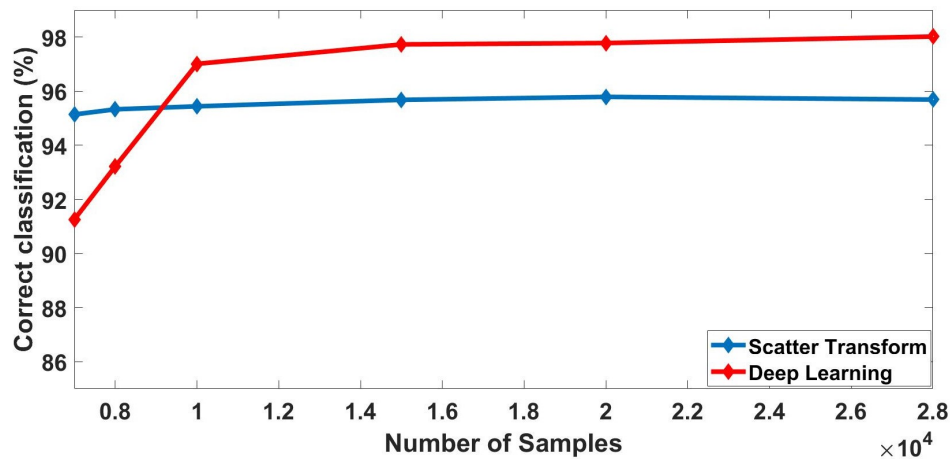


Figure 10. Comparison of the recognition accuracy between scatter transform and deep learning when the number of samples increases.

4. Discussion

So far, we focused in this article on detection of weeds in fields by the scatter transform algorithm with a comparison of other machine-learning techniques which have been trained and tested on synthetic images produced by the simulator of Figure 4. Our experimental results show that a good recognition rate of weeds detection (approximately 95%) can be achievable by the scatter transform algorithm. On the other hand, other alternative methods also work well for this problem with a minimum recognition rate around 85%. These experiments prove that texture-based algorithms can be useful for weed detection in culture crops of high density.

One may wonder how these classification results compare toward the literature on weed detection in less dense culture cited in the introduction section [12–21]. The performance in this literature varies from 75% to 99% of good detection of weed. It is, however, difficult to provide a fair comparison since in addition to the main difference with the absence of soil, the observation scales together with the acquisition conditions vary from one study to another.

One may wonder how these algorithms trained on synthetic data behave when they are applied to real images including plant background and weed not included in the synthetic data sets. We also tested our scatter transform classifier which was trained on synthetic data when applied on the real images of Figure 2. On average for all 10 real images, the accuracy found is 85.64%. Although this constitutes already interesting results, this indicates a bias between simulated data and real data. One direction could be to improve the realism of the simulator. In the version proposed here weeds were not necessarily acquired in the same lighting conditions as the plant. A simple upgrade could be to adapt the average intensity on the weed and the plant to compensate for this artifact or, since in plant and weed can indeed be of various intensity, to generate data augmentation with various contrast. However, simulators never exactly reproduce reality. Another approach to improve the performance of the training based on simulated data would be to add a step of domain adaptation after the scatter transform [30]. So far, the best and worst results obtained with scatter transform are given in Figure 11. A possible interpretation for the rather low performance in Figure 11b is the following. The density of weed in Figure 11b is very high compared to the other images in the training data-set. Consequently, the local texture in the patch may be very different from the one obtained when weeds appear as outliers. This demonstrates that the proposed algorithm, trained on synthetic data, is appropriate in the low density of weeds at an observation scale such as the one chosen for the patch where plant serves as a systematic background.

These performances could be improved in several ways. First, a large variety of weeds can be found in nature and it would be important to include more of this variability in the training data sets. Also, weeds are fast growing plants capable of winning the competition for light. Therefore,

high percentages of weed are expected to come with higher weeds than in very low percentage of the surface of weeds. This fact illustrated in Figure 11 is not included in the simulator where weeds of a fixed size are randomly picked. Such example of enrichment of the training data-set and simulator could be tested easily following the global methodology presented in this article to assess the scatter transform. Finally, we did not pay much effort on denoising the data. The proposed data have been acquired with a camera fixed on an unmanned vehicle. Compensation for variation of illumination in the data-set, or inside the images, themselves or compensation for the possible optical aberration of the camera used could also constitute directions of investigation to improve the weed/plant detection. All the methods presented in this paper (including scatter transform) have the capability to be robust to global variation of light intensity however the variation of light direction during the day may impact the captured textures. Increasing the data-set to acquire images at all hour of a working day or adding a lighting cabinet on the robot used would make the results even more robust [14,31–33].

The problem of weed detection in culture crops of high density is an open problem in agriculture which we believe deserves the organization of a challenge similar to the one organized on Arabidopsis in controlled conditions [34] for a biology community. Such challenges contribute to improving the state of the art as recently illustrated with the use of simulated Arabidopsis data to boost and speed up the training [35] in machine learning. This challenge is now open on the codalab platform (<https://competitions.codalab.org/competitions/20075>) together with the effort of proposing real data and the simulator (<https://uabox.univ-angers.fr/index.php/s/ij0knyzOUgsUV9>) developed for this article. These additional materials, therefore, contributes to the opening of the problem of weed detection in culture crops of high density to a wider computer vision community.

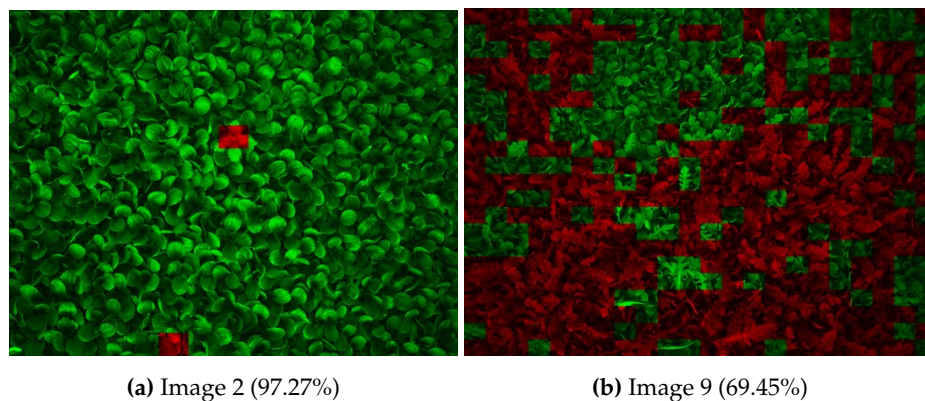


Figure 11. Visual comparison of the best and the worst recognition of weeds and plants by scatter transform.

5. Conclusions and Perspectives

In this article, we proposed the first application of the scatter transform algorithm to plant sciences with the problem of weed detection in a background of culture crops of high density. This open plant science problem is important for field robotics where the mechanical extraction of weed is a current challenge to be addressed to avoid the use of phytochemical products.

We assessed the potential of the scatter transform algorithm in comparison with single scale and multiscale techniques such as LBP, GLCM, Gabor filter, and convolutional neural network. Experimental results showed the superiority of the scatter transform algorithm with a weed detection accuracy of approximately 95% over the other single scale and multiscale techniques on this application. Though the comparison was not intended to be exhaustive among the huge literature on texture analysis, the variety of tested techniques contributes to confirm the effectiveness of using the scatter transform algorithm as a valuable multiscale technique for a problem of weed detection and opened an interesting approach for similar problems in plant sciences. Finally, an optimization method based on energy at the output of the scatter transform has been successfully proposed to select a priori the best scatter transform architecture for a classification problem.

Concerning the weed-plant detection, our optimal solution with scatter transform can serve as a first reference of performance and other machine-learning techniques could now be tested in the framework of the data challenge that we launched for this article (<https://competitions.codalab.org/competitions/20075>). As a possible perspective of the investigation, one could further optimize the scatter transform classifier proposed in this paper. For instance, the size of the grid could be fine-tuned or some hyperparameters could be added with non-linear kernels in the SVM step. Also, weed/plant detection was focused here on a binary classification since no distinction between the different weeds were included. In another direction, one could also envision to extend this work to a multiple types of weeds classification problem if more data were included.

Author Contributions: Conceptualization, P.R., A.A. and D.R.; Data curation, A.A. and S.S.; Formal analysis, E.B.; Funding acquisition, E.B. and D.R.; Methodology, D.R.; Resources, E.B.; Software, P.R., A.A. and S.S.; Supervision, D.R.; Validation, P.R., S.S. and D.R.; Visualization, P.R.; Writing—original draft, P.R. and D.R.

Funding: This research received no external funding.

Acknowledgments: Acquisitions of real-images were done in the framework of the project PUMAGri, supported from the Fonds Unique Interministeriel (FUI-BPI France). Authors thank Sixin Zhang from École Normale Supérieure, Paris France for useful discussions. Salma Samiei acknowledges Angers Loire Métropole for the funding of her PhD.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janowczyk, A.; Madabhushi, A. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *J. Pathol. Informat.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
2. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [[CrossRef](#)]
3. Bruna, J.; Mallat, S. Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1872–1886. [[CrossRef](#)] [[PubMed](#)]
4. Minaee, S.; Abdolrashidi, A.; Wang, Y. Iris recognition using scattering transform and textural features. In Proceedings of the Signal Processing and Signal Processing Education Workshop (SP/SPE), Salt Lake City, UT, USA, 9–12 August 2015; pp. 37–42.
5. Lagrange, M.; Andrieu, H.; Emmanuel, I.; Busquets, G.; Loubrié, S. Classification of rainfall radar images using the scattering transform. *J. Hydrol.* **2018**, *556*, 972–979. [[CrossRef](#)]
6. Li, B.H.; Zhang, J.; Zheng, W.S. HEp-2 cells staining patterns classification via wavelet scattering network and random forest. In Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia, 3–6 November 2015; pp. 406–410.
7. Rakotomamonjy, A.; Petitjean, C.; Salaün, M.; Thiberville, L. Scattering features for lung cancer detection in fibered confocal fluorescence microscopy images. *Artif. Intell. Med.* **2014**, *61*, 105–118. [[CrossRef](#)] [[PubMed](#)]
8. Yang, X.; Huang, D.; Wang, Y.; Chen, L. Automatic 3d facial expression recognition using geometric scattering representation. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; Voume 1, pp. 1–6.
9. Torres-Sánchez, J.; López-Granados, F.; De Castro, A.I.; Peña-Barragán, J.M. Configuration and specifications of an unmanned aerial vehicle (UAV) for early site specific weed management. *PLoS ONE* **2013**, *8*, e58210. [[CrossRef](#)] [[PubMed](#)]
10. Peña, J.M.; Torres-Sánchez, J.; Serrano-Pérez, A.; de Castro, A.I.; López-Granados, F. Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors* **2015**, *15*, 5609–5626. [[CrossRef](#)]
11. Fernández-Quintanilla, C.; Peña, J.; Andújar, D.; Dorado, J.; Ribeiro, A.; López-Granados, F. Is the current state of the art of weed monitoring suitable for site-specific weed management in arable crops? *Weed Res.* **2018**. [[CrossRef](#)]
12. Bakhshipour, A.; Jafari, A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput. Electron. Agric.* **2018**, *145*, 153–160. [[CrossRef](#)]

13. Lottes, P.; Khanna, R.; Pfeifer, J.; Siegwart, R.; Stachniss, C. UAV-based crop and weed classification for smart farming. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3024–3031.
14. Milioto, A.; Lottes, P.; Stachniss, C. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018; pp. 2229–2235.
15. Aitkenhead, M.; Dalgetty, I.; Mullins, C.; McDonald, A.J.S.; Strachan, N.J.C. Weed and crop discrimination using image analysis and artificial intelligence methods. *Comput. Electron. Agric.* **2003**, *39*, 157–171. [[CrossRef](#)]
16. Marchant, J.; Onyango, C. Comparison of a Bayesian classifier with a multilayer feed-forward neural network using the example of plant/weed/soil discrimination. *Comput. Electron. Agric.* **2003**, *39*, 3–22. [[CrossRef](#)]
17. Prema, P.; Murugan, D. A novel angular texture pattern (ATP) extraction method for crop and weed discrimination using curvelet transformation. *ELCVIA Electron. Lett. Comput. Vis. Image Anal.* **2016**, *15*, 27–59. [[CrossRef](#)]
18. Ahmad, A.; Guyonneau, R.; Mercier, F.; Belin, É. An Image Processing Method Based on Features Selection for Crop Plants and Weeds Discrimination Using RGB Images. In *International Conference on Image and Signal Processing*; Springer: Berlin, Germany, 2018; pp. 3–10.
19. Haug, S.; Michaels, A.; Biber, P.; Ostermann, J. Plant classification system for crop/weed discrimination without segmentation. In Proceedings of the 2014 IEEE Winter Conference on Applications of Computer Vision (WACV), Steamboat Springs, CO, USA, 24–26 March 2014; pp. 1142–1149.
20. Bakhshipour, A.; Jafari, A.; Nassiri, S.M.; Zare, D. Weed segmentation using texture features extracted from wavelet sub-images. *Biosyst. Eng.* **2017**, *157*, 1–12. [[CrossRef](#)]
21. Bossu, J.; Gée, C.; Jones, G.; Truchetet, F. Wavelet transform to discriminate between crop and weed in perspective agronomic images. *Comput. Electron. Agric.* **2009**, *65*, 133–143. [[CrossRef](#)]
22. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [[CrossRef](#)]
23. Haralick, R.M.; Shanmugam, K.; Its'Haik, D. Textural features for image classification. *IEEE Trans. Syst. Man Cybernet.* **1973**, *6*, 610–621. [[CrossRef](#)]
24. Shearer, S.A.; Holmes, R. Plant identification using color co-occurrence matrices. *Trans. ASAE* **1990**, *33*, 1237–1244. [[CrossRef](#)]
25. Burks, T.; Shearer, S.; Payne, F. Classification of weed species using color texture features and discriminant analysis. *Trans. ASAE* **2000**, *43*, 441. [[CrossRef](#)]
26. Chang, Y.; Zaman, Q.; Schumann, A.; Percival, D.; Esau, T.; Ayalew, G. Development of color co-occurrence matrix based machine vision algorithms for wild blueberry fields. *Appl. Eng. Agric.* **2012**, *28*, 315–323. [[CrossRef](#)]
27. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
28. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980 .
29. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin, Germany, 2015; pp. 234–241.
30. Courty, N.; Flamary, R.; Tuia, D.; Rakotomamonjy, A. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1853–1865. [[CrossRef](#)] [[PubMed](#)]
31. Slaughter, D.; Giles, D.; Downey, D. Autonomous robotic weed control systems: A review. *Comput. Electron. Agric.* **2008**, *61*, 63–78. [[CrossRef](#)]
32. Fadlallah, S.; Goher, K. A review of weed detection and control robots: A world without weeds. In *Advances in Cooperative Robotics*; World Scientific: Singapore, 2017; pp. 233–240.
33. Brown, R.B.; Noble, S.D. Site-specific weed management: sensing requirements—What do we need to see? *Weed Sci.* **2005**, *53*, 252–258. [[CrossRef](#)]

34. Scharr, H.; Minervini, M.; Fischbach, A.; Tsafaris, S.A. Annotated image datasets of rosette plants. In Proceedings of the European Conference on Computer Vision, Zürich, Switzerland, 6–12 September 2014; pp. 6–12.
35. Ubbens, J.; Cieslak, M.; Prusinkiewicz, P.; Stavness, I. The use of plant models in deep learning: An application to leaf counting in rosette plants. *Plant Methods* **2018**, *14*, 6. [[CrossRef](#)] [[PubMed](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).