AAAS
Plant Phenomics
Volume 2022, Article ID 9803570, 17 pages
https://doi.org/10.34133/2022/9803570

Plant Phenomics
A SCIENCE PARTNER JOURNAL

Research Article

SegVeg: Segmenting RGB Images into Green and Senescent Vegetation by Combining Deep and Shallow Methods

Mario Serouart,[1,2] **Simon Madec**,[1,3] **Etienne David**,[1,2,4] **Kaaviya Velumani**,[4]
Raul Lopez Lozano,[2] **Marie Weiss**,[2] **and Frédéric Baret**[2]

[1]*Arvalis, Institut du végétal, 228, route de l'aérodrome - CS 40509, 84914 Avignon Cedex 9, France*
[2]*INRAE, Avignon Université, UMR EMMAH, UMT CAPTE, 228, route de l'aérodrome - CS 40509, 84914 Avignon Cedex 9, France*
[3]*CIRAD, UMR TETIS, F-34398 Montpellier, France*
[4]*Hiphen SAS, 228, route de l'aérodrome - CS 40509, 84914 Avignon Cedex 9, France*

Correspondence should be addressed to Mario Serouart; mario.serouart@inrae.fr

Received 15 March 2022; Accepted 26 August 2022; Published 12 November 2022

Pixel segmentation of high-resolution RGB images into chlorophyll-active or nonactive vegetation classes is a first step often required before estimating key traits of interest. We have developed the SegVeg approach for semantic segmentation of RGB images into three classes (background, green, and senescent vegetation). This is achieved in two steps: A U-net model is first trained on a very large dataset to separate whole vegetation from background. The green and senescent vegetation pixels are then separated using SVM, a shallow machine learning technique, trained over a selection of pixels extracted from images. The performances of the SegVeg approach is then compared to a 3-class U-net model trained using weak supervision over RGB images segmented with SegVeg as groundtruth masks. Results show that the SegVeg approach allows to segment accurately the three classes. However, some confusion is observed mainly between the background and senescent vegetation, particularly over the dark and bright regions of the images. The U-net model achieves similar performances, with slight degradation over the green vegetation: the SVM pixel-based approach provides more precise delineation of the green and senescent patches as compared to the convolutional nature of U-net. The use of the components of several color spaces allows to better classify the vegetation pixels into green and senescent. Finally, the models are used to predict the fraction of three classes over whole images or regularly spaced grid-pixels. Results show that green fraction is very well estimated ($R^2 = 0.94$) by the SegVeg model, while the senescent and background fractions show slightly degraded performances ($R^2 = 0.70$ and 0.73, respectively) with a mean 95% confidence error interval of 2.7% and 2.1% for the senescent vegetation and background, versus 1% for green vegetation. We have made SegVeg publicly available as a ready-to-use script and model, along with the entire annotated grid-pixels dataset. We thus hope to render segmentation accessible to a broad audience by requiring neither manual annotation nor knowledge or, at least, offering a pretrained model for more specific use.

1. Introduction

The vegetation fraction (VF) is a key trait that drives the partitioning of radiation between the background and the vegetation. It is used in several studies as a proxy of crop state [1] and yield [2, 3]. The complement to unity of VF is the gap fraction that is used to estimate the plant area index. However, several ecophysiological processes such as photosynthesis and transpiration are driven by the amount of green surfaces that exchange mass and energy with the atmosphere. More specifically, the green fraction (GF) is used to estimate the green area index (GAI) [4] defined as the area of green vegetation elements per unit horizontal ground area. GF is a more relevant trait that should be used when describing crop functioning [5]. The difference between VF and GF is the senescent fraction (SF = VF − GF), sometimes called the nonphotosynthetic fraction [6, 7]. For crops, SF depends on both the growth stage and state of the plants. The SF trait is used to characterize a biotic or abiotic stress, describe nutrient recycling, and monitor the ageing process

[8–10]. Some studies have demonstrated the ability of genotypes to stay green by delaying senescence and potentially improve productivity [11, 12].

Several remote sensing methods have been developed to estimate GF and SF using the spectral variation of the signal observed at the canopy scale from metric to decametric resolution [13]. VF, GF, and SF can be also computed using very high spatial resolution images with pixel sizes from a fraction of mm to cm, i.e., significantly smaller than the typical dimension of the objects (plants, organs). RGB cameras with few to tens of millions of pixels are currently widely used as noninvasive high-throughput techniques applied to plant breeding, farm management, and yield prediction [14–16]. These cameras are borne on multiple platforms, including drones [17], ground vehicles [18], and handheld systems [19], or set on a fixed pod [16].

Several methods have been proposed to identify the green pixels in RGB images including thresholding color indices [20] and machine learning classification [21] based on few color space representations. However, these techniques are limited at least by one of the two main factors:

(i) Confounding effects: depending on the illumination conditions and on the quality of the camera optics, part of the soil may appear green due to chromatic aberration. Further, parts of the image that are saturated, with strong specular reflection or very dark, will be difficult to classify using only the color of the pixel. Finally, the soil may also appear greenish when it contains algae [22]

(ii) Continuity of colors: at the cellular scale, senescence results from the degradation of pigments that generally precedes cell death [23]. During the degradation process, changes in the pigment composition result into a wide palette of leaf color in RGB imagery, with a continuity between "green" and "senescent" states. Further, when pixels are located at the border of an organ, its color will be intermediate between organ and background. This problem is obviously enhanced when the spatial resolution of the RGB image is too coarse

It is therefore difficult to segment accurately and robustly the green vegetation parts of a RGB image using only the color information of pixels. Same limitations apply to the segmentation of the senescent vegetation parts. In addition, crop residues located in background areas are difficult to distinguish from the senescent vegetation observed on standing plants with very similar range of brownish colors. Textural and contextual information should therefore be exploited to better segment RGB images into green and senescent vegetation parts.

Semantic segmentation [24] that assigns a class to each pixel of the image appears to be an attractive approach. It is based on deep learning techniques and has been applied to several domains including urban scene description for autonomous vehicles, medical imagery [25], and agriculture [26, 27]. However, images need to be labelled exhaustively into several target classes, which requires large annotation resources [28].

The objective is then to develop and evaluate a two-step semantic segmentation approach called SegVeg. It labels each pixel of very high-resolution RGB images of vegetation scenes into three classes: background, green, and senescent vegetation. It has been designed to reduce the annotation effort by combing a convolutional neural network (CNN) that splits image into vegetation (including both green and senescent pixels) and background, to a simple support vector machine (SVM) technique that classifies the vegetation pixels into green and senescent. SegVeg will be compared to a CNN classifier that directly identifies background, green, and senescent vegetation pixels following a weak supervised training principle.

2. Materials and Methods

As shown in Figure 1, this study investigates two approaches to segment images in three classes: green, senescent, and background.

The first step consists of developing the SegVeg method that combines a binary U-net model (U-net 2C) to first separate vegetation from background. Then a SVM model will separate green from senescent vegetation, once the whole vegetation is extracted.

This stage relies on two training datasets: fully annotated patches (Dataset #1 with 2-class entire masks) for the U-net 2C training and pixel labelled datasets for the SVM approach (Dataset #2). Once the SegVeg approach is set, it is used to build a third dataset of fully nonsupervised annotated patches (Dataset #3) and train a 3-class U-net model (U-net 3C) on the same RGB images present in Dataset #1. The SegVeg and 3-class U-net performances are then compared.

2.1. The SegVeg Approach.

The SegVeg approach is made of two stages (Figures 1 and 2). In the first stage, the whole image is classified into vegetation and background mask using a U-net type deep learning network [29]. Then, vegetation pixels (predicted from the first stage) are classified into green and senescent vegetation using a SVM. The two binary outputs of each model are then merged to form a 3-class mask.

2.1.1. First Stage: Vegetation and Background Segmentation.

U-net is a deep learning model with encoder-decoder network architecture that is widely used for image semantic segmentation. The model was trained over the labelled images from Dataset #1 (Section 2.3.1) to predict two classes: vegetation (green and/or senescent) and background. EfficientNet-B2 architecture [30] with weights initialized on ImageNet was used as the backbone architecture. Patches of 512×512 pixels were used for training after data augmentation based on the Albumentations library [31]. The training process was based on a Dice loss function with an Adam optimizer.

A predefined decaying learning rate schedule (step based) was used to reach local minima, with an initial value of 0.01 and reaching at the end $10e-6$, which is an usual range in standard multilayer neural networks studies [32].

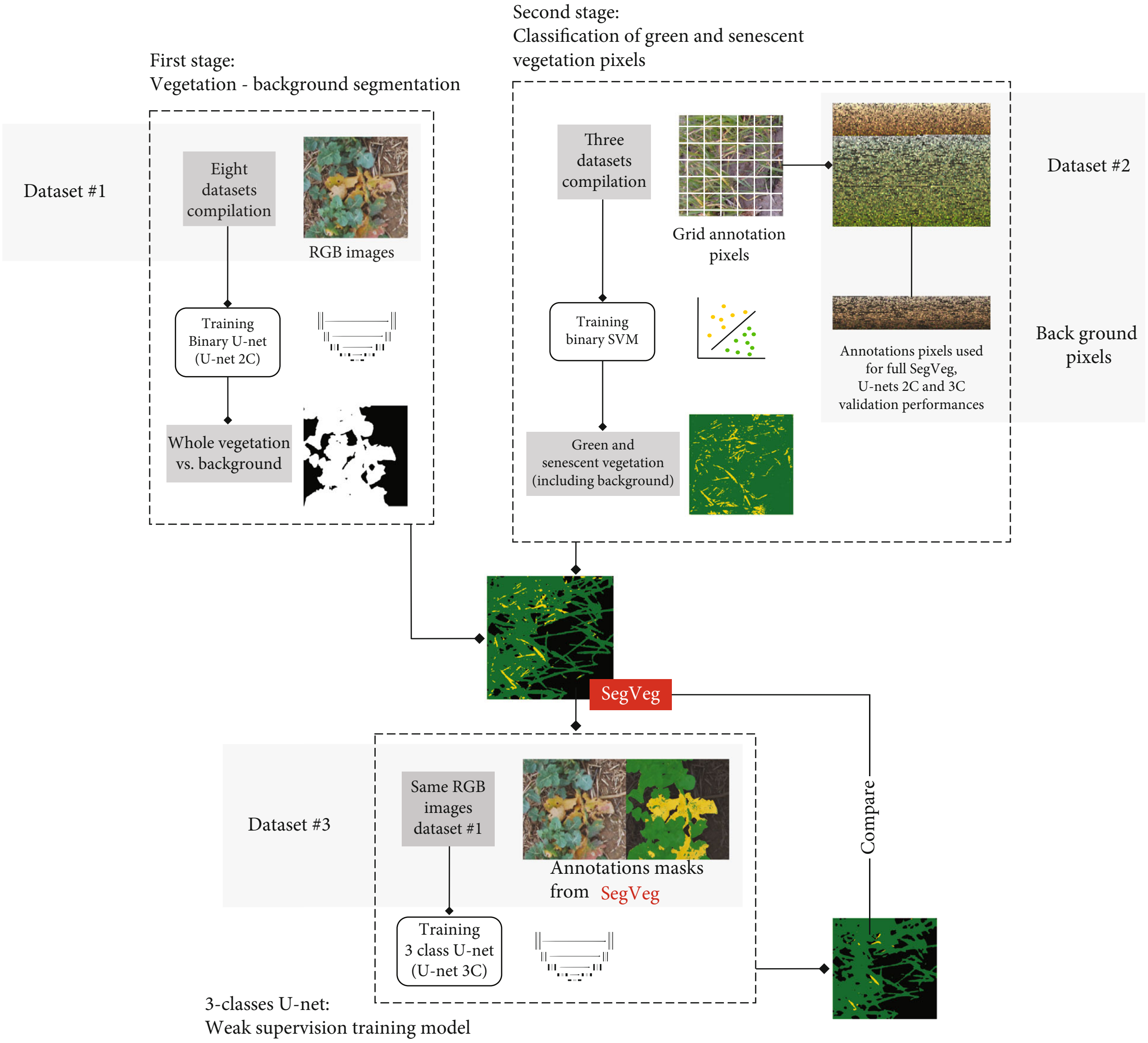

Second stage:
Classification of green and senescent vegetation pixels
First stage:
Vegetation - background segmentation
Dataset #1
Eight datasets compilation
RGB images
Training Binary U-net (U-net 2C)
Whole vegetation vs. background
Three datasets compilation
Grid annotation pixels
Dataset #2
Training binary SVM
Annotations pixels used for full SegVeg, U-nets 2C and 3C validation performances
Back ground pixels
Green and senescent vegetation (including background)
SegVeg
Compare
Dataset #3
Same RGB images dataset #1
Annotations masks from SegVeg
Training 3 class U-net (U-net 3C)
3-classes U-net:
Weak supervision training model

Figure 1: Flowchart describing the overall approach of the study.

The minibatch size was set to 32 for computational purpose. Finally, early stopping was implemented to set number of training iterations. The Python Segmentation Models library under PyTorch was used [33] with GPU activation (GeForce RTX 3090).

2.1.2. Second Stage: Classification of Green and Senescent Vegetation Pixels. The support vector machine (SVM) is an efficient machine learning classification method widely used for image segmentation [34–36]. It maps the original features to some higher-dimensional space where the training dataset is separable. Several color spaces and transformations [37] were used to classify green and senescent pixels including RGB, HSV, CIELab, grayscale, luminances, CMYK, YCbCr, and YIQ derived from the original RGB values. A total of 23 potential input features were thus computed, namely R, G, and B; H, S, and V; L, a, and b; GE; LA, LB, and LC; C, M, Y, and K; Yi, Cb, and Cr; and Yj, I, and Q. However, the possible redundancy and irrelevancy of some features may decrease the accuracy of the classification. We then selected the most appropriate inputs using the step forward wrapper method [38]. Finally, 14 input features were retained: R, G, B, H, S, a, b, GE, M, YE, Cb, Cr, I, and Q.

This second-stage SVM was 4calibrated over labelled pixels from Dataset #2 (see Section 2.3.2). The hyperparameters were tuned using a grid search algorithm following a leave-one-out cross-validation principle. This process led to the optimal values C: 1 and γ: 10^{-3}, and kernel *rbf* was set according to prior knowledge that data are not linearly separable. Scikit 0.23.2 with Python 3.7 was used for implementation [39].

2.2. The 3-Class U-net Model (U-net 3C). A three-class U-net model was used as a reference to evaluate the proposed SegVeg approach (Figure 1).

However, due to the unavailability of a dataset containing entire images annotated into three classes (background, green, and senescent vegetation), we prepared 3-class masks

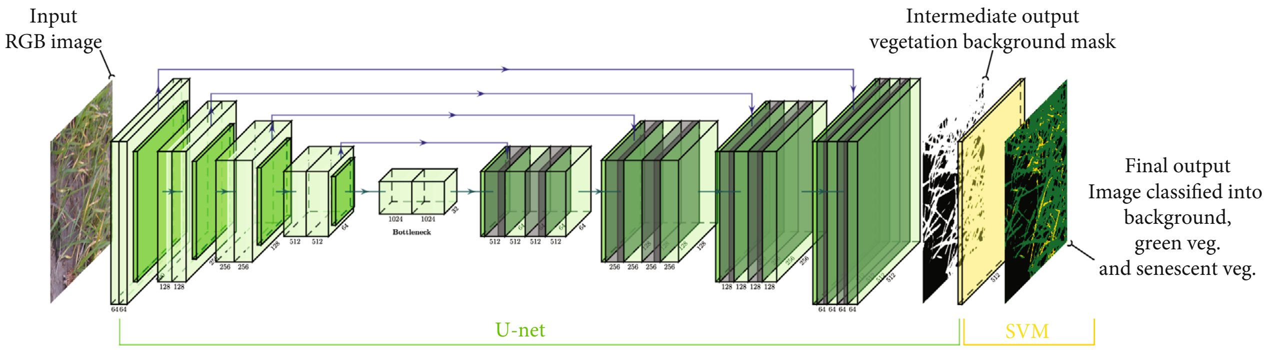


FIGURE 2: Illustration of the SegVeg architecture inputs and outputs. The first stage is a U-net model that predicts vegetation and background masks. The second stage is a SVM that classifies the vegetation mask into green and senescent pixels. The two stages were trained over two independent datasets.

TABLE 1: Characteristics of the subdatasets composing the final dataset.

Subdatasets	Country	Year	Crops	Stage	Reference
UTokyo	Japan	2019 2012	Rice, wheat	Vegetative	[40, 41]
P2S2	France	2018	Wheat, rapeseed, sugar beet, and potato	All	[42]
	Belgium		Maize, grassland, sunflower, rice, and soya		
Wuhan	China	2012 2015	Cotton, maize, and rice	Vegetative	[43]
CVPPP 1 and 2	Italy	2012 2013	Arabidopsis, tobacco	All	[44]
GEVES	France	2020	Maize	Vegetative	—
Phenofix	France	2020	Maize	All	—
Phenomobile	France	2020	Wheat	Early	—
Bonirob	Germany	2016	Sugar beet	Early	[45]

by applying SegVeg over the RGB images used to train U-net 2C (i.e., Dataset #1). Indeed, to reduce the annotation effort, the second-stage SVM was trained over pixels extracted from regularly spaced grids, explained in the following dataset sections. Therefore, no manually annotated 3-class masks were available as groundtruth references.

The same U-net architecture and hyperparameters used for U-net 2C of the SegVeg approach were also employed here during training.

2.3. Training and Testing Datasets

2.3.1. Dataset #1: Vegetation and Background Fully Annotated Patches. Eight subdatasets from previous studies were compiled to get a wide range of acquisition conditions, species, crop states, and stages (Table 1).

The images were acquired with several cameras equipped with different focal length optics and variable distances from the ground. All blurred images or those with poor quality were excluded from our study. The original images were then split into several square patches of 512 × 512 pixels, a size selected to keep sufficient context. A total of 2015 patches were extracted, showing a large diversity as illustrated in Table 2. The ground sampling distance (GSD) ranges were between 0.3 and 2 mm to capture enough details (Figure 3).

Considering that image annotation is time consuming, it was subcontracted to a private company, imageannotation.ai. Each original image was carefully segmented by several operators into vegetation (green and senescent combined) and background pixels. We then verified the resulting classified images and reannotated the few wrongly annotated ones.

2.3.2. Dataset #2: Green, Senescent, and Background Annotated Pixels. Dataset #2 is composed of annotated pixels only, extracted from images on which we have affixed regular square matrix (grids) of 8 to 11 pixels.

This dataset was used to train and test the SVM stage of the SegVeg approach (on green and senescent pixels). After adding the background pixels, it was also used to evaluate the performances of both the SegVeg and the U-nets (2C and 3C).

(1) Image Acquisition and Extraction. Three independent datasets (LITERAL, PHENOMOBILE, and P2S2) were used to train and evaluate the proposed methods.

Table 2: Characteristics of the subdatasets used to compose the training dataset. UGV means unmanned ground vehicle.

Subdatasets	Platform	Camera	Image size (px)	Distance to ground (m)	GSD (mm)	No. of images
UTokyo	Gantry	Canon EOS Kiss X5	5184 × 3456	1.5-1.8	0.2-0.6	534
		Garden Watch Camera	1280 × 1024			
P2S2	Handheld	SONY ILCE-5000	5456 × 3632	2	0.5	170
		SONY ILCE-6000	6000 × 5000			
		Canon EOS 400D	3888 × 2592			
		Canon EOS 60D	5184 × 3456			
		Canon EOS 750D	6000 × 4000			
Wuhan	Gantry	Olympus E-450	3648 × 2736	0.3-5	0.4-0.5	343
CVPPP 1 & 2	Gantry	Canon PowerShotSD1000	3108 × 2324	1	0.1-0.3	752
GEVES	Handheld	SAMSUNG SM-A705FN	3264 × 1836	2	0.2	50
Phenofix	Gantry	SONY RX0 II	4800 × 3200	2	0.6	30
Phenomobile	UGV	SONY RX0 II	4800 × 3200	1.7	0.8-1.4	76
Bonirob	UGV	JAI AD-130GE	1296 × 966	0.85	0.3	60
Total						2015

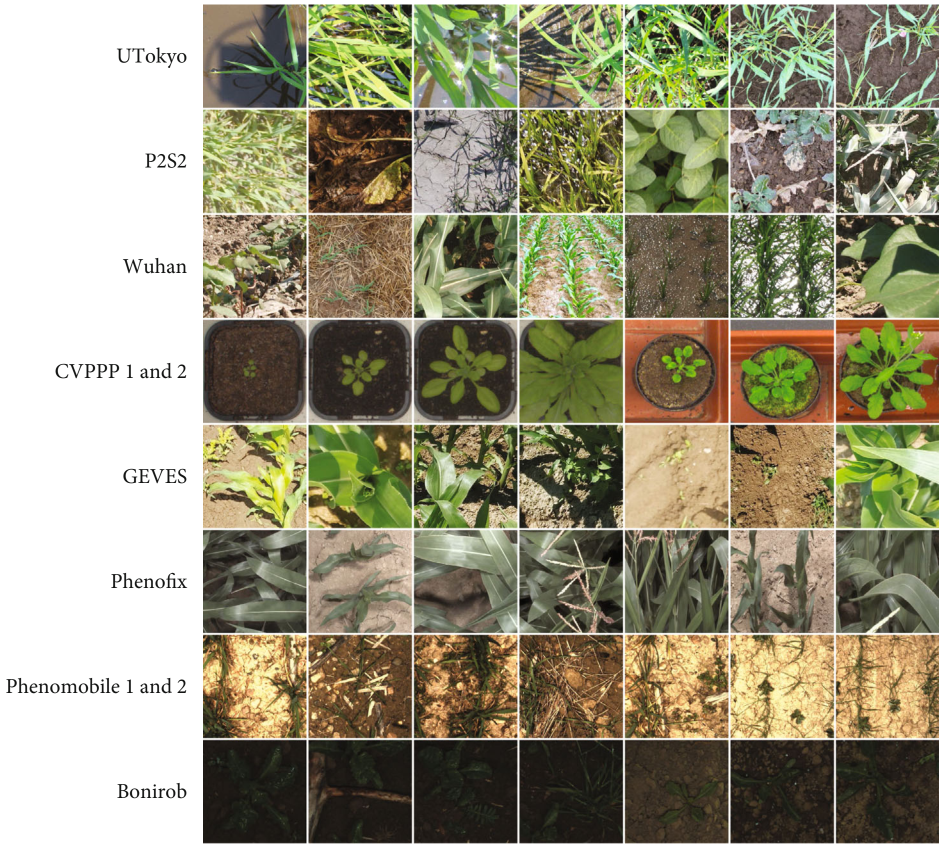


Figure 3: Sample of 512 × 512 pixels patches extracted from the eight subdatasets (Dataset #1).

(i) The LITERAL dataset was acquired with a handheld system called LITERAL (Figure 4). An operator maintains a boom with a Sony RX0 camera fixed at its extremity. The camera faced the ground from nadir at an approximately fixed distance (Table 3). The 68 available annotated images covered a wide range of wheat genotypes grown at several locations in France, representing different growth stages, soil backgrounds, and illumination conditions

(ii) The PHENOMOBILE dataset was acquired with the Phenomobile system (Figure 4), an unmanned

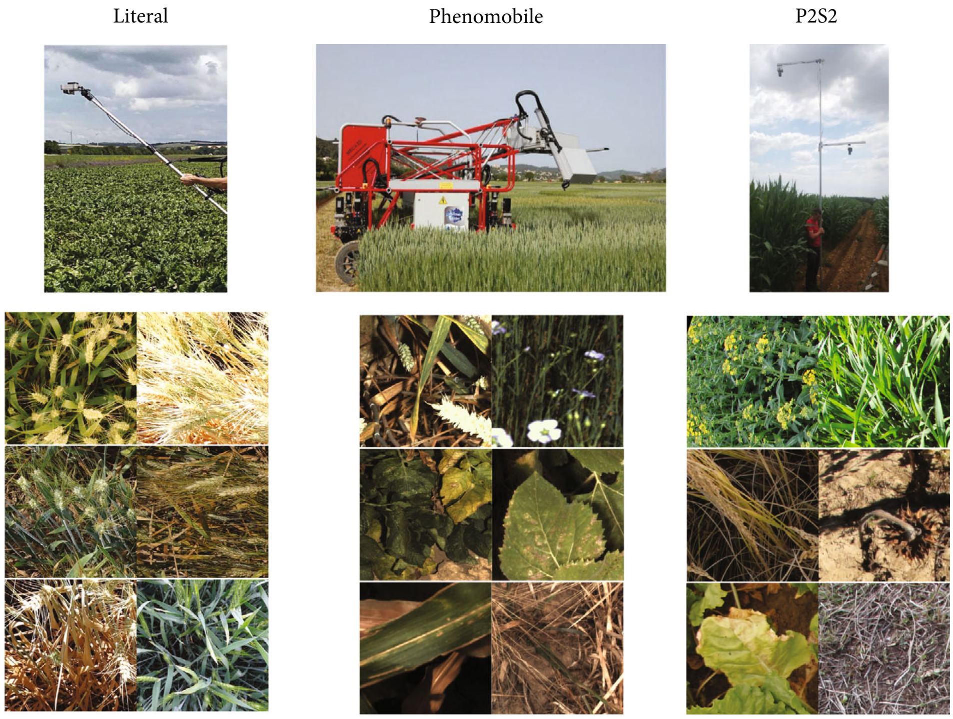


Figure 4: The acquisition systems used for the three independent datasets in Dataset #2: LITERAL, PHENOMOBILE, and P2S2 and their respective examples of 512 × 512 images patches extracted from the three systems.

Table 3: Second-stage dataset description.

Datasets	LITERAL	PHENOMOBILE	P2S2
Latitude, longitude	43.7° N, 5.8° E 49.7° N, 3.0° E 43.5° N, 1.5° E	43.7° N, 6.7° E 47.4° N, 2.3° E 43.7° N, 5.8° E 43.4° N, 0.4° W	43.6° N, 4.5° E 43.4° N, 1.2° E 48.3° N, 2.4° E 50.6° N, 4.7° E
Year	2017-2020	2018-2020	2018
Crops	Wheat	Wheat, sunflower, sugar beet, maize, potato, and flax	Wheat, sunflower, sugar beet, maize, potato, rapeseed, grassland, rice, and soya
Vector	Handheld	Phenomobile	Handheld
Focal length (mm)	8	16–25	50
Camera	Sony RX0 II	Baumer VCXG-124C	ILCE-6000 SONY Canon EOS 750D
Image size (pixels)	4800 × 3200	4096 × 3000	6000 × 4000 3888 × 2592
Pixel size (μm)	2.74	3.45	3.72
Distance to ground (m)	1.5–2.5	2–4.5	1.5–2
GSD (mm)	0.65	1.3	0.5

ground vehicle [46]. This system uses flashes for image acquisition making the measurements independent of the natural illumination conditions. Images are acquired from nadir at a fixed distance from the top of the canopy (Table 3). The 173 available annotated images covered six crops grown in four phenotyping platforms in France (Table 3)

(iii) The P2S2 dataset is composed of 200 hemispherical and nadir images. The acquisition was designed to provide a large dataset over a wide range of crops, observed under contrasted growth conditions, throughout the crop growth cycle, covering crucial phenological stages. More details on the dataset can be found in [42]

Several cameras were used for the acquisition of the three datasets, resulting in differences in image quality and GSD (Table 4). Note that the GSD of this dataset (Table 4) is consistent with that of the previous dataset (Table 2). A total of 441 images of 512 × 512 pixels were finally selected to represent a wide diversity (Figure 3).

Table 4: Distribution of labelled pixel for the three datasets.

Datasets	No. of labelled images	No. of labelled pixels	% classes					
			Green veg.	Sen. Veg.	Background	Green/Sen. Veg. unsure	Unknown	Other
LITERAL	68	4260	46.5	15.8	15.0	13.1	9.5	0.1
PHENOMOBILE	173	8266	40.3	31.1	27.6	0.1	0.8	0.1
P2S2	200	18559	43.6	16	15.2	11.1	13	1.1
Total	441	31085	43.4	20.5	19.75	8.1	7.8	0.45

Table 5: Distribution of the labelled pixels into the training and testing datasets. Only the pixels labelled as *Green Veg.* and *Sen. Veg.* were used for the SVM SegVeg training.

Datasets	No. of labelled pixels	% classes		No. of pixels train	No. of pixels test	% train	% test
		Green Veg.	Sen. Veg.				
LITERAL	2655	75	25	0	2655	0	100
PHENOMOBILE	5883	60	40	1803	4080	30	70
P2S2	11200	75	25	4329	6871	39	61
Total	19738	70	30	6132	13606	32	68

(2) Pixel Labelling. The previously mentioned pixel grids were classified into one of the following six classes, namely, *green vegetation*, *senescent vegetation*, *background*, *green/senescent vegetation unsure*, *unknown*, and *others*. This allowed us to remove pixels with uncertain annotations and potential bias in the training phase. The *green/senescent vegetation unsure*, *unknown*, and *others* were for instance not used in the training and evaluation of the proposed models. However, because of the complexity, subjectivity, and time required to assign pixels into the six classes listed above, the annotation was limited to a small number of pixels per patches (i.e., not building full 3-classes ground-truth masks). This sampled annotation is possible because the second stage of SegVeg (shallow machine learning SVM method) does not require context or local information and therefore not demanding entire patches to be exhaustively annotated. We used a grid displayed on each 512 × 512 images, where the pixels to be classified were located at the intersection of the grid points. A video recording the annotation process of a few pixels is available in Supplementary Material (figure S1). The regular square matrix can vary from 8 to 11 pixels on a side, depending on images. The web based platform, Datatorch [47], was used by 2 annotators. A second round of pixel labelling was performed by 2 other reviewers to find a better consensus on the uncertain pixels and to avoid potential bias in building Dataset #2.

Among the 441 annotated images (Table 4), the unsure classes represented about 16% of the total number of pixels. It can be noticed that for the PHENOMOBILE dataset, the use of integrated flashes during image acquisition provided better pixel interpretation leading to fewer confusions. This dataset is publicly available on Zenodo and can be accessed by following the guidelines at this link https://github.com/mserouar/SegVeg.

(3) Split between Training and Testing Datasets. A total of 19,738 pixels were finally available to perform the training and testing of the SegVeg SVM stage, of which 6132 were used for training and 13,606 for testing (Table 5). Note that for the evaluation of U-net approaches (2C and 3C), the test Dataset #2 evolves by adding the almost 6000 background pixels annotated from the grid (Figure 1, Supplementary Material figure S1), which are naturally absent in the green/senescent SVM training and evaluation.

The LITERAL dataset that represented only a small fraction of the available patches over wheat crops was kept entirely for testing. The PHENOMOBILE dataset was split randomly into training (30%) and testing (70%) dataset (Table 5), resulting in 1803 pixels used to train the SVM model. Similarly, P2S2 was randomly split into 4329 pixels for training (about 40%) and the remaining for testing. This allows to get a balanced distribution between the contributions of PHENOMOBILE and P2S2 datasets to the training process as well as maintain a balanced Green/Senescent pixels fraction. The splitting scheme was chosen according to the concrete theoretical foundation of the SVM algorithm. SVMs are usually not chosen for large-scale data studies because their training complexity is highly dependent on the dataset size (quadratic to the number of observations), which also comes with calculation time issues [48–50]. Moreover, the concept of hyperplane and margins does not require a lot of observations during training, and adding observations could lead to poor generalisation properties. A big amount of initial data was hence kept for the validation step, to ensure robustness in predictions and model performances.

2.4. Evaluation Metrics. Since semantic segmentation classifies each individual pixel, three standard classification metrics derived from the confusion matrix were used to quantify the performances of the methods at the class level: precision, recall, and $F1$-score (Table 6). Further, the overall accuracy and overall $F1$-score were also computed to get a more global evaluation of the segmentation performances

Table 6: Metrics used to evaluate the performances of the models.

Metrics	Name	Definition
True positive	Tp_{class}	Number of pixels well predicted in the given class
True negative	Tn_{class}	Number of pixels well predicted as not in the given class
False positive	Fp_{class}	Number of pixels wrongly predicted in the given class (confusion)
False negative	Fn_{class}	Number of pixels wrongly predicted as not in the given class (missing pixels)
Precision	Prec_{class}	$Tp/Tp + Fp$
Recall	$\mathrm{Re\ c}_{class}$	$Tp/(Tp + Fn)$
Accuracy	Acc_{class}	$((Tp + Tn)/(Tp + Tn + Fp + Fn)) \times 100$
$F1$-score	$\mathrm{F1}_{class}$	$((2 \times Tp)/(2 \times Tp + Fp + Fn)) \times 100$
Overall $F1$-score	$\mathrm{F1}_{All}$	$(1/N)\sum_{i=0}^{N} F1 - score_i \times 100$
% confidence interval error	CI	$1.96 \times \sqrt{(F1_{score} \times (1 - F1_{score}))/n}$
RMSE	RMSE	$\sqrt{(1/n)\sum_{i=1}^{n} \left(y_i^{theorical} - y_i^{predicted}\right)^2}$
R^2	R^2	$1 - \left(\left(\sum\left(y_i^{predicted} - y_i^{theorical}\right)^2\right)/\left(\sum\left(y_i^{theorical} - y_i^{mean}\right)^2\right)\right)$
Canopy fraction	CF_class	$\sum_{i=1}^{I}\sum_{j=1}^{J}(image(i, j) = class)/\sum_{i=1}^{I}\sum_{j=1}^{J}(image(i, j))$ (where i and j are, respectively, the width and height of the image in pixels)

Table 7: Performances of the U-net 2C model to classify vegetation (*Green Veg.+Sen.Veg.*) and background (Back.) pixels over test Dataset #2. The elements of the confusion matrix, $F1class$ and $F1_{all}$, are presented.

Sub-datasets	Predicted Classes	Labelled Classes (in %)		$\mathrm{F1}_{class}$	$\mathrm{F1}_{all}$
		Veg.	Back.		
LITERAL	Veg.	97	40	94	82
	Back.	3	60	70	
PHENOMOBILE	Veg.	84	13	88	85
	Back.	16	87	81	
P2S2	Veg.	95	12	95	92
	Back.	5	88	88	

(Table 6). We also considered the fraction of pixels of a certain class in an image in a given viewing direction. This trait is widely used as a proxy of crop development [51] particularly for the green parts characteristic of the photosynthetically active elements [52]. Finally, regression results RMSE and R^2 were also considered to evaluate the methods. All these metrics were computed over the test dataset (Table 5), either directly on the test pixels from the image grids, for grid canopy fractions directly on image grids from which the training pixels have been removed, or finally, on the whole images for U-net 3C step.

3. Results

3.1. Performances of the SegVeg Approach

3.1.1. Separation of Vegetation | Background with First-Stage U-net 2C Model. Results (Table 7) on background and combined green/senescent vegetation pixel grids show that U-net 2C first-stage model classifies well the vegetation from the background pixels, with an overall mean $F1$-score between 82% and 92%. The $F1class$ values are higher for the vegetation class. Misclassifications are observed when either the background corresponding to algae/moss is classified as vegetation (Supplementary Material figure S2, bottom) or senescent vegetation is confounded with crop residues (Supplementary Material figure S2, top). The P2S2 subdatasets, achieved the best $F1_{all}$ performances.

3.1.2. Green and Senescent Vegetation Classification Performances of the SVM Only and Full SegVeg Approach. The pixel classification performances were evaluated on the following: (i) applying only the second-stage SegVeg SVM and (ii) applying the full SegVeg approach. Results (Table 8) show that the green vegetation pixels are generally well identified for the three subdatasets.

Table 8: Confusion matrix (in % of the labelled pixels), accuracy, and $F1_{all}$ values computed for the SVM classification only and using the full SegVeg approach for the three subdatasets (e.g., pixels from Dataset #2). The diagonal terms of the confusion matrix are indicated in gray color. The colors of the two last columns correspond to the accuracy and $F1_{all}$ values (dark green, highest; dark red, lowest).

Sub-dataset	Model	Predicted	Labelled Classes (%)				
		Classes	Sen. Veg.	Green Veg.	Back.	Acc	$F1_{all}$
LITERAL	SVM alone	Sen. Veg.	78	4	37	74	52
		Green Veg.	22	96	63		
	SegVeg	Sen. Veg.	82	6	13	84	79
		Green Veg.	9	93	26		
		Backg.	9	1	61		
PHENOMOBILE	SVM alone	Sen. Veg.	90	4	61	60	47
		Green Veg.	10	96	39		
	SegVeg	Sen. Veg.	61	3	5	83	80
		Green Veg.	5	89	7		
		Backg.	34	8	87		
P2S2	SVM alone	Sen. Veg.	89	2	54	69	48
		Green Veg.	11	98	46		
	SegVeg	Sen. Veg.	77	2	6	91	86
		Green Veg.	5	96	6		
		Backg.	18	2	88		

When using only the SVM, the senescent vegetation pixels show significant confusion with the green vegetation for the LITERAL subdataset. The background pixels are preferentially classified as senescent vegetation, except for the LITERAL subdataset (in SVM rows Table 8). This highlights the importance of separating first the vegetation from the background with the U-net 2C model since without using contextual information, e.g., using only the RGB color information, does not allow to separate well the vegetation from the background pixels, particularly for the senescent vegetation and the darkest pixels as illustrated in Supplementary Material figure S3.

3.1.3. Performances of the Full SegVeg Approach. Results obtained over the pixels of test Dataset #2 show that the accuracy and $F1_{all}$ score of the SegVeg model are high for the three subdatasets. The SegVeg approach classifies generally well the pixels into the three classes because of the good performances of the two stages demonstrated earlier (Table 8 and Figure 5).

However, a significant amount of misclassification is still observed between the senescent vegetation and the background for the PHENOMOBILE subdataset and between the background and the green vegetation for the LITERAL one (Figure 6). A 95% confidence interval (CI) error of 2.7 and 2.1%, respectively, for senescent vegetation and background was quantified. This CI is two times higher than that of the green vegetation class. The degraded performances observed on LITERAL images could be explained by the complexity of the images due to the presence of awns that are smaller than the pixel size, inducing confusion between classes (Figure 6).

The classification performances of SegVeg seem to slightly degrade when the green fraction decreases and when the senescent fraction increases (Supplementary Material figure S4). These situations are underrepresented in the U-net 2C training database, which may contribute to the degraded performances observed.

3.2. Comparison of the SegVeg Approach with the U-net 3C. Results show that U-net 3C (Table 9) performs similarly to SegVeg (Table 8) on Dataset #2.

The *Similitude* between the two models has been further studied by looking at differences in each pixel predictions between SegVeg and U-net 3C models. SegVeg pixel predictions were used as groundtruth, i.e., reference values, in confusion matrix of Table 9*Similitude* case.

The average accuracy and $F1_{all}$ values for the *Similitude* are quite high, 90 and 85, respectively, with high values in the diagonal terms of the confusion matrix. However, on average, SegVeg approach exhibits slightly higher performances compared to U-net 3C. Tables 8 and 9 reveal that the best performances for SegVeg come mostly from a better identification of the background pixels, particularly for the LITERAL dataset.

Both models achieve the best performances on the P2S2 subdataset, whereas the worst performances are observed on the LITERAL subdataset. The poor performances are particularly due to larger confusion over the background class predicted by U-net 2C (Tables 8 and 9).

4. Discussion

4.1. Use of Different Color Spaces to Better Separate the Green and Senescent Vegetation. Differences in eye sensitivity among operators impact the perception of colors [53] and may therefore induce disagreement among them. Further, first stages of senescence may also create differences between

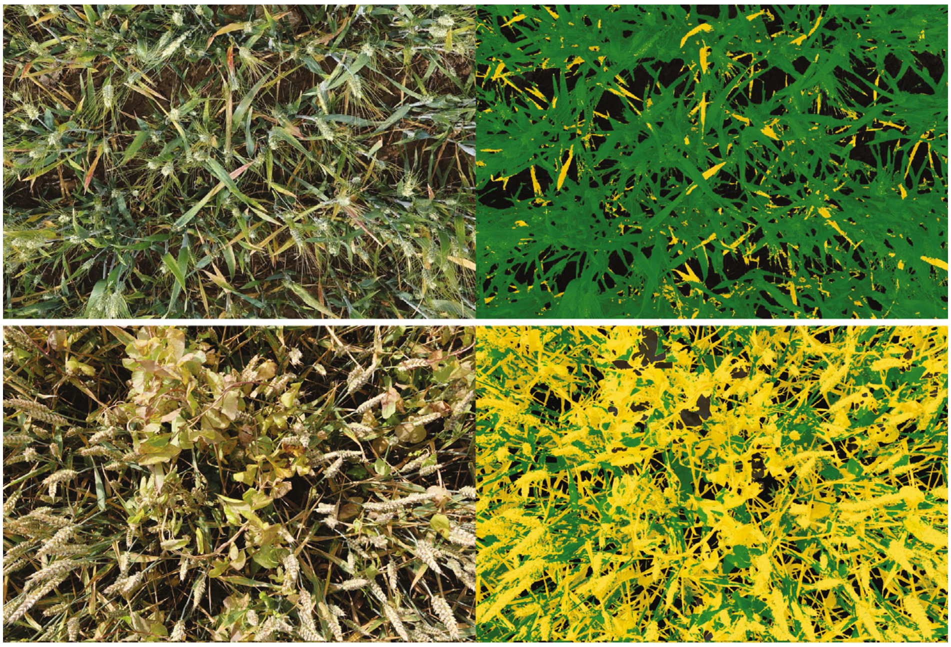


Figure 5: Examples of SegVeg model predictions over entire images of wheat acquired with LITERAL during early (top) and late (bottom) senescence stage. On the left, the original RGB images. On the right, the corresponding segmented images where the background, and the green and senescent vegetation are represented, respectively, in black, green, and yellow.

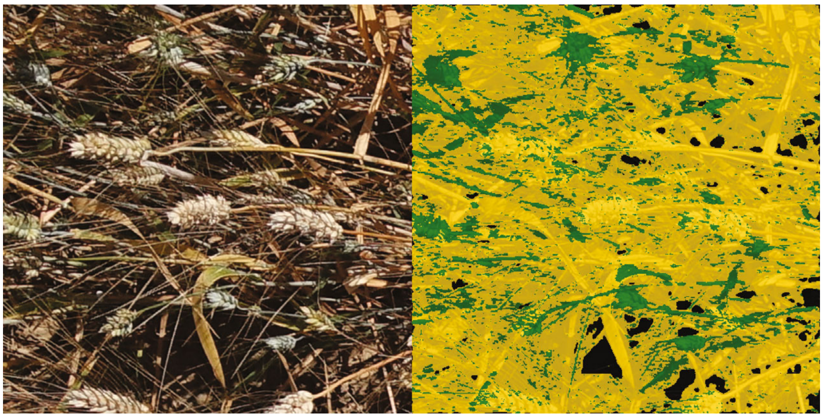


Figure 6: Example of misclassification with the SegVeg approach on a complex image presenting lots of thin spikes acquired with LITERAL.

the labelling of operators, since the yellow and reddish colors observed are in continuity with the green ones in the color space. To take into account this effect, the labelling was done using several operators to get more consensual labelling.

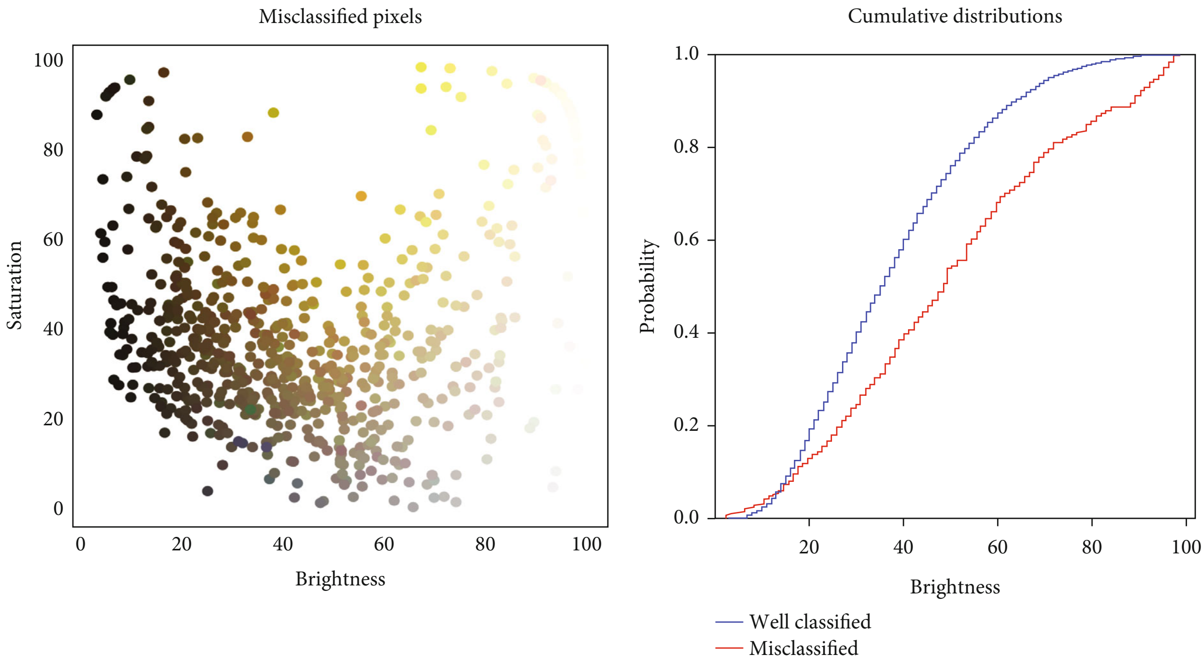
The colors identified as senescent vegetation during the SVM classification of the vegetation pixels show that simple thresholds in the RGB space are not sufficient to get a satisfactory separation. Reciprocally, the same applies to the green vegetation. The combined use of certain components of other color representations seem to be useful to segment the green vegetation as proposed by other authors such as R, S, a, b, Cb, and Cr in [21], sRGB space used for CIELab transformation, in [54], or H and S in [55]. Likewise, additional features may also be used to better separate the senescent vegetation such as the CMYK color space or the quadrature from YIQ that were selected as input features to the SVM (Supplementary Material figure S5).

To better highlight, qualitatively, the model performances using these features and the corresponding theoretical boundaries, a 3D RGB cube of 35^3 voxels was created. It contains a huge panel of color shades, which helps to discern visually where the SegVeg approach locates the senescent vegetation within the color spaces (Supplementary Material figure S6).

4.2. Impact of Illumination Conditions on the Segmentation Performances. The pixels misclassified by the SegVeg approach correspond mostly to brownish colors representative of the senescent vegetation or background (Figure 7(a)). The few green pixels observed with high brightness and

Table 9: Performances of U-net 3C model and Similitude to SegVeg model evaluation (in %). Similitude confusion matrix was built with SegVeg outputs as groundtruth pixel values on Dataset #2. The diagonal terms of the confusion matrix are indicated in gray cells. The colors of accuracy and $F1_{all}$ are related to their performances (dark red the lowest; dark green the highest).

Sub-dataset	Model	Predicted	Labelled Classes(%)			Acc	$F1_{all}$
		Classes	Sen. Veg.	Green Veg.	Back.		
LITERAL	U-net 3C	Sen. Veg.	86	4	28	83	74
		Green Veg.	10	95	31		
		Backg.	4	1	41		
	Similitude	Sen. Veg.	87	2	30	89	83
		Green Veg.	12	97	10		
		Backg.	1	1	60		
PHENOMOBILE	U-net 3C	Sen. Veg.	67	3	9	83	81
		Green Veg.	12	89	7		
		Backg.	21	8	84		
	Similitude	Sen. Veg.	76	3	12	87	83
		Green Veg.	13	94	5		
		Backg.	11	3	83		
P2S2	U-net 3C	Sen. Veg.	74	2	6	90	85
		Green Veg.	11	95	6		
		Backg.	15	3	88		
	Similitude	Sen. Veg.	80	2	4	93	90
		Green Veg.	14	96	3		
		Backg.	6	2	93		

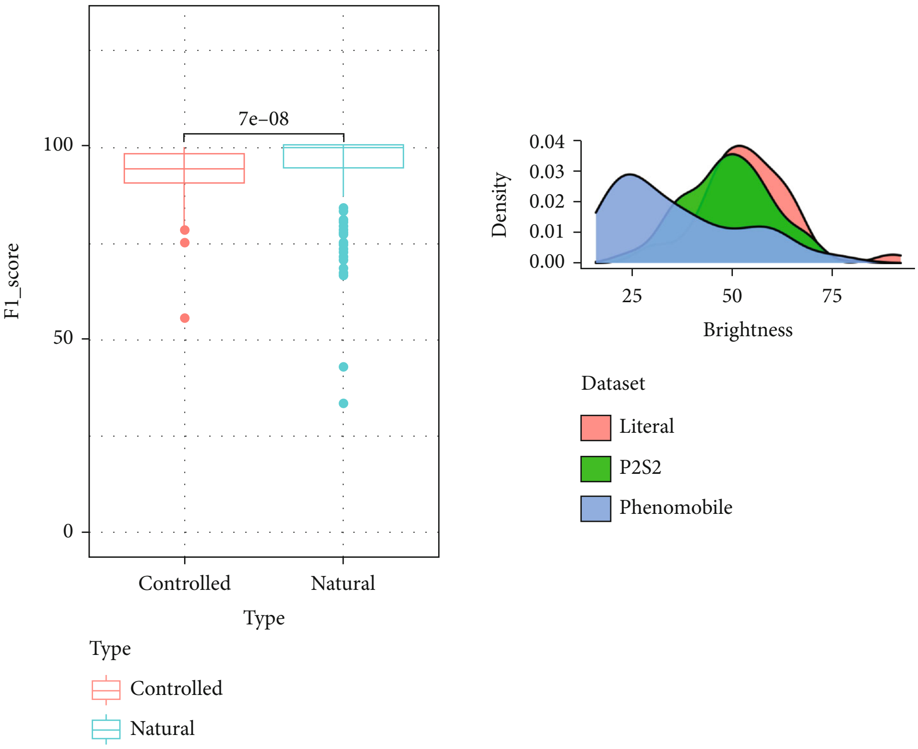


(a)

(b)

Figure 7: (a) Distribution of the brightness (V from HSV) and saturation (S from HSV) for the misclassified pixels by the SegVeg model. Each point corresponds to a misclassified pixel from the grids of the test dataset. They are represented by their actual RGB color. (b) Cumulated distribution of the brightness of the misclassified (red) and well-classified (blue) pixel.

saturation may correspond either to errors in the labelling or to mixed pixels very close to the limit between the green and senescent vegetation (Supplementary Material figure S6). Illumination conditions may also strongly impact the quality of the classification. Misclassified pixels are preferentially observed for the small brightness values (Figure 7(b)) where



(a)

(b)

Figure 8: (a) Distribution of the performance ($F1_{all}$) for both controlled (PHENOMOBILE) and natural (P2S2 and LITERAL datasets) illumination conditions (with p value expressed above boxplots). (b) Distribution over brightness (V from HSV) for the three datasets.

the dynamics of the color values may be too limited to get an accurate classification based both on the color spaces or on the spatial features, inducing confusion among the three classes. This applies both to the labelling process and to the model predictions. Misclassified pixels are also observed preferentially in the highest brightness values (Figure 7(b)). In such conditions, some authors [56] propose to assign the saturated pixels to the most frequently saturated class. In our case, this would degrade the segmentation performances since the saturated pixels may belong to any of the three classes. However, a larger representation of green vegetation particularly with glossy leaves under either clear sky conditions or using flashes is often saturated.

The confusions observed for the PHENOMOBILE subdataset and leading to slightly degraded segmentation performances (Figure 8(a)) are partly due to the use of flashes instead of the natural illumination as in LITERAL and P2S2 subdatasets. The noncollimated nature of the light emitted by the flashes induces a decrease in the intensity of the radiation that varies as the inverse of the square of the distance to the source. When the source is too close to the top of canopy, pixels tend to be saturated with limited classification potential. To limit this saturation effect, images taken from the PHENOMOBILE were slightly underexposed. Further, the pixels located at the bottom of the scene receive very little illumination and are therefore very dark. The distribution of the brightness for the PHENOMOBILE dataset (Figure 8(b)) shows more darker pixels than the other subdatasets acquired under natural illumination conditions. This is in agreement to the higher confusion between the vegetation and the background presented earlier (Tables 7 and 8).

4.3. Weak Supervision Is Promising. Because of the unavailability of images fully labelled into the three classes, U-net 3C was trained over masks predicted by the SegVeg model. This weak supervision approach could lead to biased predictions, since SegVeg predicted masks are not perfect as demonstrated previously in Table 8, and obviously, training will converge to similar SegVeg results. Moreover, U-net 3C was trained over whole images compared to 6132 pixels for SVM classification model. However, the performances of U-net 3C (Table 9) are quite close to those of SegVeg (Table 8) for the PHENOMOBILE and P2S2 subdatasets, while SegVeg performs slightly better over the LITERAL subdataset. Comparison between SegVeg and U-net 3C (Table 9, "Similitude" case) confirms the consistency between the two models, as expected. Weak supervision appears to be a promising way to pretrain deep learning algorithms by reducing the labelling process by the operators. The larger number of images therefore available to train the model is expected to partly compensate for the lower quality of the "automatic" labelling. However, the main differences lie in the patterns of the green and senescent vegetation masks (Figure 9) where SegVeg appears crisper than U-net 3C

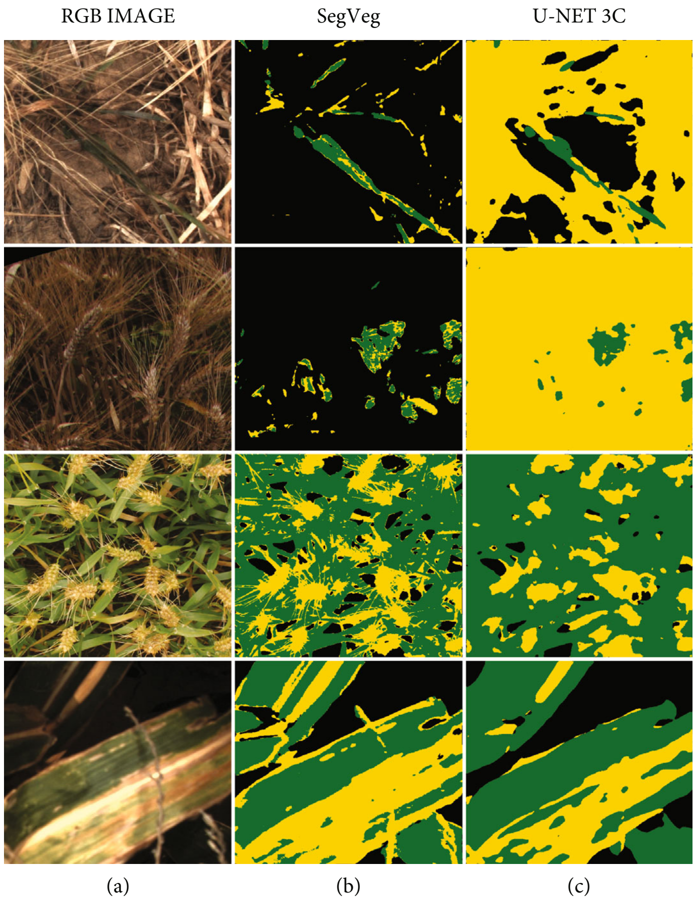
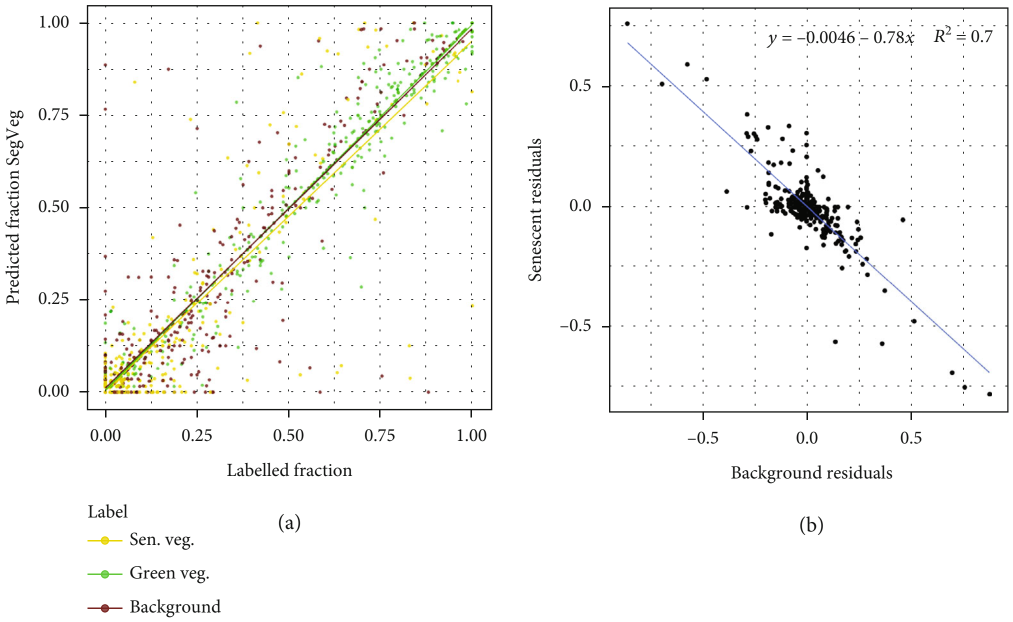


Figure 9: Results of the segmentation using SegVeg (b) or U-net 3C (c). Background, green vegetation, and senescent vegetation are represented, respectively, in black, green, and yellow. (a) The original RGB image.

Table 10: Performances of SegVeg and U-net 3C to estimate the background, green, and senescent vegetation fractions over grids. "Similitude" for comparison of model performances was computed using either the labelled grids or whole images. R^2 is the determination coefficient. The colors of R^2 and RMSE are related to their column values (dark green, the best; dark red, the worst).

Fraction	Model	Grid/Image	R^2	RMSE	Slope	Offset (abs.)
Background	SegVeg	Grid	0.73	0.14	0.97	0.01
	U-net 3C	Grid	0.74	0.14	0.99	0.01
	Similitude	Grid	0.82	0.13	0.92	0.00
		Image	0.80	0.13	0.94	0.01
Green Veg.	SegVeg	Grid	0.94	0.08	0.99	0.00
	U-net 3C	Grid	0.90	0.10	1.03	0.01
	Similitude	Grid	0.95	0.07	1.04	0.01
		Image	0.95	0.07	1.07	0.02
Sen. Veg.	SegVeg	Grid	0.70	0.13	0.95	0.00
	U-net 3C	Grid	0.70	0.13	1.14	0.02
	Similitude	Grid	0.74	0.14	1.02	0.00
		Image	0.73	0.13	1.07	0.00

which shows fuzzier masks. Indeed, the kernel filters used in U-net 3C to separate the green from the senescent vegetation tend to omit the small elements in the images and render more diffused patches. Conversely, the pixel-based separation between the green and senescent vegetation allows to better describe the small details (Figure 9).



(a)

(b)

Figure 10: (a) Comparison between the fractions predicted by the SegVeg approach and the labelled ones over test Dataset #1: green vegetation (green), senescent vegetation (yellow), and background (brown). The best fit for each fraction is represented by a solid line of the same color. (b) Relationship between the residual of the background and senescent fractions.

4.4. Predicting the Fractions of Green and Senescent Vegetation. The evaluation of the performances over pixels that have been labelled by the operators has been presented. However, the grid-pixels correspond to a subsample of the image which questions their representativeness in regard to the entire image. We therefore evaluated the agreement between the segmentation predicted by SegVeg and by U-net 3C over both the grid-pixels and the entire images, following the same exact principle as Table 9, SegVeg pixels as reference. Results show (Table 10, "Similitude" case) that R^2, RMSE, slope, and offset for the grids and the images are in good agreement of each of the three fractions considered. This indicates the fraction of background, green, and senescent vegetation computed over the pixel subsampling represents quite well the whole images.

SegVeg and U-net 3C show similar performances. The best agreement is observed for the green vegetation fraction (Table 10), with a slight advantage for SegVeg, confirming the slightly better performances in the segmentation of this class (Tables 8 and 9). The estimates are not biased, according to slopes (Table 8 and Figure 10(a)). Conversely, the estimation of the background and senescent vegetation fractions show degraded performances for U-net 3C, which are related to the degraded performances observed previously in the segmentation of these two classes. The confusion between the background and the senescent vegetation pixels by U-net 3C may be quite large as highlighted by the number of outliers, with a quasiexact compensation between these two fractions since the green vegetation fraction is well predicted (Figure 10(b)). Small biases are observed in these fractions predicted by SegVeg and U-net 3C models, except for the senescent fraction of U-net 3C for which the bias (Table 10) mostly comes from the distribution of the outliers (Figure 10).

The SegVeg approach and U-net 3C segmentation appear efficient to compute the fractions of the different elements of the image. However, the SegVeg model offers a slight advantage with better performances for green fraction and smaller biases in senescent vegetation fraction.

4.5. Limitation of the Study. This study is based on segmentation models using shallow and deep learning techniques. It is therefore constrained by the availability of training and testing datasets. The first-stage SegVeg U-net 2C model was trained over a relatively large and diverse database (Table 1) containing 2015 images of 512 × 512 pixels. The SegVeg SVM is trained over 6132 pixels extracted from grids applied to the original images, thus showing a wide diversity in species, phenological stages, canopy state, and acquisition conditions. However, the pixels labelled as uncertain (green/Sen. Veg. unsure, unknown, and other) were not used, forcing the SVM model to extrapolate for these situations. Finally, the training was completed over two subdatasets where the P2S2 is overrepresented as compared to PHENOMOBILE. This is why the results were presented per subdataset. This also partly explains the differences in performances observed over the three test datasets, with a general trend: P2S2 > PHENOMOBILE > LITERAL.

The evaluation of the models was performed at the pixel level. A large number of pixels was considered here (more than 20,000 pixels, including background class, Table 5), along with those extracted from the LITERAL subdataset that were not used in training. The "unsure" pixels were not used to compute the performances, which may also induce small biases in the results since the "unsure" pixels may not be evenly distributed between the three classes of interest. However, we did not have other alternatives, since "unsure" pixels correspond mostly to extremely dark, bright (S3), or mixed pixels. Indeed, great attention should be paid to the image spatial resolution and exposure during image acquisition. Studies based on 3D scenes rendered realistically should be conducted to better understand the unsure classes and their possible distribution among the three classes of interest.

Data Availability

SegVeg pixels dataset, images, and their corresponding segmentation masks are be publicly available. All the SegVeg scripts for computation and analysis are also public: https://github.com/mserouar/SegVeg. For simplicity, dataset download links (including Zenodo) will be specified in the above repository.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Authors' Contributions

Mario Serouart and Simon Madec have written the code, analyzed the results, and conducted the reviews. Mario Serouart, Simon Madec, Kaaviya Velumani, and Etienne David have conducted the pipeline. Frederic Baret, Marie Weiss, and Raul Lopez Lozano have supervised experiments at all stages. All authors contributed to editing, reviewing, and refining the manuscript.

Acknowledgments

This work received support from ANRT for the CIFRE grant of Mario Serouart, cofunded by Arvalis. The study was partly supported by several projects including ANR PHENOME (Programme d'investissement d'avenir), Digitag (PIA Institut Convergences Agriculture Numérique ANR-16-CONV-0004), CASDAR LITERAL, and P2S2 funded by CNES. Many thanks are due to the people who annotated the datasets, including Frederic Venault, Micheline Debroux, Kamran Irfan, and Gaetan Daubige.

Supplementary Materials

Supplementary 1. Video record of grid-pixel annotation process. The annotator classifies every pixel at the intersection of the grid to the one of the six classes.

Supplementary 2. Example of classification errors with the vegetation and background first-stage U-net 2C model. Left: original images. Middle: vegetation masked images. Right: background masked images. Top: almost all the senescent vegetation is classified as soil. Bottom: background algae zones are classified as vegetation.

Supplementary 3. Distribution of the colors among the six classes as observed over the labelled pixels of the test and training datasets. For each class, pixels are sorted according to their brightness from the HSV color space.

Supplementary 4. Performances ($F1$ all) of the SegVeg approach as a function of the green (left) and senescent fraction (right) per image.

Supplementary 5. Use of different meaningful color spaces to describe the image content: RGB original image (left), Y component from CMYK (middle), and Q component from YIQ (right). Y and Q images are in gray scale.

Supplementary 6. Boundaries of SegVeg colors inferred on a 35 3-voxel RGB cube thanks to the SegVeg model second-stage SVM. On the right, yellow predicted pixels. On the left, the rest that includes the green predicted pixels.

References

[1] T. Sakamoto, A. A. Gitelson, A. L. Nguy-Robertson et al., "An alternative method using digital cameras for continuous monitoring of crop status," *Agricultural and Forest Meteorology*, vol. 154-155, pp. 113–126, 2012.

[2] P. Steduto, T. C. Hsiao, D. Raes, and E. Fereres, "AquaCrop—the FAO crop model to simulate yield response to water: I. concepts and underlying principles," *Agronomy Journal*, vol. 101, no. 3, pp. 426–437, 2009.

[3] K. Donohue, "Development in the wild: phenotypic plasticity," in *Annual Plant Reviews Volume 45: The Evolution of Plant Form*, pp. 321–355, JohnWiley & Sons, Ltd, 2013.

[4] I. Jonckheere, S. Fleck, K. Nackaerts et al., "Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography," *Agricultural and Forest Meteorology*, vol. 121, no. 1-2, pp. 19–35, 2004.

[5] W. Li, H. Fang, S. Wei, M. Weiss, and F. Baret, "Critical analysis of methods to estimate the fraction of absorbed or intercepted photosynthetically active radiation from ground measurements: application to rice crops," *Agricultural and Forest Meteorology*, vol. 297, article 108273, 2021.

[6] M. J. Hill and J. P. Guerschman, "Global trends in vegetation fractional cover: hotspots for change in bare soil and non-photosynthetic vegetation," *Agriculture, Ecosystems and Environment*, vol. 324, article 107719, 2022.

[7] M. Weiss, F. Baret, G. Smith, I. Jonckheere, and P. Coppin, "Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling," *Agricultural and Forest Meteorology*, vol. 121, no. 1-2, pp. 37–53, 2004.

[8] N. Sade, M. Del Mar Rubio-Wilhelmi, K. Umnajkitikorn, and E. Blumwald, "Stress-induced senescence and plant tolerance to abiotic stress," *Journal of Experimental Botany*, vol. 69, no. 4, pp. 845–853, 2018.

[9] S. Munné-Bosch and L. Alegre, "Die and let live: leaf senescence contributes to plant survival under drought stress," *Functional Plant Biology*, vol. 31, no. 3, pp. 203–216, 2004.

[10] H. G. Sedigheh, M. Mortazavian, D. Norouzian et al., "Oxidative stress and leaf senescence," *BMC Research Notes*, vol. 4, no. 1, p. 477, 2011.

[11] J. T. Christopher, M. J. Christopher, A. K. Borrell, S. Fletcher, and K. Chenu, "Stay-green traits to improve wheat

adaptation in well-watered and water-limited environments," *Journal of Experimental Botany*, vol. 67, no. 17, pp. 5159–5172, 2016.

[12] J. Anderegg, K. Yu, H. Aasen, A. Walter, F. Liebisch, and A. Hund, "Spectral vegetation indices to track senescence dynamics in diverse wheat germplasm," *Frontiers in Plant Science*, vol. 10, p. 1749, 2020.

[13] C. Ji, X. Li, H. Wei, and S. Li, "Comparison of different multispectral sensors for photosynthetic and non-photosynthetic vegetation-fraction retrieval," *Remote Sensing*, vol. 12, no. 1, p. 115, 2020.

[14] E. David, M. Serouart, D. Smith et al., "Global wheat head detection 2021: An improved dataset for benchmarking wheat head detection methods," *Plant Phenomics*, vol. 2021, article 9846158, 2021.

[15] S. Liu, F. Baret, B. Andrieu, P. Burger, and M. Hemmerlé, "Estimation of wheat plant density at early stages using high resolution imagery," *Frontiers in Plant Science*, vol. 8, 2017.

[16] K. Velumani, S. Madec, B. de Solan et al., "An automatic method based on daily *in situ* images and deep learning to date wheat heading stage," *Field Crops Research*, vol. 252, article 107793, 2020.

[17] S. Jay, F. Baret, D. Dutartre et al., "Exploiting the centimeter resolution of UAV multispectral imagery to improve remote-sensing estimates of canopy structure and biochemistry in sugar beet crops," *Remote Sensing of Environment*, vol. 231, article 110898, 2019.

[18] A. Comar, P. Burger, B. De Solan, F. Baret, F. Daumard, and J.-F. Hanocq, "A semi-automatic system for high throughput phenotyping wheat cultivars in-field conditions: description and first results," *Functional Plant Biology*, vol. 39, no. 11, pp. 914–924, 2012.

[19] B. Frederic, B. de Solan, R. Lopez-Lozano, K. Ma, and M. Weiss, "GAI estimates of row crops from downward looking digital photos taken perpendicular to rows at 57.5° zenith angle: theoretical considerations based on 3D architecture models and application to wheat crops," *Agricultural and Forest Meteorology*, vol. 150, no. 11, pp. 1393–1401, 2010.

[20] G. E. Meyer and J. C. Neto, "Verification of color vegetation indices for automated crop imaging applications," *Computers and Electronics in Agriculture*, vol. 63, no. 2, pp. 282–293, 2008.

[21] W. Guo, U. Rage, and S. Ninomiya, "Illumination invariant segmentation of vegetation for time series wheat images based on decision tree model," *Computers and Electronics in Agriculture*, vol. 96, pp. 58–66, 2013.

[22] T. L. Starks, L. E. Shubert, and F. R. Trainor, "Ecology of soil algae: a review," *Phycologia*, vol. 20, no. 1, pp. 65–80, 1981.

[23] H. Thomas, H. J. Ougham, C. Wagstaff, and A. D. Stead, "Defining senescence and death," *Journal of Experimental Botany*, vol. 54, no. 385, pp. 1127–1132, 2003.

[24] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," *International Journal of Multimedia Information Retrieval*, vol. 7, no. 2, pp. 87–93, 2018.

[25] F. Jiang, A. Grigorev, S. Rho et al., "Medical image semantic segmentation based on deep learning," *Neural Computing and Applications*, vol. 29, no. 5, pp. 1257–1265, 2018.

[26] A. Milioto, P. Lottes, and C. Stachniss, "Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs," 2018, http://arxiv.org/abs/1709.06764.

[27] C. Miao, A. Pages, Z. Xu, E. Rodene, J. Yang, and J. C. Schnable, "Semantic segmentation of sorghum using hyperspectral data identifies genetic associations," *Plant Phenomics*, vol. 2020, article 4216373, 2020.

[28] R. Barth, J. IJsselmuiden, J. Hemming, and E. J. V. Henten, "Data synthesis methods for semantic segmentation in agriculture: A *Capsicum annuum* dataset," *Computers and Electronics in Agriculture*, vol. 144, pp. 284–296, 2018.

[29] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," 2015, http://arxiv.org/abs/1505.04597.

[30] M. Tan and Q. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105–6114, ICML 2019 Long Beach, 2019.

[31] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2018.

[32] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," 2012, https://arxiv.org/abs/1206.5533.

[33] P. Yakubovskiy, "Qubvel/segmentation models," *Pytorch*, vol. 21, p. 21Z, 2021.

[34] H.-Y. Yang, Wang, Wang, and Zhang, "LS-SVM based image segmentation using color and texture information," *Journal of Visual Communication and Image Representation*, vol. 23, pp. 1095–1112, 2012.

[35] T.-C. Zhang, J. Zhang, J.-P. Zhang, and H. Wang, "Review of methods of image segmentation based on quantum mechanics," *Journal of Electronic Science and Technology*, vol. 16, no. 3, pp. 243–252, 2018.

[36] S. Sakurai, H. Uchiyama, A. Shimada, D. Arita, and R.-i. Taniguchi, "Two-step transfer learning for semantic plant segmentation," in *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods - ICPRAM*, pp. 332–339, ICPRAM 2018, Funchal Madeira - Portugal, 2018.

[37] G. Joblove and D. P. Greenberg, "Color spaces for computer graphics," *ACM SIGGRAPH Computer Graphics*, vol. 12, no. 3, pp. 20–25, 1978.

[38] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[40] S. V. Desai, V. N. Balasubramanian, T. Fukatsu, S. Ninomiya, and W. Guo, "Automatic estimation of heading date of paddy rice using deep learning," *Plant Methods*, vol. 15, no. 1, p. 76, 2019.

[41] W. Guo, B. Zheng, T. Duan, T. Fukatsu, S. Chapman, and S. Ninomiya, "Easypcc: benchmark datasets and tools for high-throughput measurement of the plant canopy coverage ratio under field conditions," *Sensors*, vol. 17, no. 4, 2017.

[42] S. Madec, K. Irfan, E. David et al., "The P2S2 segmentation dataset: annotated in-field multi-crop RGB images acquired under various conditions," in *7th International Workshop on Image Analysis Methods in the Plant Sciences (IAMPS)*, Lyon, France, 2019.

[43] Y. Li, Z. Cao, H. Lu, Y. Xiao, Y. Zhu, and A. B. Cremers, "In-field cotton detection via region-based semantic image

segmentation," *Computers and Electronics in Agriculture*, vol. 127, pp. 475–486, 2016.

[44] M. Minervini, A. Fischbach, H. Scharr, and S. A. Tsaftaris, "Finely-grained annotated datasets for image-based plant phenotyping," *Pattern Recognition Letters*, vol. 81, pp. 80–89, 2016.

[45] N. Chebrolu, P. Lottes, A. Schaefer, W. Winterhalter, W. Burgard, and C. Stachniss, "Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields," *The International Journal of Robotics Research*, vol. 36, no. 10, pp. 1045–1052, 2017.

[46] S. Madec, F. Baret, B. de Solan et al., "High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates," *Frontiers in Plant Science*, vol. 8, pp. 1–15, 2017.

[47] DataTorch, "Tools and community for building machine learning models," https://datatorch.io/.

[48] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[49] H. Yu, J. Yang, J. Han, and X. Li, "Making SVMS scalable to large data sets using hierarchical cluster indexing," *Data Mining and Knowledge Discovery*, vol. 11, no. 3, pp. 295–321, 2005.

[50] Y.-J. Lee and O. L. Mangasarian, "RSVM: reduced support vector machines," in *Proceedings of the 2001 SIAM International Conference on Data Mining (SDM)*, pp. 1–17, Chicago, USA, 2001.

[51] A. Gitelson, Y. Kaufman, R. Stark, and D. Rundquist, "Novel algorithms for remote estimation of vegetation fraction," *Remote Sensing of Environment*, vol. 80, no. 1, pp. 76–87, 2002.

[52] M. D. Marcial-Pablo, A. Gonzalez-Sanchez, S. I. Jimenez-Jimenez, R. E. Ontiveros-Capurata, and W. Ojeda-Bustamante, "Estimation of vegetation fraction using RGB and multispectral images from UAV," *International Journal of Remote Sensing*, vol. 40, no. 2, pp. 420–438, 2019.

[53] R. Pridmore, "Complementary colors theory of color vision: physiology, color mixture, color constancy and color perception," *Color Research and Application*, vol. 36, no. 6, pp. 394–412, 2011.

[54] H. K. Suh, J. W. Hofstee, and E. J. van Henten, "Improved vegetation segmentation with ground shadow removal using an HDR camera," *Precision Agriculture*, vol. 19, no. 2, pp. 218–237, 2018.

[55] E. Hamuda, B. Mc Ginley, M. Glavin, and E. Jones, "Automatic crop detection under field conditions using the HSV colour space and morphological operations," *Computers and Electronics in Agriculture*, vol. 133, pp. 97–107, 2017.

[56] S. Dandrifosse, A. Bouvry, V. Leemans, B. Dumont, and B. Mercatoris, "Imaging Wheat canopy through stereo vision: overcoming the challenges of the laboratory to field transition for morphological features extraction," *Frontiers in Plant Science*, vol. 11, 2020.