Article

Plant Detection in RGB Images from Unmanned Aerial Vehicles Using Segmentation by Deep Learning and an Impact of Model Accuracy on Downstream Analysis

Mikhail V. Kozhekin [1,2], Mikhail A. Genaev [1,2], Evgenii G. Komyshev [1,2,3], Zakhar A. Zavyalov [4] and Dmitry A. Afonnikov [1,2,3,*]

1 Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, 630090 Novosibirsk, Russia
2 Kurchatov Center for Genome Research, Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences, 630090 Novosibirsk, Russia
3 Department of Mechanics and Mathematics, Novosibirsk State University, 630090 Novosibirsk, Russia
4 GeosAero LLC, 440000 Penza, Russia
* Correspondence: ada@bionet.nsc.ru; Tel.: +7-(383)-363-4963

Abstract: Crop field monitoring using unmanned aerial vehicles (UAVs) is one of the most important technologies for plant growth control in modern precision agriculture. One of the important and widely used tasks in field monitoring is plant stand counting. The accurate identification of plants in field images provides estimates of plant number per unit area, detects missing seedlings, and predicts crop yield. Current methods are based on the detection of plants in images obtained from UAVs by means of computer vision algorithms and deep learning neural networks. These approaches depend on image spatial resolution and the quality of plant markup. The performance of automatic plant detection may affect the efficiency of downstream analysis of a field cropping pattern. In the present work, a method is presented for detecting the plants of five species in images acquired via a UAV on the basis of image segmentation by deep learning algorithms (convolutional neural networks). Twelve orthomosaics were collected and marked at several sites in Russia to train and test the neural network algorithms. Additionally, 17 existing datasets of various spatial resolutions and markup quality levels from the Roboflow service were used to extend training image sets. Finally, we compared several texture features between manually evaluated and neural-network-estimated plant masks. It was demonstrated that adding images to the training sample (even those of lower resolution and markup quality) improves plant stand counting significantly. The work indicates how the accuracy of plant detection in field images may affect their cropping pattern evaluation by means of texture characteristics. For some of the characteristics (GLCM mean, GLRM long run, GLRM run ratio) the estimates between images marked manually and automatically are close. For others, the differences are large and may lead to erroneous conclusions about the properties of field cropping patterns. Nonetheless, overall, plant detection algorithms with a higher accuracy show better agreement with the estimates of texture parameters obtained from manually marked images.

Keywords: crop; field image; plant counting; UAV; deep learning; semantic segmentation; image texture analysis

1. Introduction

Unmanned aerial vehicle (UAV)-based crop monitoring is one of the most important technologies for controlling plant growth in modern precision agriculture [1,2]. The high mobility of UAVs allows for the monitoring of large fields and the collection of data from all areas. Due to the ability to use different sensors, this technology has found a wide range of applications in agriculture [3,4] and field plant phenomics [5,6]. These applications include estimation of the soil moisture content [7], weed detection [8], and assessment of the leaf area index (LAI) and plant biomass [9,10]. Methods have been developed to monitor plant nitrogen status [11] and plant height [12], to detect pathogens [13], and to evaluate plant phenology in the field [14,15]. A promising approach is to employ data obtained from UAV monitoring for the yield prediction of crops [16–18].

One of the important and common tasks in field monitoring is plant stand counting in images obtained from UAVs [19]. The number of plants per unit area is closely related to yield. The evaluation of these parameters can be performed at different stages of plant development. At the early stages, it characterizes germination and allows for the planning of subsequent agronomic measures to achieve the highest yield. The estimation of plant numbers via ground observations is labor-intensive and time-consuming. Visual inspection is prone to human error and subjectivity. Even the use of hand-held digital cameras allows only small areas of fields to be inspected [20]. Methods based on RGB image acquisition by means of a UAV are much more productive. As a rule, they include several basic steps [6,21–23]: (1) the UAV flight and raw-image capture, (2) the stitching of raw images into a single orthomosaic, (3) the development of a digital-image-processing method to identify plants in the orthomosaic, and (4) subsequent analysis for the formulation of recommendations to agronomists. The main efforts of researchers in the field of image analysis are focused on step 3: techniques for fast and accurate plant identification [19,24].

Two types of methods are used for plant identification based on the analysis of digital orthophoto images: those involving computer vision and those based on deep machine learning [19]. The first type necessitates binarization/segmentation algorithms, which are typically implemented via a combination of R, G, and B components to compute indices in order to distinguish plant pixels from soil pixels [21,25–27]. During the postprocessing of plant areas in the images, the separation of objects and the removal of noise (usually small objects) take place. Several approaches are employed for these purposes: the evaluation of an object's shape/size features and machine learning algorithms [21], peak detection in plant rows [26–28], and searching for statistical correlations [29]. The image objects filtered by postprocessing are thought to correspond to plants, and their counting is then performed.

Deep learning methods are based on networks characterized by a multilayered architecture where subsequent layers utilize the output of a preceding layer as an input to extract features related to the analyzed objects [24,30,31]. These approaches enable the automatic extraction of image features with regression or classification within a single pipeline, trained simultaneously from end to end [32].

Deep machine learning techniques can solve three of the most frequent image processing problems: image classification, image segmentation, and searching for objects in an image [33–35]. The two latter approaches are used to count plants in UAV images [24,36]. When the segmentation problem is being solved by deep machine learning methods, no data preprocessing or construction of various indices is required because it is used in computer vision methods. Necessary plant parameters are extracted automatically.

For semantic segmentation, architectures based on convolutional neural networks (CNNs) are currently used [37]: fully convolutional networks, convolutional networks with graphical models, encoder–decoder-based models, and some others. These methods have been successfully utilized in plant counting problems using different network topologies.

Fully convolutional networks have been employed for plant identification in field images of corn and strawberry [38] and for plant and weed identification [39]. U-Net-based topologies have been used to count maize plants at an early developmental stage [40], to identify plants in high-elevation ecosystems [41], and to recognize a maize plant in the field [42]. A network with a SegNet architecture has been utilized to count bolls of cotton to estimate its yield [43]. Note that semantic segmentation approaches require the further processing of the neural network results in order to identify individual plants in the images [38]. They are most suitable for counting plants at the early stages of development, while the contours of individual plants in the images are not touching each other.

Instance segmentation methods are convenient for solving the problem of counting plants in an image because they enable an investigator to select image regions corresponding to different plants. These techniques are actively used in tasks of object extraction from an image [37]. One of the most popular network architectures for instance segmentation is Region CNN (R-CNN) [44] and its modifications. A Mask R-CNN network has been used for counting the medicinal plants *Lamiophlomis rotata* in mountain landscapes [45] as well as potato and lettuce plants [46]. Faster R-CNN networks have been employed for the counting of maize, sugar beet, sunflower plants [47], corn plants [48], and potato plants [49] in field images.

Object detection methods are also actively used for plant counting. In particular, these are networks based on You Only Look Once (YOLO) architectures [50]. They have a higher speed and have been applied to count cotton [51], sorghum [52], and maize [53] plants. Several neural network architectures have been used for the detection of citrus trees in images [54]. Detection and instance segmentation approaches tend to identify individual plants more accurately, even if the plants touch each other. Nonetheless, they are more computationally and memory-demanding [35].

It should be noted that the identification of plants in images is not the only purpose of such projects for agronomists but serves as a basis for the subsequent assessment of crop characteristics in order to select optimal agronomic treatments. The mean absolute error and coefficients of correlation between a predicted number of plants and the true number of plants give an estimate of the bias in plant density when machine learning algorithms are applied [21]. Other characteristics include sowing uniformity [55] and row regularity for weed identification [56–58].

One approach to the estimation of the uniformity of objects in an image is to use texture features [59–61]. In crop analysis, texture parameters are employed in particular for between- and within-crop-row weed detection [56,57]. Nevertheless, how the accuracy of plant identification by machine learning affects the characteristics of plant arrangement remains unclear.

Here, a method is presented for detecting the plants of five species in images acquired from a UAV on the basis of image segmentation via deep learning algorithms (CNNs). Twelve orthomosaics were collected and marked at several sites in Russia to train and test neural network algorithms. Additionally, 17 external datasets from the Roboflow service were utilized to extend image sets. Finally, we compared several texture features for manually assessed and neural-network-estimated plant masks.

2. Materials and Methods

2.1. Image Acquisition and Construction of Orthomosaics

The locations of the fields (Penza Oblast, Krasnodar Krai, and Stavropol Krai, Russia) are given in Table S1 (Supplementary Materials). Image acquisition in each field was performed within 1 day by means of three UAVs of the Geoscan 201 Agrogeodesia model of the flying wing type (Figure 1). The UAVs are equipped with a Sony RX1R II RGB

camera (Sony Corporation, Tokyo, Japan) with a resolution of 42.4 Mpix. The flight altitude was 50 m. Each UAV is capable of imaging 8000 ha per day. The flights and aerial image acquisition were carried out by GeosAero LLC (Penza, Russia).

Figure 1. A UAV launching from the launcher.

Orthomosaics were built from images using Agisoft Metashape Professional (https://www.agisoft.com (accessed on 1 October 2023)). The coordinate system was WGS:84 (EPSG:4326), and the file format was geotiff. Orthomosaics were obtained with a spatial resolution of at least 2 cm/pix.

2.2. Image Datasets

2.2.1. Data from Russian Regions During 2019–2023

The main dataset included orthomosaic images of different crops at the early seedling stage from several Russian regions.

The process of markup of these images was carried out in the QGIS Desktop software (https://qgis.org (accessed on 1 October 2023)). Plant centers were manually marked, and the resulting images were saved in vector format as a markup file (Figure 2). Because the scale of the orthomosaics differed in some cases by almost a factor of 2, all orthomosaics were resized to the same scale (1 pix = 1 cm) before analysis. This procedure was performed using the resize() function of the Python Imaging Library by means of an appropriate normalization factor. Because the input of the neural network has to be supplied with a markup image in raster format, raster masks were generated from the vector data at the same scale. The mask included the background (black color) and plant centers (white circles with a radius of 4 cm).
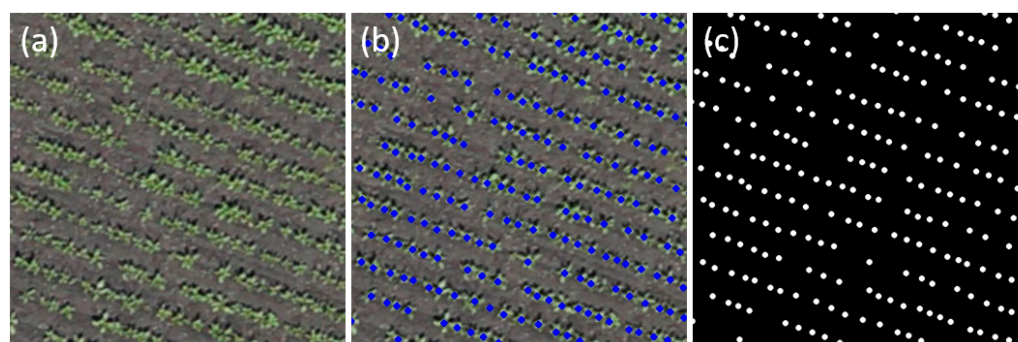

Figure 2. Examples of images for the analysis. (**a**) A fragment of an orthomosaic before markup; (**b**) the same fragment with vector markup applied in QGIS; (**c**) a generated raster mask showing the location of plant centers.

The list of 12 orthomosaic images is given in Table 1. The total area of the crop fields in the orthomosaics was 23.494 ha, and the total number of plants was 610,067.

Table 1. Orthomosaic images of fields in Russian regions in 2019–2023.

Dataset Name	Crop	Location	Imaging Date	Inter-Row Distance, cm	Resolution GSD cm/px	Area, ha	Number of Plants
Beet_marat_0	Sugar beet	MARAT	2023.06.10	45	1.966	0.924	64,750
Beet_marat_1	Sugar beet	MARAT	2023.06.10	45	1.966	0.924	74,394
Krasnodar	Sugar beet	IVANOVSK	2022.06.23	72	1.778	6.801	75,653
Krasnodar_1	Sunflower	NOVOMINSK	2022.06.23	93	1.032	4.156	92,709
Stavropol_1_7	Sunflower	VINODEL	2019.05.25	70	1.791	1.337	59,044
Stavropol_2_7	Potato	VINODEL	2019.05.25	90	1.874	1.336	30,518
Stavropol_4_1	Potato	VINODEL	2019.05.25	90	1.778	1.336	42,148
Stavropol_4_3	Potato	VINODEL	2019.05.25	90	1.778	1.336	35,326
Stavropol_4_9	Potato	VINODEL	2019.05.25	90	1.778	1.336	42,148
Stavropol_2_0	Potato	VINODEL	2019.05.25	90	1.874	1.336	30,283
Stavropol_2_2	Potato	VINODEL	2019.05.25	90	1.874	1.336	33,225
Stavropol_4_0	Potato	VINODEL	2019.05.25	90	1.778	1.336	29,869

Thus, the acquired images provide a large amount of data for the training of deep learning neural network algorithms.

2.2.2. Public Datasets

Additional image datasets were included in this work to extend training sets. They were obtained from the Roboflow service (https://universe.roboflow.com (accessed on 25 November 2023)). The datasets were searched for keywords (crop name, growth stages, field survey, and UAV) in November 2023. In the retrieved datasets, duplicates and images of markup with insufficient quality were excluded. As a result, 14 datasets for five plant species were selected: sugar beet, corn, sunflower, potato, and tobacco. The list of datasets is summarized in Table S2 (Supplementary Material).

Some images were of a high spatial resolution, ~0.42 cm/pix [28] (marked with an asterisk in Table S2). These images were used to build orthomosaics (UBONN_Sb1_2015, UBONN_Sb2_2015, and UBONN_Sb3_2015), which were resized to a common scale and marked as described above.

The rest of the images were converted to a scale of 1 pix = 1 cm. Note that information on spatial resolution was not available for most of the additional images with a lower resolution. Therefore, information about row spacing for different crops was employed to determine an image scale (Table S3, Supplementary Material). The added images contain plants marked by rectangles. Plant centers were identified as the centers of rectangles and marked with a circle having a radius of 4 cm in these images. Weed plants were excluded from these datasets. Based on this marking, a raster mask for each image was generated.

Thus, in this study, the total image sample consisted of 7456 field images in which 362,230 plants of five agricultural crops were marked.

2.2.3. Data Stratification

The data were collected in Russia (Table 1) and the additional data (Table S2, Supplementary Material) were stratified into four datasets. Three datasets (HQ1, HQ2, and HQ3) consisted of orthomosaics; most of them were obtained in Russia in 2019–2023 via a unified acquisition protocol. They have an initial spatial resolution of at least 2 cm/pix, and their labeling was performed manually. They represent high-quality markup images. The fourth dataset (LQ) includes individual frames acquired with the help of UAVs or ground-based imaging. It contains public images from the Roboflow service without the additional manual correction of plant centers in the markup files. These four datasets were balanced in terms of the number of images. The list of included datasets is given in Table 2.

Table 2. Image stratification into four datasets.

Dataset Name	List of Images	Number of Plants	Number of Images
HQ1	Beet_marat_0, Stavropol_4_1, UBONN_Sb1_2015, Krasnodar, Stavropol_1_7, Stavropol_2_0	281,325	1102
HQ2	Stavropol_4_3, UBONN_Sb2_2015, Krasnodar_1, Stavropol_2_2	170,629	721
HQ3	Beet_marat_1, Stavropol_2_7, Stavropol_4_9, UBONN_Sb3_2015, Stavropol_4_0	186,173	459
LQ	BW_C_2021, DSC_SbCSf_2023, FYXDDS_C_2023, HZH_C, NWE_C_2022, NWE_Sb1_2022, NWE_Sb2_2022, NWE_Sf_2022, SEV_Sb_2022, UFMS_C_2023, URLTBK_P_2024, USM_T_2023, VW_C_2022, VW_Sf_2022	334,170	7453

2.3. Neural Network Architecture and Learning Algorithms

In the analysis, each image was divided into 512×512 px tiles using the Random-Crop(512,512) function of the PyTorch 2.0.1 package. Tile images served as the input for a neural network. The U-Net architecture with the ResNet-18 encoder [62] was chosen as a baseline model. The network structure is shown in Figure 3 and includes an encoder part and a decoder part. The encoder consists of convolution layers, with each layer performing convolution operations, normalization batch, ReLU activation functions, and subsampling operations. The output of the encoder has a dimensionality of $512 \times 16 \times 16$. The U-Net decoder is composed of upsampling (backpropagation), convolution layers, concatenation with the corresponding encoder layers, and normalization. The outputs of each decoder layer are concatenated with corresponding encoder layers of the same dimensionality; the last decoder layer has a dimensionality of 512×512, corresponding to a single-channel segmentation mask. Additionally, encoders with ResNet-34 and ResNet-50 architectures were used. They have different numbers of layers and a larger number of parameters. The characteristics of the encoders are given in Table S4 (Supplementary Material).
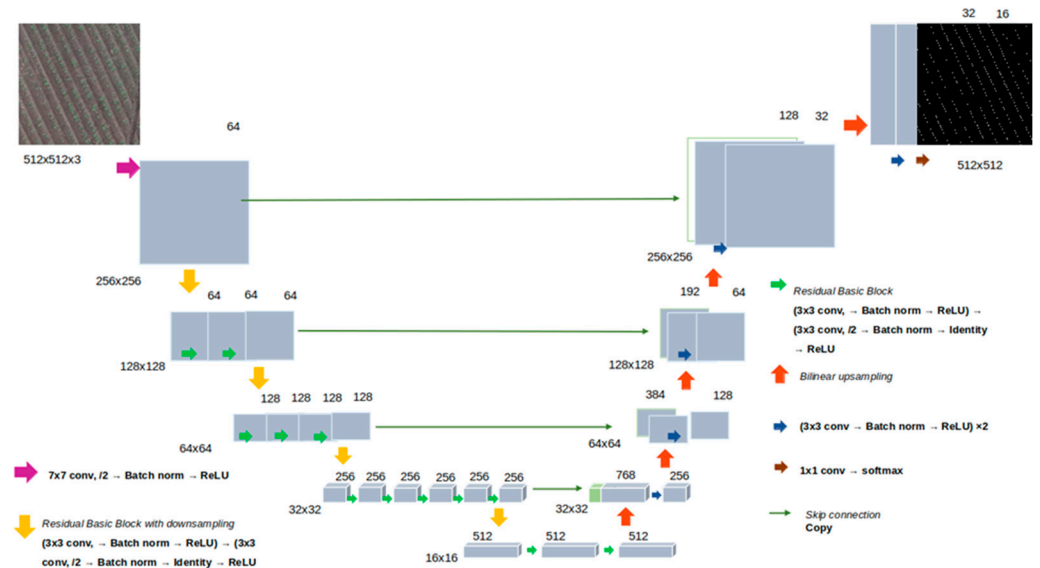

Figure 3. The architecture of the U-Net network used in this work for plant identification.

The Adam adaptive optimizer [63] was utilized to fit network parameters. The learning rate decreased linearly from 10^{-4} to 10^{-6} with a batch size of 8. To optimize model weights, the combined loss function DiceCE was used. It is defined as the sum of Cross Entropy [62] and Dice [64]. Cross Entropy evaluates the quality of a classification. Dice computes a measure of similarity between a predicted mask and the true segmentation mask.

In the training process, we carried out the procedure of image augmentation using the Albumentations library [65]: rotate an image by a random angle in the range from 0 to 90° (method Rotate()), randomly change the scale by a value below 30% (method Random-Scale()), randomly change brightness and contrast (methods RandomBrightnessContrast()

and RandomGamma()), and perform random vertical and horizontal mapping (methods HorizontalFlip() and VerticalFlip()). Data processing was implemented in Python 3.12 and the Pytorch 2.0.1 framework.

The network model was trained for 100 epochs. The best model was selected based on the metrics computed for a validation sample.

Finally, after image segmentation, plant contours were determined in the predicted mask images by the cv2.findContours() method from the OpenCV library [66]. Segmented objects whose area was smaller than a given threshold were excluded as possible noise.

Several experiments were conducted to evaluate the influence of network structure and a combination of datasets in the training/validation and test samples on the accuracy of plant recognition (Table 3). In experiments code-named "HR," validation and testing were based on high-resolution images. In the RN18-LR experiment, low-resolution external images served as the training sample. Experiments RN34-HR-LR and RN50-HR-LR3 differed (from the two other experiments) in that encoders with a larger number of parameters were used: ResNet-34 and ResNet-50 (Table 3).

Table 3. A description of the experiments conducted during the analysis.

Experiment	Training Sample	Validation Sample	Test Sample	Encoder Architecture
RN18-HQ	HQ1	HQ2	HQ3	ResNet-18
RN18-LQ	LQ	HQ2	HQ3	ResNet-18
RN18-HQ-LQ	HQ1+LQ	HQ2	HQ3	ResNet-18
RN34-HQ-LQ	HQ1+LQ	HQ2	HQ3	ResNet-34
RN50-HQ-LQ	HQ1+LQ	HQ2	HQ3	ResNet-50

2.4. Evaluating Accuracy of Plant Identification

The *IoU* (intersection over union) metric was employed to assess segmentation quality for true *X* and predicted *Y* contours:

$$IoU = \frac{X \cap Y}{X \cup Y}.$$

To evaluate the quality of model performance, we used Pearson's correlation coefficient *r*, Spearman's correlation coefficient r_s, mean absolute error *MAE*, and mean absolute percentage error *MAPE*, which were calculated via the formulas

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|,$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|x_i - y_i|}{y_i}$$

where x_i and y_i are the number of plants in the *i*th orthomosaic fragments obtained by manual markup and a neural network, respectively; *n* is the number of such tiles. We utilized non-overlapping tiles of 20×20 m in size.

2.5. The Downstream Analysis of the Processed Orthomosaics

The evaluation of several texture characteristics was carried out as the downstream analysis of the crop image masks obtained either manually or by neural networks. These characteristics are dependent on a mutual arrangement of plants in the images, its regularity, proximity, and other factors. Second-order texture characteristics that are determined on the basis of the Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run-length Matrix (GLRM) [67–69] were chosen for the analysis.

Initially, we evaluated 10 characteristics; however, most of them turned out to be highly correlated. Therefore, four texture features were selected for our analysis: GLCM

mean, GLCM correlation, GLRM long run, and GLRM run ratio. They are defined in Table S5 (Supplementary Material). Eight main directions were used to calculate texture features for neighboring pixels (GLCM features) and for a series (GLRM features): up, down, left, right, and four diagonal directions [70].

Texture characteristics were estimated for images of the test sample obtained in Russia (dataset HR3 except for UBONN_Sb3_2015, Table 2). Nonoverlapping tiles of size 1000×1000 px from orthomosaics were employed to evaluate texture characteristics. Statistical associations were evaluated between true mask tiles (manual markup) and tiles marked by a neural network. Statistical analysis was performed with the help of the numpy library of the Python language.

3. Results

3.1. The Evaluation of the Accuracy of Plant Identification in Different Experiments

Changes in the loss function and *IoU* for plant identification in training and validation samples during training are shown in Figure 4 for four experiments.
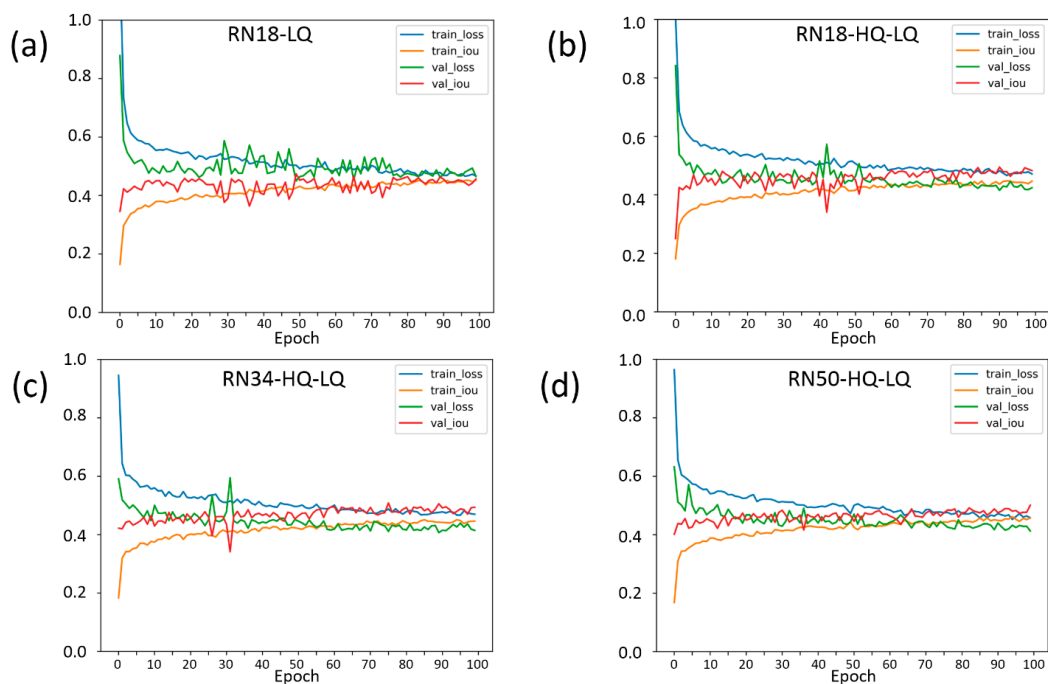

Figure 4. The learning curves of the models corresponding to the experiments: (**a**) RN18-LQ; (**b**) RN18-HQ-LQ; (**c**) RN34-HQ-LQ; and (**d**) RN50-HQ-LQ. On the X-axis, the ID numbers of epochs during training are plotted. The Y-axis shows parameters characterizing the magnitude of error obtained with the training and validation samples (see the panels in the top-right corner of the graphs). Blue curve: change in the loss function on the training sample; green curve: change in the loss function on the validation sample; yellow curve: change in the *IoU* metric on the training sample; red curve: change in the *IoU* metric on the validation sample.

As presented in the graph in Figure 4a, only on a low-resolution image sample (RN18-LQ) network topology does the training ResNet-18 network yield high-amplitude fluctuations in the loss function and *IoU* at the initial stage of the training process. The variation in these parameters stabilized around epoch 75 (Figure 4a). In an experiment with the same network architecture and both high- and low-resolution data as the training sample (RN18-HQ-LQ, Figure 4b), the loss and *IoU* stabilized at 50 epochs. For the RN34-HQ-LQ experiment (Figure 4c), where a network architecture with a larger number of parameters was used, ResNet-34, the loss, and *IoU* stabilized at epoch 35. The use of the ResNet-50

network architecture and low- and high-resolution images in the training process showed a smooth change in accuracy characteristics without substantial spikes (Figure 4d).

Table 4 summarizes the accuracy assessment metrics for the experiments described in Table 4. The table contains the average values of the metrics for the five orthomosaics included in the test sample. As one can see in the table, the use of low-resolution data and the ResNet-50 encoder yielded the best results (experiment RN50-HQ-LQ): the Pearson's coefficient of correlation between the number of plants determined manually and the predicted number of plants was greater than 0.98. The parameters *MAPE* and *MAE* for the best model were the smallest. The model is able to segment sunflower, potato, and sugar beet plants at an early stage of growth, regardless of light and other conditions. Nonetheless, Spearman's correlation coefficient and *IoU* were not the greatest for the RN50-HQ-LQ experiment. In terms of r_s, it was ranked #2 and in terms of *IoU* #3 by value.

Table 4. Accuracy measures for different experiments and images from the HR3 sample. The best parameter value in a column is shown in bold, and the worst value is underlined.

Experiment	Epoch Number for the Best Model	*MAE*	*MAPE*, %	*r*	r_s	*IoU*
RN18-HQ	79	96	6.22	0.9816	<u>0.9363</u>	**0.3753**
RN18-LQ	50	<u>127</u>	7.90	0.9850	0.9597	<u>0.3166</u>
RN18-HQ-LQ	79	105	5.57	0.9883	**0.9600**	0.3693
RN34-HQ-LQ	76	95	5.84	0.9878	0.9489	0.3600
RN50-HQ-LQ	83	**78**	**5.20**	0.9885	0.9571	0.3688

A comparison of the accuracy parameters between RN18-HQ and RN18-LQ revealed a difference between the cases of HQ and LQ data used for training. RN18-HR yielded better estimates for *MAPE*, *MAE*, and *IoU*. In the RN18-LQ experiment, these parameters showed lower performance (note that the test dataset contained HQ data only). On the other hand, both correlation coefficients were smaller in the RN18-HR experiment and larger in RN18-LQ. This finding implies that using the LQ dataset decreases the accuracy of finding a plant center (a lower value of *IoU* in all experiments). Nevertheless, the larger number of images in this dataset (Table 3) decreased the variance for the estimate of the number of plants in the image (higher correlation coefficients). Of note, using both the HQ and LQ datasets in training (experiment RN18-HQ-LQ) yielded better plant count estimates (*MAE*, *MAPE*, *r*, and r_s) but a lower *IoU* value in comparison with the RN18-HQ experiment.

Examples of plant images for different crops and manual and automatic markups are shown in Figure 5. The agreement between automatic and manual markups is high. Nonetheless, some errors could be detected in the masks obtained by means of neural networks: some contours for plants were found to be elongated due to the merging of neighboring plant contours; in some cases, small contours appeared between plants.

Table 5 shows the results of accuracy estimation for the best model (RN50-HQ-LQ) for orthomosaics from the test sample. For most images, accuracy is approximately the same (*MAPE* 1–6%), except for plant recognition for the Stavropol_4_9 orthomosaic. The latter showed lower performance (*MAPE* above 12%). This could be explained by a larger proportion of touching plants in the images obtained from this field (Figure 5e).

Table 5. Accuracy metrics of the best model for the RN50-HQ-LQ experiment with orthomosaics from test samples. The best parameter value in a column is highlighted in bold, and the worst is underlined.

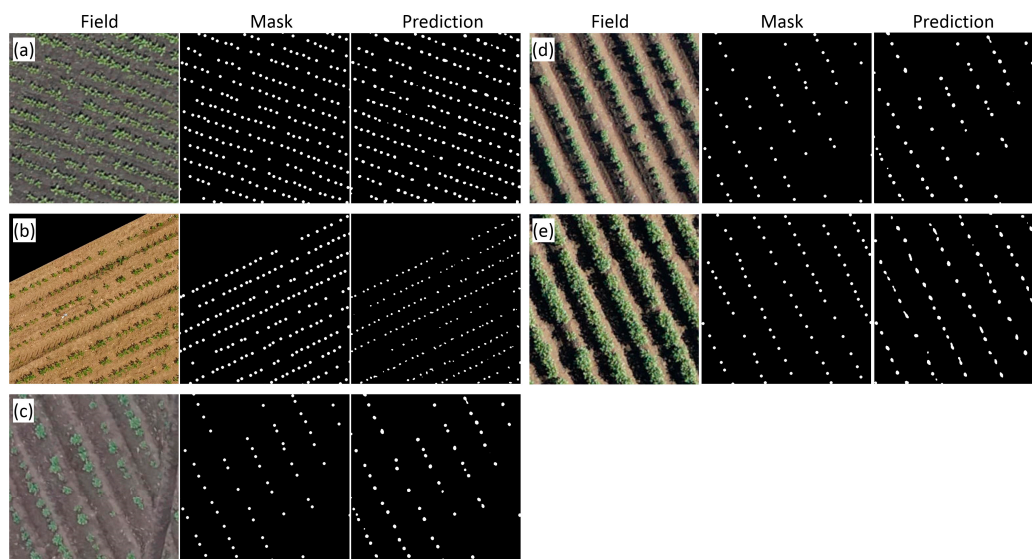| Orthomosaic Image	*MAE*	*MAPE*, %	*R*	r_s
Beet_marat_1	66	**1.27**	<u>0.9734</u>	0.9464
Stavropol_2_7	18	2.30	**0.9995**	0.9870
Stavropol_4_9	<u>223</u>	<u>12.75</u>	0.9880	<u>0.8790</u>
UBONN_Sb3_2015	**4**	3.19	0.9987	**0.9950**
Stavropol_4_0	79	6.51	0.9828	0.9781

Figure 5. Examples of RN50-HQ-LQ model performance on the test sample for different crops and high-resolution orthomosaics. (**a**) Sugar beet, Beet_marat_1; (**b**) sugar beet, UBONN_Sb3_2015; (**c**) potato, Stavropol_2_7; (**d**) potato, Stavropol_4_0; (**e**) potato, Stavropol_4_9. Images in rows from left to right: original (Field); manual plant marking (Mask); automatic marking by the RN50-HQ-LQ network.

We evaluated the performance of the algorithms on a desktop PC with the following configuration: CPU, Intel® Core™ i5-8265U @ 1.60 GHz (Intel Corporation, Santa Clara, CA, USA); GPU, NVIDIA GeForce RTX 2080 12 G (Nvidia Corporation, Santa Clara, CA, USA); GPU environment, CUDA 11.2; OS, Windows 10; software, Python 3.12; and framework, Pytorch 2.0.1. The performance was tested for files larger than 3 GB with an image size of 50,000 × 50,000 px. The estimated computational performance of the RN50-HQ-LQ network is enough to process a large amount of data within a reasonable time, at least 50 ha per hour on the desktop PC with the GPU.

3.2. The Influence of Plant Identification Accuracy on Subsequent Analysis

A downstream analysis of the plants in orthomosaics via the evaluation of the texture features was performed next. The plant masks obtained manually and predicted by the networks of the best (RN50-HQ-LQ) and lower (RN18-HQ) performance were used for texture analysis.

A comparison of the mean values for the estimates of the four texture features among different orthomosaics and markups is presented in Table S6 (Supplementary Material). First of all, this table indicates that there are noticeable differences (all of them statistically significant) in mean texture parameters (for the manually marked images) between Beet_marat_1 and the other three datasets. For example, the mean value of the GLCM mean parameter for the Beet_marat_1 dataset is 14.36, while for the other datasets it ranges from 4.5 to 6.5. Similar differences were noted in the other parameters. This is because the Beet_marat_1 dataset has a row spacing of 45 cm, whereas for the other orthomosaics, it is 90 cm (see Table 1). Thus, the estimated texture characteristics may characterize the cropping patterns in the images.

A comparison of the mean values of the features obtained from manually and automatically labeled masks for different datasets revealed that, for the predicted masks, the differences are significant in most cases. Nonetheless, deviations of the mean values of texture features for the more accurate mapping, RN50-HQ-LQ, were overall smaller as compared to the less accurate one, RN18-HQ. For example, for the GLCM mean parameter for the Beet_marat_1 data (14.36), the differences in the estimates between the manually made

mask and RN50-HQ-LQ (14.15) are insignificant, whereas in comparison with RN18-HQ (15.00), the differences are significant. Note that for the predicted masks and features based on the length of the gray level series (GLRM longRun and GLRM runRatio), the estimates of the mean based on the manual masks and the predicted ones differ by a factor of almost 2, while the differences in the predicted masks among different datasets are small. This is probably because, unlike the manual marking, in which plant centers are marked with circles that do not touch each other, the marking obtained with the help of neural networks gives elongated areas for plants that often touch each other (Figure 5e). Errors of this kind are more common for the less accurate RN18-HQ network.

The characteristics of the linear relationship between the manual markup (mask) and the predicted markups are shown in Figure 6. The figure suggests that the estimates obtained with neural networks—compared to the manual approach—have systematic biases (the slope and intercept differ from 1.0 and 0, respectively). The correlation coefficients for parameters and datasets in some cases are close to 1.0. In other cases, they deviate strongly from 1.0, even to the point of not being significant (RN18-HQ predictions, the Beet_marat_1 dataset; Figure 6d).
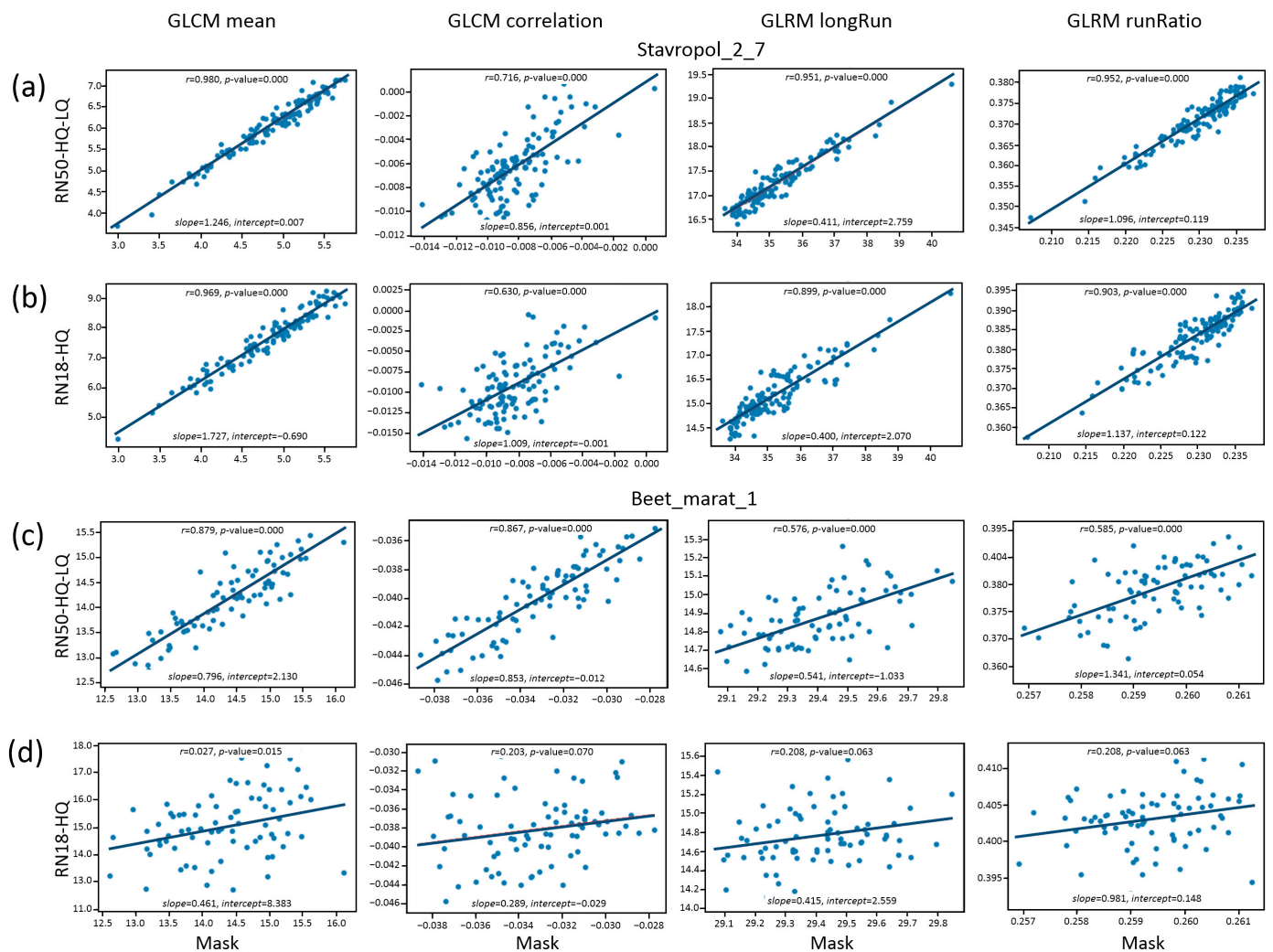

Figure 6. A comparison of crop texture characteristics estimates between markup obtained by the manual approach (X axis) and markup obtained by neural network algorithm (Y axis). The names of characteristics are shown at the top of the figure. (**a**) Stavropol_2_7, prediction by the RN50-HQ-LQ method; (**b**) Stavropol_2_7, prediction by the RN18-HQ method; (**c**) Beet_marat_1, prediction by the RN50-HQ-LQ method; (**d**) Beet_marat_1, prediction by the RN18-HQ method.

Examples of the masks of some tiles having strong deviations of the GLCM mean parameter for the predicted masks are shown in Figure S1 (Supplementary Material). This figure presents characteristic errors in the marking of orthomosaics by the neural network: the shape of markers differs from the circular one, touching contours are observed, and some markers deviate from a row's direction. These errors are more pronounced for the markup obtained by the RN18-HQ neural network. Apparently, such errors lead to the deviation of the texture parameters described above.

Visual analysis of the examples of mask images and the results of neural network prediction (Figure S1) demonstrates that for the network with a lower quality of prediction (RN-18-HQ), the images contain more noise (in comparison with a more accurate algorithm, RN50-HQ-LQ). In particular, this noise results from the inclusion of small objects between regularly spaced plant centers. Because of the noise, the RN-18-HQ images appear less homogeneous and regular. This appears to lead to significantly greater distortions in the estimation of the texture characteristics by the RN-18-HQ model (Figure 6). Note also that for the Beet_marat_1 dataset, the plant spacing is smaller than for Stavropol_2_7. This may further amplify the effect of noise when evaluating the texture characteristics (images become less homogeneous). For example, for Stavropol_2_7 images, the less accurate RN-18-HQ network gives estimates of textural characteristics with rather high values of correlation coefficients when compared to the mask, and for the RN-18-HQ and Beet_marat_1 dataset, the correlation coefficients lose significance. In general, it can be assumed that the result of the comparison of texture features is influenced on the one hand by the accuracy of prediction (the presence of noise) and, on the other hand, by the specificity of the plant arrangement pattern (in particular, the distance between rows).

Thus, the effect of estimation accuracy on the subsequent analysis of recognized plant masks strongly depends on which texture parameters are estimated. Nevertheless, overall, for a prediction by a more accurate network, the deviation of mean values of texture features appears to be closer to the deviation obtained via manual partitioning.

4. Discussion

In this work, a method was developed for counting plants in field images obtained from a UAV. The neural network approach was used to solve the problem of image segmentation based on the U-Net architecture and the ResNet encoder. The algorithm was tested on several agricultural crops and manifested high accuracy. The estimates of plant-counting accuracy for the best model proved to be comparable to the accuracy of both computer vision-based and deep-learning-based techniques. Computer vision-based approaches achieve relative root mean square error (RRMSE) values between 2 and 4%, depending on the field and crop (maize and sunflower) [25]. When maize plants were counted at different growth stages, RRMSE accuracy estimates of 2–6% were achieved in that study, and more accurately for an earlier stage of plant development. For later stages, when plants begin to touch each other, the accuracy of this method drops severalfold [71]. When plants are counted in images by means of the U-Net neural network, *MAPE* estimates range from 4% to 16%, depending on the field and crop [40]. Using the RiceNet neural network to count rice shoots results in *MAE* values ranging from 3 to 4 [72].

It has been shown that image spatial resolution affects the performance of neural network deep learning algorithms for various tasks [73–76]. Spatial resolution is critical for an analysis of field images obtained from UAVs for plant counting [23,25,77] and sizing [46]. In the present work, images of different resolutions as well as plant markup quality were employed for the training and testing of neural network algorithms. The HQ set was acquired by a UAV camera via a uniform protocol; the images were assembled into orthophotos followed by manual markup. The LQ set represented fragmented images of

fields at different resolutions, whose markup was determined by the recalculation of the positions of the centers of the frames marking the plants. Using only the HQ set for training allows for the more accurate localization of plants in an image. By contrast, using the LQ set, due to the larger number of images, enables us to obtain higher correlations between true and predicted plant counts. Therefore, combining these data for training—despite the differences in resolution and markup quality—significantly improves the accuracy of plant recognition.

Errors in plant identification in the field images inevitably cause biases in crop density and plant spacing estimation [21,55]. The present work illustrates how the precision of plant detection in field images may affect their regularity estimates by means of several texture characteristics. For some characteristics and datasets, the correlation of estimates between images marked manually and images marked automatically is high (GLCM mean, GLRM long run, GLRM run ratio, Stavropol_2_7 dataset). For others, the differences are large and may lead to erroneous conclusions about the properties of field cropping patterns. Thus, overall, methods with a higher accuracy of automatic plant prediction and counting should give estimates of texture parameters close to those derived from manually marked images.

5. Conclusions

Segmentation method using deep learning algorithms was developed for detecting and counting plants of five species in RGB images acquired from a UAV. Several CNN models based on the U-Net architecture with different encoders (ResNet-18, ResNet-34, ResNet-50) were implemented. They were trained using orthomosaics with high quality markup obtained at several locations in Russia and additional datasets of various spatial resolutions and markup quality from the Roboflow service. The performance of several neural networks and training datasets was evaluated. The advantage of usage both the high- and low-quality marked images in neural networks training was demonstrated. This strategy yielded better plant count estimates but a lower performance of the plant location in the images (*IoU*). Several texture features characterizing cropping patterns were estimated and compared for manually evaluated and neural-network-estimated plant masks. For some of the texture characteristics (GLCM mean, GLRM long run, GLRM run ratio) the estimates between images marked manually and automatically are close. For others, the differences are large and may lead to erroneous conclusions about the properties of field cropping patterns. In general, plant detection algorithms with a higher accuracy show better agreement with the estimates of texture parameters obtained from manually marked images.

Author Contributions: Conceptualization, M.A.G., Z.A.Z. and D.A.A.; methodology, M.V.K., M.A.G., E.G.K. and Z.A.Z.; software, M.V.K., M.A.G. and E.G.K.; validation, M.A.G., Z.A.Z. and D.A.A.; formal analysis, M.V.K., M.A.G. and E.G.K.; investigation, M.V.K. and M.A.G.; resources, D.A.A.; data curation, M.V.K. and Z.A.Z.; writing—original draft preparation, M.V.K. and D.A.A.; writing—review and editing, D.A.A.; visualization, M.V.K. and E.G.K.; supervision, Z.A.Z. and D.A.A.; project

References

1. Tsouros, D.C.; Bibi, S.; Sarigiannidis, P.G. A Review on UAV-Based Applications for Precision Agriculture. *Information* **2019**, *10*, 349. [CrossRef]
2. Maes, W.H.; Steppe, K. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [CrossRef] [PubMed]
3. Velusamy, P.; Rajendran, S.; Mahendran, R.K.; Naseer, S.; Shafiq, M.; Choi, J.-G. Unmanned Aerial Vehicles (UAV) in Precision Agriculture: Applications and Challenges. *Energies* **2021**, *15*, 217. [CrossRef]
4. Radoglou-Grammatikis, P.; Sarigiannidis, P.; Lagkas, T.; Moscholios, I. A Compilation of UAV Applications for Precision Agriculture. *Comput. Netw.* **2020**, *172*, 107148. [CrossRef]
5. Shu, M.; Fei, S.; Zhang, B.; Yang, X.; Guo, Y.; Li, B.; Ma, Y. Application of UAV Multisensor Data and Ensemble Approach for High-Throughput Estimation of Maize Phenotyping Traits. *Plant Phenomics* **2022**, *2022*, 9802585. [CrossRef]
6. Guo, W.; Carroll, M.E.; Singh, A.; Swetnam, T.L.; Merchant, N.; Sarkar, S.; Singh, A.K.; Ganapathysubramanian, B. UAS-Based Plant Phenotyping for Research and Breeding Applications. *Plant Phenomics* **2021**, *2021*, 9840192. [CrossRef] [PubMed]
7. Ge, X.; Ding, J.; Jin, X.; Wang, J.; Chen, X.; Li, X.; Liu, J.; Xie, B. Estimating Agricultural Soil Moisture Content through UAV-Based Hyperspectral Images in the Arid Region. *Remote Sens.* **2021**, *13*, 1562. [CrossRef]
8. Bah, M.D.; Hafiane, A.; Canals, R. Deep Learning with Unsupervised Data Labeling for Weed Detection in Line Crops in UAV Images. *Remote Sens.* **2018**, *10*, 1690. [CrossRef]
9. Liu, S.; Jin, X.; Nie, C.; Wang, S.; Yu, X.; Cheng, M.; Shao, M.; Wang, Z.; Tuohuti, N.; Bai, Y.; et al. Estimating Leaf Area Index Using Unmanned Aerial Vehicle Data: Shallow vs. Deep Machine Learning Algorithms. *Plant Physiol.* **2021**, *187*, 1551–1576. [CrossRef] [PubMed]
10. Hu, P.; Chapman, S.C.; Zheng, B. Coupling of Machine Learning Methods to Improve Estimation of Ground Coverage from Unmanned Aerial Vehicle (UAV) Imagery for High-Throughput Phenotyping of Crops. *Funct. Plant Biol.* **2021**, *48*, 766–779. [CrossRef]
11. Sahoo, R.N.; Rejith, R.G.; Gakhar, S.; Ranjan, R.; Meena, M.C.; Dey, A.; Mukherjee, J.; Dhakar, R.; Meena, A.; Daas, A.; et al. Drone Remote Sensing of Wheat N Using Hyperspectral Sensor and Machine Learning. *Precis. Agric.* **2024**, *25*, 704–728. [CrossRef]
12. Volpato, L.; Pinto, F.; González-Pérez, L.; Thompson, I.G.; Borém, A.; Reynolds, M.; Gérard, B.; Molero, G.; Rodrigues, F.A. High Throughput Field Phenotyping for Plant Height Using UAV-Based RGB Imagery in Wheat Breeding Lines: Feasibility and Validation. *Front. Plant Sci.* **2021**, *12*, 591587. [CrossRef] [PubMed]
13. Antolínez García, A.; Cáceres Campana, J.W. Identification of Pathogens in Corn Using Near-Infrared UAV Imagery and Deep Learning. *Precis. Agric.* **2023**, *24*, 783–806. [CrossRef]
14. Yang, Q.; Shi, L.; Han, J.; Yu, J.; Huang, K. A near Real-Time Deep Learning Approach for Detecting Rice Phenology Based on UAV Images. *Agric. For. Meteorol.* **2020**, *287*, 107938. [CrossRef]
15. Sadeghi-Tehran, P.; Sabermanesh, K.; Virlet, N.; Hawkesford, M.J. Automated Method to Determine Two Critical Growth Stages of Wheat: Heading and Flowering. *Front. Plant Sci.* **2017**, *8*, 252. [CrossRef] [PubMed]
16. Feng, A.; Zhou, J.; Vories, E.; Sudduth, K.A. Prediction of Cotton Yield Based on Soil Texture, Weather Conditions and UAV Imagery Using Deep Learning. *Precis. Agric.* **2024**, *25*, 303–326. [CrossRef]

17. Xu, X.; Nie, C.; Jin, X.; Li, Z.; Zhu, H.; Xu, H.; Wang, J.; Zhao, Y.; Feng, H. A Comprehensive Yield Evaluation Indicator Based on an Improved Fuzzy Comprehensive Evaluation Method and Hyperspectral Data. *Field Crops Res.* **2021**, *270*, 108204. [CrossRef]
18. Yang, T.; Zhang, W.; Zhou, T.; Wu, W.; Liu, T.; Sun, C. Plant Phenomics & Precision Agriculture Simulation of Winter Wheat Growth by the Assimilation of Unmanned Aerial Vehicle Imagery into the WOFOST Model. *PLoS ONE* **2021**, *16*, e0246874. [CrossRef]
19. Pathak, H.; Igathinathane, C.; Zhang, Z.; Archer, D.; Hendrickson, J. A Review of Unmanned Aerial Vehicle-Based Methods for Plant Stand Count Evaluation in Row Crops. *Comput. Electron. Agric.* **2022**, *198*, 107064. [CrossRef]
20. Booth, D.T.; Cox, S.E.; Meikle, T.; Zuuring, H.R. Ground-Cover Measurements: Assessing Correlation Among Aerial and Ground-Based Methods. *Environ. Manag.* **2008**, *42*, 1091–1100. [CrossRef] [PubMed]
21. Jin, X.; Liu, S.; Baret, F.; Hemerlé, M.; Comar, A. Estimates of Plant Density of Wheat Crops at Emergence from Very Low Altitude UAV Imagery. *Remote Sens. Environ.* **2017**, *198*, 105–114. [CrossRef]
22. Valente, J.; Sari, B.; Kooistra, L.; Kramer, H.; Mücher, S. Automated Crop Plant Counting from Very High-Resolution Aerial Imagery. *Precis. Agric.* **2020**, *21*, 1366–1384. [CrossRef]
23. Buters, T.; Belton, D.; Cross, A. Seed and Seedling Detection Using Unmanned Aerial Vehicles and Automated Image Classification in the Monitoring of Ecological Recovery. *Drones* **2019**, *3*, 53. [CrossRef]
24. Lu, D.; Ye, J.; Wang, Y.; Yu, Z. Plant Detection and Counting: Enhancing Precision Agriculture in UAV and General Scenes. *IEEE Access* **2023**, *11*, 116196–116205. [CrossRef]
25. Bai, Y.; Nie, C.; Wang, H.; Cheng, M.; Liu, S.; Yu, X.; Shao, M.; Wang, Z.; Wang, S.; Tuohuti, N.; et al. A Fast and Robust Method for Plant Count in Sunflower and Maize at Different Seedling Stages Using High-Resolution UAV RGB Imagery. *Precis. Agric.* **2022**, *23*, 1720–1742. [CrossRef]
26. Calvario, G.; Alarcón, T.E.; Dalmau, O.; Sierra, B.; Hernandez, C. An Agave Counting Methodology Based on Mathematical Morphology and Images Acquired through Unmanned Aerial Vehicles. *Sensors* **2020**, *20*, 6247. [CrossRef]
27. Alt, V.V.; Pestunov, I.A.; Melnikov, P.V.; Elkin, O.V. Automated Detection of Weeds and Evaluation of Crop Sprouts Quality Based on RGB Images. *Sib. Her. Agric. Sci.* **2019**, *48*, 52–60. [CrossRef]
28. Chebrolu, N.; Labe, T.; Stachniss, C. Robust Long-Term Registration of UAV Images of Crop Fields for Precision Agriculture. *IEEE Robot. Autom. Lett.* **2018**, *3*, 3097–3104. [CrossRef]
29. García-Martínez, H.; Flores-Magdaleno, H.; Khalil-Gardezi, A.; Ascencio-Hernández, R.; Tijerina-Chávez, L.; Vázquez-Peña, M.A.; Mancilla-Villa, O.R. Digital Count of Corn Plants Using Images Taken by Unmanned Aerial Vehicles and Cross Correlation of Templates. *Agronomy* **2020**, *10*, 469. [CrossRef]
30. Osco, L.P.; Marcato Junior, J.; Marques Ramos, A.P.; De Castro Jorge, L.A.; Fatholahi, S.N.; De Andrade Silva, J.; Matsubara, E.T.; Pistori, H.; Gonçalves, W.N.; Li, J. A Review on Deep Learning in UAV Remote Sensing. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102456. [CrossRef]
31. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [CrossRef] [PubMed]
32. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
33. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
34. Mohimont, L.; Alin, F.; Rondeau, M.; Gaveau, N.; Steffenel, L.A. Computer Vision and Deep Learning for Precision Viticulture. *Agronomy* **2022**, *12*, 2463. [CrossRef]
35. Hafiz, A.M.; Bhat, G.M. A Survey on Instance Segmentation: State of the Art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189. [CrossRef]
36. Hoeser, T.; Kuenzer, C. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sens.* **2020**, *12*, 1667. [CrossRef]
37. Mittal, P.; Singh, R.; Sharma, A. Deep Learning-Based Object Detection in Low-Altitude UAV Datasets: A Survey. *Image Vis. Comput.* **2020**, *104*, 104046. [CrossRef]
38. Barreto, A.; Lottes, P.; Ispizua Yamati, F.R.; Baumgarten, S.; Wolf, N.A.; Stachniss, C.; Mahlein, A.-K.; Paulus, S. Automatic UAV-Based Counting of Seedlings in Sugar-Beet Field and Extension to Maize and Strawberry. *Comput. Electron. Agric.* **2021**, *191*, 106493. [CrossRef]
39. Lottes, P.; Behley, J.; Milioto, A.; Stachniss, C. Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2870–2877. [CrossRef]
40. Vong, C.N.; Conway, L.S.; Zhou, J.; Kitchen, N.R.; Sudduth, K.A. Early Corn Stand Count of Different Cropping Systems Using UAV-Imagery and Deep Learning. *Comput. Electron. Agric.* **2021**, *186*, 106214. [CrossRef]
41. Zhang, C.; Atkinson, P.M.; George, C.; Wen, Z.; Diazgranados, M.; Gerard, F. Identifying and Mapping Individual Plants in a Highly Diverse High-Elevation Ecosystem Using UAV Imagery and Deep Learning. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 280–291. [CrossRef]

42. Kitano, B.T.; Mendes, C.C.T.; Geus, A.R.; Oliveira, H.C.; Souza, J.R. Corn Plant Counting Using Deep Learning and UAV Images. *IEEE Geosci. Remote Sens. Lett.* **2024**, 1–5. [CrossRef]

43. Li, F.; Bai, J.; Zhang, M.; Zhang, R. Yield Estimation of High-Density Cotton Fields Using Low-Altitude UAV Imaging and Deep Learning. *Plant Methods* **2022**, *18*, 55. [CrossRef] [PubMed]

44. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

45. Ding, R.; Luo, J.; Wang, C.; Yu, L.; Yang, J.; Wang, M.; Zhong, S.; Gu, R. Identifying and Mapping Individual Medicinal Plant Lamiophlomis Rotata at High Elevations by Using Unmanned Aerial Vehicles and Deep Learning. *Plant Methods* **2023**, *19*, 38. [CrossRef] [PubMed]

46. Machefer, M.; Lemarchand, F.; Bonnefond, V.; Hitchins, A.; Sidiropoulos, P. Mask R-CNN Refitting Strategy for Plant Counting and Sizing in UAV Imagery. *Remote Sens.* **2020**, *12*, 3015. [CrossRef]

47. David, E.; Daubige, G.; Joudelat, F.; Burger, P.; Comar, A.; De Solan, B.; Baret, F. Plant Detection and Counting from High-Resolution RGB Images Acquired from UAVs: Comparison between Deep-Learning and Handcrafted Methods with Application to Maize, Sugar Beet, and Sunflower. *bioRxiv* **2022**. [CrossRef]

48. Hosseiny, B.; Rastiveis, H.; Homayouni, S. An Automated Framework for Plant Detection Based on Deep Simulated Learning from Drone Imagery. *Remote Sens.* **2020**, *12*, 3521. [CrossRef]

49. Mhango, J.K.; Harris, E.W.; Green, R.; Monaghan, J.M. Mapping Potato Plant Density Variation Using Aerial Imagery and Deep Learning Techniques for Precision Agriculture. *Remote Sens.* **2021**, *13*, 2705. [CrossRef]

50. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

51. Oh, S.; Chang, A.; Ashapure, A.; Jung, J.; Dube, N.; Maeda, M.; Gonzalez, D.; Landivar, J. Plant Counting of Cotton from UAS Imagery Using Deep Learning-Based Object Detection Framework. *Remote Sens.* **2020**, *12*, 2981. [CrossRef]

52. Li, H.; Wang, P.; Huang, C. Comparison of Deep Learning Methods for Detecting and Counting Sorghum Heads in UAV Imagery. *Remote Sens.* **2022**, *14*, 3143. [CrossRef]

53. Wang, B.; Zhou, J.; Costa, M.; Kaeppler, S.M.; Zhang, Z. Plot-Level Maize Early Stage Stand Counting and Spacing Detection Using Advanced Deep Learning Algorithms Based on UAV Imagery. *Agronomy* **2023**, *13*, 1728. [CrossRef]

54. Osco, L.P.; Nogueira, K.; Marques Ramos, A.P.; Faita Pinheiro, M.M.; Furuya, D.E.G.; Gonçalves, W.N.; De Castro Jorge, L.A.; Marcato Junior, J.; Dos Santos, J.A. Semantic Segmentation of Citrus-Orchard Using Deep Neural Networks and Multispectral UAV-Based Imagery. *Precis. Agric.* **2021**, *22*, 1171–1188. [CrossRef]

55. Vong, C.N.; Conway, L.S.; Feng, A.; Zhou, J.; Kitchen, N.R.; Sudduth, K.A. Corn Emergence Uniformity Estimation and Mapping Using UAV Imagery and Deep Learning. *Comput. Electron. Agric.* **2022**, *198*, 107008. [CrossRef]

56. Pérez-Ortiz, M.; Peña, J.M.; Gutiérrez, P.A.; Torres-Sánchez, J.; Hervás-Martínez, C.; López-Granados, F. Selecting Patterns and Features for Between- and within- Crop-Row Weed Mapping Using UAV-Imagery. *Expert Syst. Appl.* **2016**, *47*, 85–94. [CrossRef]

57. 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; IEEE: Singapore, 2017; pp. 3024–3031.

58. Kamath, R.; Balachandra, M.; Prabhu, S. Crop and Weed Discrimination Using Laws' Texture Masks. *Int. J. Agric. Biol. Eng.* **2020**, *13*, 191–197. [CrossRef]

59. Chetverikov, D. Pattern Regularity as a Visual Key. *Image Vis. Comp.* **2000**, *18*, 975–985. [CrossRef]

60. Ngan, H.Y.T.; Pang, G.K.H. Regularity Analysis for Patterned Texture Inspection. *IEEE Trans. Automat. Sci. Eng.* **2009**, *6*, 131–144. [CrossRef]

61. Sun, H.-C.; Kingdom, F.A.A.; Baker, C.L. Perceived Regularity of a Texture Is Influenced by the Regularity of a Surrounding Texture. *Sci. Rep.* **2019**, *9*, 1637. [CrossRef] [PubMed]

62. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241, ISBN 978-3-319-24573-7.

63. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.

64. Drozdzal, M.; Vorontsov, E.; Chartrand, G.; Kadoury, S.; Pal, C. The Importance of Skip Connections in Biomedical Image Segmentation. In *Deep Learning and Data Labeling for Medical Applications*; Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., et al., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2016; Volume 10008, pp. 179–187, ISBN 978-3-319-46975-1.

65. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and Flexible Image Augmentations. *Information* **2020**, *11*, 125. [CrossRef]

66. Bradski, G.; Kaehler, A. *Learning OpenCV: Computer Vision with the OpenCV Library*; O'Reilly Media, Inc.: Newton, MA, USA, 2008.

67. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural Features for Image Classification. *IEEE Trans.Syst. Man Cybern.* **1973**, *6*, 610–621. [CrossRef]

68. Majumdar, S.; Jayas, D.S. Classification of Bulk Samples of Cereal Grains Using Machine Vision. *J. Agric. Eng. Res.* **1999**, *73*, 35–47. [CrossRef]

69. Galloway, M.M. Texture Analysis Using Gray Level Run Lengths. *Comput. Graph. Image Process.* **1975**, *4*, 172–179. [CrossRef]

70. Hall-Beyer, M. GLCM Texture: A Tutorial. *NCGIA Remote Sens. Core Curric.* **2000**, *3*, 75.

71. Che, Y.; Wang, Q.; Zhou, L.; Wang, X.; Li, B.; Ma, Y. The Effect of Growth Stage and Plant Counting Accuracy of Maize Inbred Lines on LAI and Biomass Prediction. *Precis. Agric.* **2022**, *23*, 2159–2185. [CrossRef]

72. Bai, X.; Liu, P.; Cao, Z.; Lu, H.; Xiong, H.; Yang, A.; Cai, Z.; Wang, J.; Yao, J. Rice Plant Counting, Locating, and Sizing Method Based on High-Throughput UAV RGB Images. *Plant Phenomics* **2023**, *5*, 0020. [CrossRef] [PubMed]

73. Dodge, S.; Karam, L. Understanding How Image Quality Affects Deep Neural Networks. In Proceedings of the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016; IEEE: Lisbon, Portugal, 2016; pp. 1–6.

74. Koziarski, M.; Cyganek, B. Impact of Low Resolution on Image Recognition with Deep Neural Networks: An Experimental Study. *Int. J. Appl. Math. Comput. Sci.* **2018**, *28*, 735–744. [CrossRef]

75. Artemenko, N.V.; Genaev, M.A.; Epifanov, R.U.; Komyshev, E.G.; Kruchinina, Y.V.; Koval, V.S.; Goncharov, N.P.; Afonnikov, D.A. Image-Based Classification of Wheat Spikes by Glume Pubescence Using Convolutional Neural Networks. *Front. Plant Sci.* **2024**, *14*, 1336192. [CrossRef]

76. Neupane, B.; Horanont, T.; Hung, N.D. Deep Learning Based Banana Plant Detection and Counting Using High-Resolution Red-Green-Blue (RGB) Images Collected from Unmanned Aerial Vehicle (UAV). *PLoS ONE* **2019**, *14*, e0223906. [CrossRef]

77. De Castro, A.; Torres-Sánchez, J.; Peña, J.; Jiménez-Brenes, F.; Csillik, O.; López-Granados, F. An Automatic Random Forest-OBIA Algorithm for Early Weed Mapping between and within Crop Rows Using UAV Imagery. *Remote Sens.* **2018**, *10*, 285. [CrossRef]