

## Article

# Estimation of Greenhouse Lettuce Growth Indices Based on a Two-Stage CNN Using RGB-D Images

Min-Seok Gang <sup>1,2</sup> , Hak-Jin Kim <sup>1,2,3,\*</sup> and Dong-Wook Kim <sup>1</sup>

<sup>1</sup> Department of Biosystems Engineering, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea; msg1907@snu.ac.kr (M.-S.G.); dwk8033@snu.ac.kr (D.-W.K.)

<sup>2</sup> Integrated Major in Global Smart Farm, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

<sup>3</sup> Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

\* Correspondence: kimhj69@snu.ac.kr; Tel.: +82-2-880-4604

**Abstract:** Growth indices can quantify crop productivity and establish optimal environmental, nutritional, and irrigation control strategies. A convolutional neural network (CNN)-based model is presented for estimating various growth indices (i.e., fresh weight, dry weight, height, leaf area, and diameter) of four varieties of greenhouse lettuce using red, green, blue, and depth (RGB-D) data obtained using a stereo camera. Data from an online autonomous greenhouse challenge (Wageningen University, June 2021) were employed in this study. The data were collected using an Intel RealSense D415 camera. The developed model has a two-stage CNN architecture based on ResNet50V2 layers. The developed model provided coefficients of determination from 0.88 to 0.95, with normalized root mean square errors of 6.09%, 6.30%, 7.65%, 7.92%, and 5.62% for fresh weight, dry weight, height, diameter, and leaf area, respectively, on unknown lettuce images. Using red, green, blue (RGB) and depth data employed in the CNN improved the determination accuracy for all five lettuce growth indices due to the ability of the stereo camera to extract height information on lettuce. The average time for processing each lettuce image using the developed CNN model run on a Jetson SUB mini-PC with a Jetson Xavier NX was 0.83 s, indicating the potential for the model in fast real-time sensing of lettuce growth indices.

**Keywords:** convolutional neural network (CNN); growth estimation; growth index; growth monitoring; greenhouse; lettuce; red, green, blue, and depth (RGB-D); stereo camera



**Citation:** Gang, M.-S.; Kim, H.-J.; Kim, D.-W. Estimation of Greenhouse Lettuce Growth Indices Based on a Two-Stage CNN Using RGB-D Images. *Sensors* **2022**, *22*, 5499. <https://doi.org/10.3390/s22155499>

Academic Editors: Jiangtao Qi and Haiye Yu

Received: 19 May 2022

Accepted: 20 July 2022

Published: 23 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Fresh weight, dry weight, height, diameter, and leaf area are growth indicators of farmland and horticultural crops because they are closely associated with transpiration and photosynthesis [1–3]. The accurate measurement of these indices enables us to quantify and model crop growth for bioenvironmental factors. Climate, nutrient, and irrigation control strategies can be established based on well-developed crop models to obtain optimal productivity. Growth indices can also help farmers determine appropriate transplanting or harvesting times to increase their economic status. These applications of growth index measurements may be even more effective, especially in greenhouses, where strict environmental controls are possible.

Traditional methods for measuring growth indicators are costly because operators typically collect plant samples and use analytical instruments in a laboratory. In addition, the reliability of the observed values depends on the sampling methods and sample conditions [4]. Some indices can be calculated from physical traits, such as leaf length and width [5,6]. However, direct manual measurements are still required, and the nonlinear relationship between traits can negatively affect model accuracy. Therefore, methods that estimate growth indices using sensors and crop image data have been effective alternatives

to traditional methods for many researchers. These methods allow for the efficient collection of various indices [7–10].

Recently, the focus has been on deep learning techniques to automatically extract and use key features from numerous images. Deep neural networks are advantageous for solving nonlinear problems and can learn a relational model based on input and output through multiple hidden layers. Deep learning techniques do not require handcrafted features selected by experts and provide a relatively fast processing speed in inference compared to traditional computer vision-based methods [11]. Research has been conducted on crop species classification [12–14], plant and fruit localization [14,15], and pest diagnosis [16] by applying deep learning techniques. Studies have also been conducted on estimating quantitative growth indicators [14,17–22] and classifying or detecting objects.

Several researchers have applied segmentation using deep learning techniques to specify crop areas from red, green, blue (RGB) and three-dimensional (3D) data to estimate crop growth indices [18–21]. However, only the linear regression, random forest, and simple artificial neural network (ANN) models were used to estimate growth indices using features extracted from the segmented area. There are potential limitations in expressing the relationship between growth indicators and various features, such as color and texture. In addition, the segmentation was computationally costly and could increase image processing time, making it difficult to apply to real-time agricultural applications.

In this regard, a regression using a convolutional neural network (CNN) can be an alternative. The CNN has advantages in image processing and can reduce the processing time by effectively learning image features through kernel structures in the form of a 2D array. Zhang et al. developed a CNN model using RGB images and estimated the leaf fresh weight, leaf dry weight, and leaf area of greenhouse lettuce. They demonstrated that the estimation of these indicators could be superior to the estimation in the support vector machine, random forest, and linear regression models [22].

In addition, 3D data, such as RGB-D images, may provide more information on vertical changes in leaves and stems [23–26]. Quan et al. developed a novel two-stream CNN model based on RGB-D images to estimate the fresh weight of aboveground weeds in a field on a high-end graphics processing unit (GPU) (2080Ti, NVIDIA, Santa Clara, CA, USA) [27]. The two-stream CNN model used a multi-input single-output (MISO) structure and dense network in the network blocks. Quan et al. revealed that using depth images and a CNN effectively estimated the vertical changes as weeds grow. However, the dataset consisted of weeds with small-scale fresh weights and relatively short growth periods.

Several researchers used the public RGB-D dataset of greenhouse lettuces [28,29] published by Hemming et al. [30], whose dataset was built for the online challenge session of the Third Autonomous Greenhouse Challenge. Raja et al. developed deep learning models, including multi-input multi-output (MIMO) and MISO models, using a CNN and a deformable CNN (DCNN) to estimate growth indices [28]. Raja et al. reported a normalized mean squared error (NMSE) of 0.068 for MIMO and 0.069 for MISO. Li and Wang generated point cloud data from the same RGB-D dataset [30] and estimated the height of lettuce through a CNN and recurrent neural network [29].

Nevertheless, developing a CNN model using 3D data to estimate growth indices more effectively was necessary due to relatively poor performance in predicting the biophysical properties of lettuces in the late growth stage. It is important to investigate model applicability over the entire growth period of harvestable crops with quantitative comparisons. In many crops, vertical growth increases toward the later growth stages [31–33]. Moreover, [31–33] found that the height of the leafy and fruit vegetables increases quadratically or exponentially in the late growth stages until saturation. The estimation accuracy for the late stage may affect the performance of deep learning models [34]. In addition, the attainment of real-time processing speed on the embedded board should also be considered for agricultural applications of the neural network model that can handle various growth indicators.

Studies using the CNN and RGB-D models have applied MISO and MIMO structures. A two-stage architecture connects the two individual models in series and may extend

and improve these structures. Two-stage architecture might be effective for data with time-series characteristics [30,35–38].

In this study, a two-stage CNN model was presented using RGB-D data to estimate various growth indicators (i.e., fresh weight, dry weight, height, diameter, and leaf area) of greenhouse lettuces in real time. In particular, the model estimation performance for the entire growth period was quantitatively presented. The RGB-D images published by Heming et al. [30] for four different varieties of lettuces (i.e., Lugano, Salanova, Aphylion, and (Satine) were used. The coefficient of determination ( $R^2$ ), root mean square error (RMSE), and normalized RMSE (NRMSE) were determined to evaluate the performance of the developed model for each growth index. In addition, the processing speeds per image were analyzed to investigate the real-time performance of the developed model on an embedded board.

## 2. Materials and Methods

### 2.1. Dataset

The dataset in this study was originally produced by Hemming et al. [30] for the online challenge session of the Third Autonomous Greenhouse Challenge. The dataset is publicly available at the 4TU.ResearchData website and consists of 388 RGB images, depth images, and actual aboveground data corresponding to a pair of RGB images and depth images, respectively. The RGB images are 24-bit portable network graphic (PNG) images with three channels, and the depth images are 8-bit PNG images with a single channel. The resolution of all the images is  $1080 \times 1920$  pixels (Figure 1).



**Figure 1.** RGB images of sampled Salanova, Lugano, Satine, and Aphylion cultivar lettuce grown for seven weeks (growing stages at one-week intervals after transplanting). The images are cropped to  $512 \times 512$  pixels from the center. The dataset is available in online: <https://doi.org/10.4121/15023088.v1> [30].

An Intel camera (RealSense D415, Intel, Santa Clara, CA, USA) was used for image acquisition. The D415 is equipped with RGB and stereo cameras to extract depth information. In principle, the D415 calculates depth information using the difference in the distance between the corresponding points of the left and right camera images. The data were collected by fixing the camera 0.9 m from the top of the plant through a metal frame to photograph the plant from top to bottom, maintaining the same shooting height and resolution.

In the dataset, four varieties of lettuce, namely Lugano, Salanova, Aphylion, and Satine, were photographed. The number of image pairs for each variety was 96, 102, 92, and 98, respectively. The plants were hydroponically grown in an experimentally controlled greenhouse in Bleiswijk, Netherlands. Data were acquired weekly from approximately 70 sampled plants at each growth stage for seven weeks from transplanting the seedlings until the crops reached a harvestable size.

Shoot fresh weight, shoot dry weight, height, diameter, and leaf area were obtained using a destructive measurement method (Table 1). Fresh weight was obtained by measuring the weight of lettuce harvested from the point where the first leaf was attached, and dry weight was measured after 3 days of drying in an oven after obtaining the fresh weight. Leaf area was calculated through the surface area projected onto a plane after separating the leaves from the stem, and the increase in leaf area due to leaf curvature was not considered. The diameter of the lettuce projected onto a plane was measured, and the height was measured from where the first leaf was attached to the highest point of the plant. Table 2 lists the range of each data measurement for the lettuce varieties.

**Table 1.** Growth index data format for each RGB image of the lettuce measured using destructive methods.

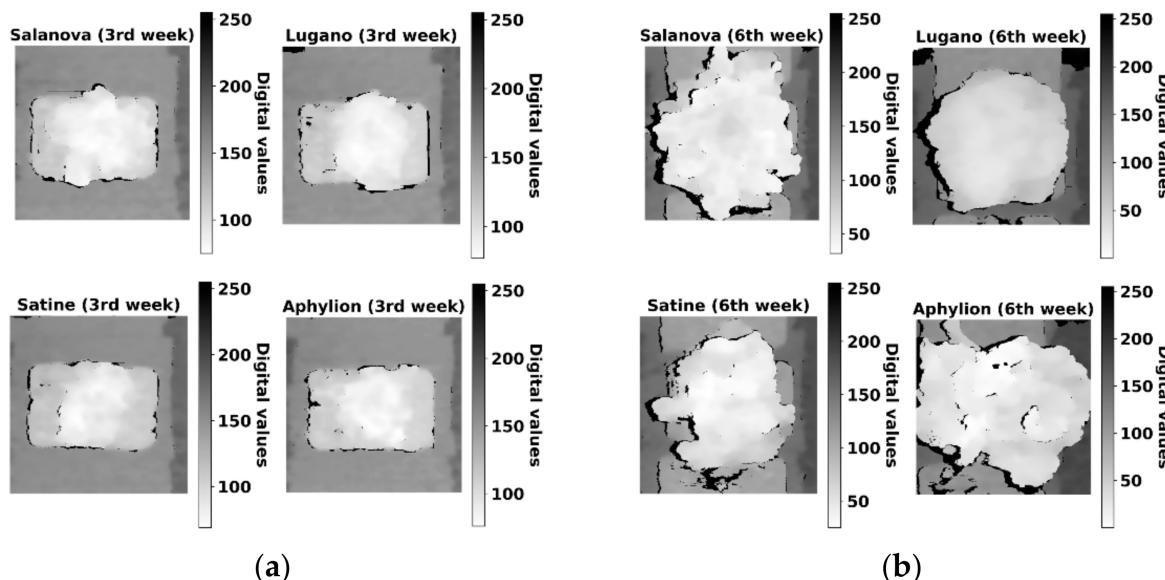
Week	RGB-Image (Salanova)	Fresh Weight (g)	Dry Weight (g)	Height (cm)	Diameter (cm)	Leaf Area (cm <sup>2</sup> )
1		5.2	0.58	9.8	17.2	202.7
2		16.4	1.22	6.8	18.5	520.1
3		69.8	4.16	9.0	25.1	1694.3
4		85.0	5.02	8.0	25.0	2008.8
5		110.0	6.04	12.0	28.0	2414.7
6		133.2	6.84	15.8	32.0	3089.2
7		236.5	11.04	17.0	33.5	5348.1

**Table 2.** Maximum and minimum values for each growth index within the dataset measured using destructive methods.

	Fresh Weight (g)	Dry Weight (g)	Height (cm)	Diameter (cm)	Leaf Area (cm <sup>2</sup> )
Minimum	1.4	0.09	4.3	8.2	57.6
Maximum	459.7	20.1	25	37	5868

Figure 2 illustrates an example of the lettuce depth image data. The depth images were aligned to have the same pixel area as the RGB images. As the value of each pixel in the depth image data was between 681 and 1306, it was difficult to visualize the depth data directly. Therefore, the pixel values were normalized using the maximum and minimum values of Equation (1) and were multiplied by 255, the maximum reflection value for the grayscale expression:

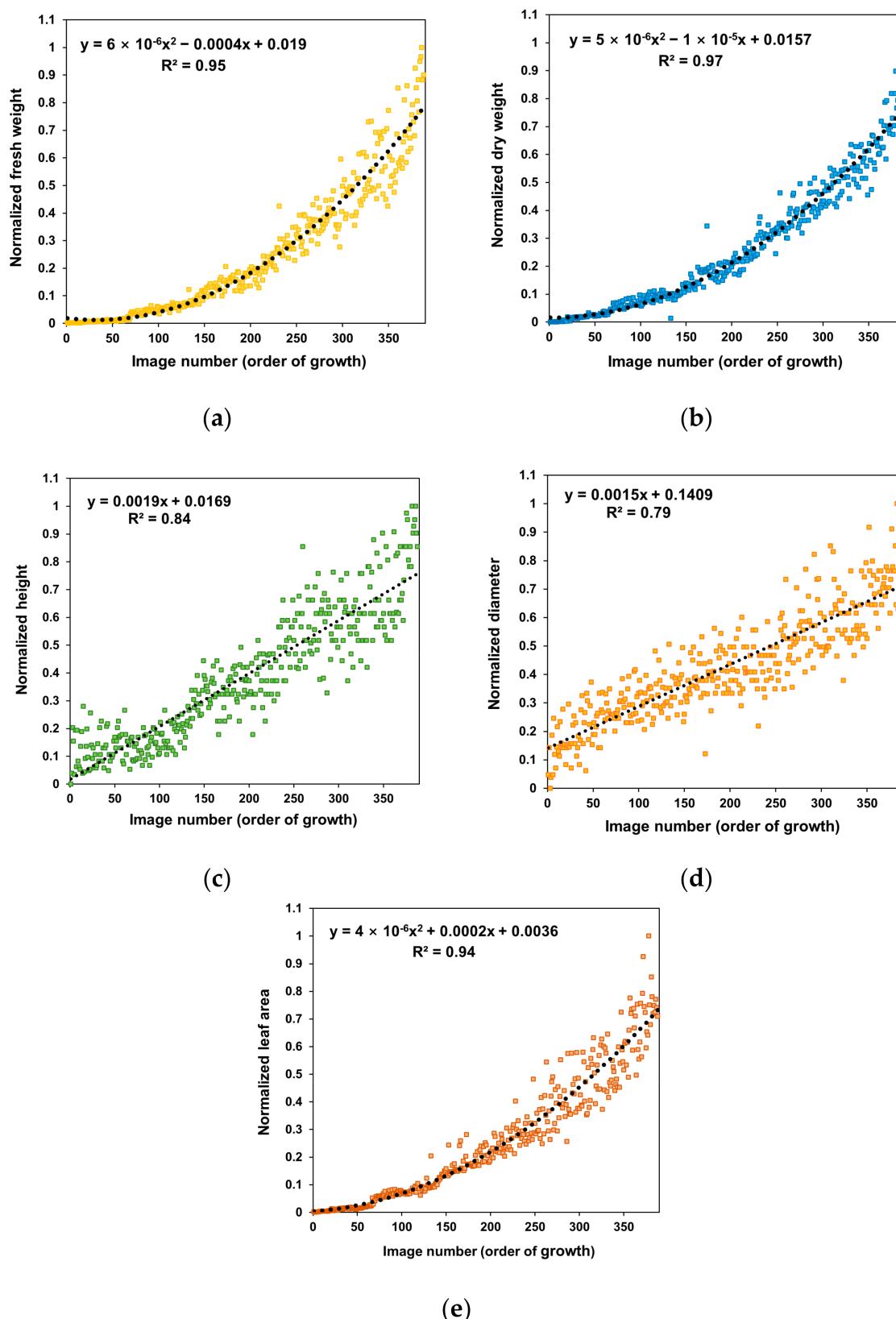
$$x_{\text{normalized}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$



**Figure 2.** Normalized depth images with grayscale to visualize digital values from 681 to 1306 (high: close to the camera, low: far from the camera): (a) depth images of the four lettuce varieties 3 weeks after transplanting; and (b) depth images of the four lettuce varieties 6 weeks after transplanting.

The normalized values were converted into a grayscale image. A value closer to zero indicates a closer distance between the plant and camera, and larger values indicate a farther distance between the plant and camera.

We also examined the characteristics of the dataset. Figure 3 depicts the normalized growth index values during the cultivation period of the entire dataset. As the crop grew, the height increased with a steeper slope than the diameter. As a result, the change rates of the fresh weight, dry weight, and leaf area increased in a quadratic equation, and the rates became higher toward the later growth stages. These growth patterns of the lettuce were consistent with previous studies [8,39].



**Figure 3.** Change rates of growth indices in the dataset according to the crop growth ordered as the largest normalized growth index effect obtained from the product of each index: (a) fresh weight, (b) dry weight, (c) height, (d) diameter, and (e) leaf area.

## 2.2. Image Preprocessing

A total of 388 image pairs were used to build the model: 338 pairs for learning and 50 pairs for estimating unknown samples. A parallel input comprising RGB and depth images was generated for model training. First, the RGB images were cropped to  $512 \times 512$  pixels from the center points of the plants so that the area information could be used for learning while maintaining the resolution of the input image to preserve the input pixel information. The center position of the plants was manually specified. For the depth images, the crop area was decreased to  $256 \times 256$  pixels to minimize the effect of the background. The k-nearest neighbor method was applied to fill the pixels for which the depth image data were not acquired, and the average value of nine pixels was used [27,40].

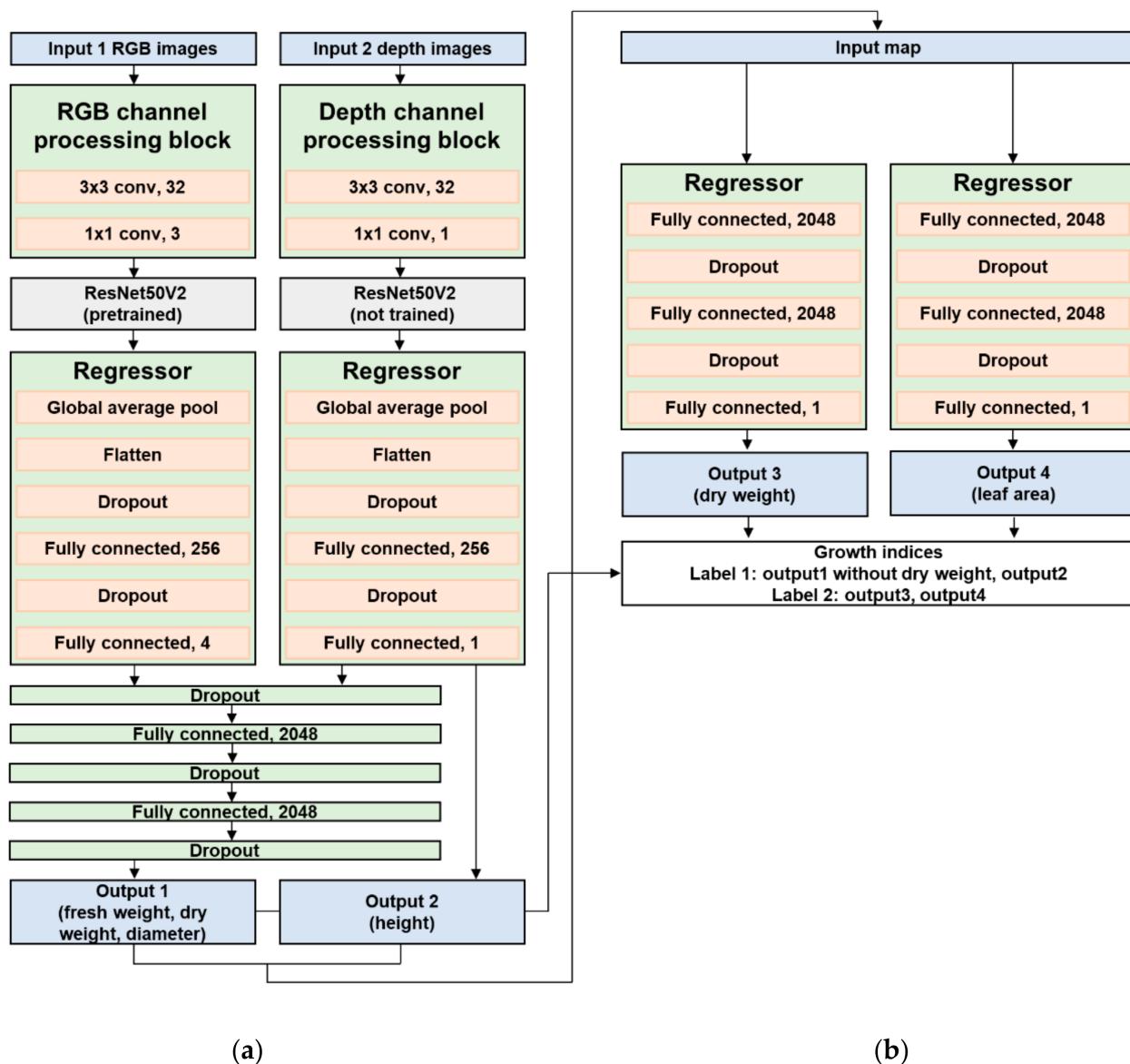
All data were normalized to a value between zero and one through Equation (1). The RGB images were normalized using a maximum value of 255 and a minimum value of zero, which are the reflection values in the visible band. The depth images were normalized using a maximum value of 1306 and a minimum value of 681 from the measurements. The actual aboveground data used as labels were normalized to values between zero and one through Equation (1).

If data are insufficient for deep learning model training, errors increase greatly, leading to underfitting and poor model performance on the training and unknown datasets. In addition, overfitting may occur, deteriorating model performance on unknown data due to the large variance in the predicted values [41]. In this context, data augmentation was applied to sufficiently supplement the data. Data augmentation during learning involved rotating the original images at an arbitrary angle within  $180^\circ$ , horizontal inversion, and arbitrary image translation. The original train dataset of 338 pairs was extended to 119,868 pairs through augmentation.

## 2.3. Network Architecture and Training

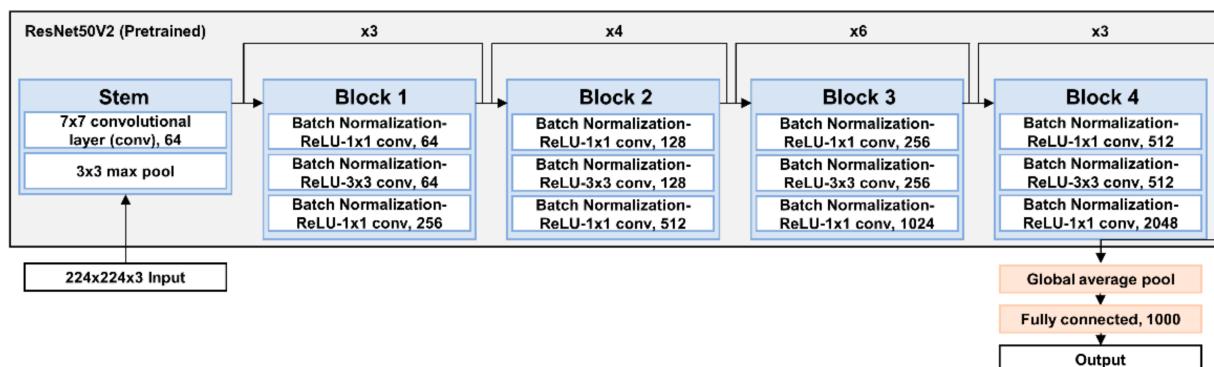
A two-stage CNN model was proposed to obtain fresh weight, dry weight, height, diameter, and leaf area values for the input RGB and depth images (Figure 4). The model architecture was developed using the TensorFlow 2.5.1 framework and Python 3.8. The model structure consists of a model containing two individual models, unlike the MISO and MIMO models in the previous studies [27,28]. Two-stage architecture might be effective for data with time-series characteristics by connecting the two models in series [35–38]. In the first stage, the model performed learning and inference on fresh weight, dry weight, diameter, and height using RGB and depth images. The estimated fresh weight, diameter, and height were used as final model outputs. In the second stage, the results estimated in the previous model are used as input for a model to estimate the leaf area and correct the dry weight values.

In the first stage, transfer learning was applied to the CNN structure of the RGB image input model, and a pretrained model with the ImageNet dataset [42] was used. ResNet50V2 [43] (i.e., the improved structure of ResNet50 [44]) was used as the pretrained model architecture. When a CNN is trained through backpropagation, as the number of layers becomes deeper, the differential values for the loss between the target and predicted data decrease, and the weights of each layer filter that affect the model output are learned with small values. This phenomenon, called gradient vanishing, made the input data untrainable, consequently degrading model accuracy. However, ResNet solved the gradient vanishing problem by dividing each neural network layer into residual blocks and omitting each block through a skip connection. Moreover, ResNetV2 improved ResNet using the activation function in front of the convolutional layer in the residual block (Figure 5).



(a) (b)

**Figure 4.** Two-stage architecture. (a) First stage: parallel RGB and depth input models based on untrained ResNet50V2 to estimate fresh weight, dry weight, diameter, and height. (b) Second stage: parallel artificial neural network regression model to estimate the dry weight and leaf area.



**Figure 5.** CNN architecture of ResNet50V2, which was used as backbone model of presented model.

Two separate convolutional layers ( $3 \times 3$  and  $1 \times 1$ ) were added before the pretrained model because the image size was larger than the original input size of the pretraining model. After the pretrained model was applied, an additional fully connected layer was placed to perform training suitable for the input dataset. We also tried to prevent overfitting by setting the dropout layers.

The architecture for the depth input model was designed similarly to the RGB model. However, without pretraining for ResNet50V2, the depth images were newly learned. The results of the RGB and depth input models were concatenated and passed through two fully connected layers with 2048 nodes and dropout layers.

In the second stage, the ANN model with two fully connected layers with 2048 nodes and dropout layers was deployed identically to the last part of the first stage to use the values inferred in the first stage for final growth prediction. Losses in all layers were calculated using the mean squared error measurements between the predicted value and actual label. Adam was used as an optimization function to minimize the result value of the loss function [45], and the rectified linear unit was applied as an activation function [46].

The model was trained using a desktop with Windows (Windows 10, Microsoft, Redmond, WA, USA), an Intel i7 central processing unit (CPU; i7-11700, Intel, Santa Clara, CA, USA), and a GeForce RTX GPU (RTX 3090, NVIDIA, Santa Clara, CA, USA) with 64 GB of RAM. The hyperparameter settings were as follows: The initial learning rate was set to 0.001, and 200-epoch learning was performed. If the validation loss did not decrease for eight iterations, the learning rate was halved by setting the learning rate scheduler. If the validation loss did not change for 30 iterations, early stopping was applied to reduce unnecessary learning time. The input batch size was set to 32. When the loss on the validation dataset became the smallest value, the model weights were preserved. In addition, to avoid overfitting from fixing a validation dataset,  $k$ -fold cross-validation was applied [47]. In  $k$ -fold cross-validation, a training dataset is divided into  $k$  pieces, and training is performed  $k - 1$  times. For each training iteration, training is performed with  $k - 1$  datasets, and one dataset is used for validation. This study applied the commonly used five-fold cross-validation [48,49].

#### 2.4. Model Evaluation

The growth indices were estimated using a testing dataset not used for training, and the  $R^2$ , RMSE, and NRMSE between the estimated and actual values were analyzed to evaluate the estimation accuracy of the developed model during the entire cultivation period. The NRMSE was calculated using Equation (2) with the maximum and minimum values for the entire dataset.

Demonstrating the effectiveness of the developed model in the late growth period can improve model applicability. The training results using only RGB images were compared with those using RGB and depth images in 20 samples of late growth. We analyzed an overall change of five growth indices using Equations (1) and (3), and the change increased rapidly in the later 20 images:

$$\text{NRMSE} = \frac{\text{RMSE}}{x_{\max} - x_{\min}} \quad (2)$$

$$\begin{aligned} \text{Normalized growth index effect} &= \prod \text{Growth index}_{\text{normalized}} \\ \text{Growth index} &= \{\text{Fresh weight}, \text{Dry weight}, \text{Height}, \text{Diameter}, \text{Leaf area}\} \end{aligned} \quad (3)$$

In addition, to numerically compare the performance with a previous study [28], the NMSE value was calculated using Equation (4):

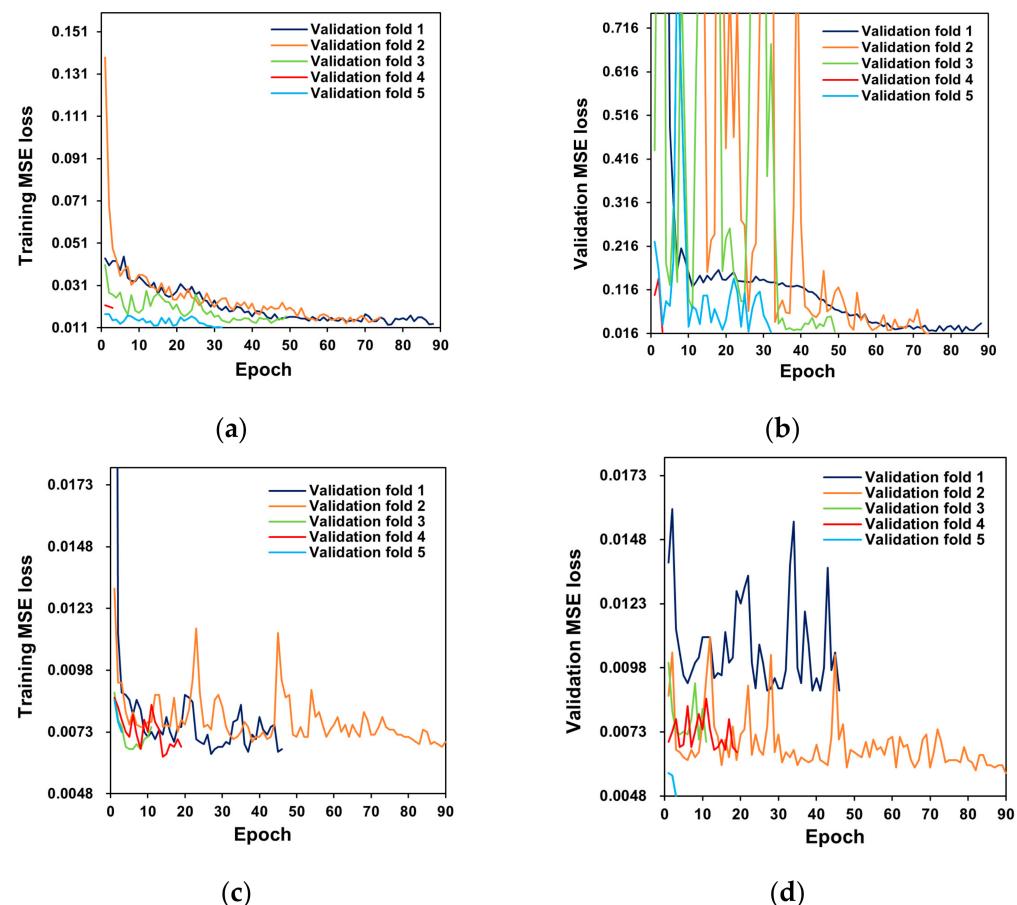
$$\text{NMSE} = \sum_{j=1}^m \frac{\sum_{i=1}^n (gt_{ij} - p_{ij})^2}{\sum_{i=1}^n (gt_{ij})^2} \quad (4)$$

where  $gt$  is the actual growth indices,  $p$  is the estimated value,  $n$  is the number of images, and  $m$  is the number of growth indices.

The loading and inference times per image were measured on the desktop for training, and an embedded board (Jetson SUB mini-PC, Nanshan, Shenzhen, China) based on a Jetson Xavier (Jetson Xavier NX, NVIDIA, Santa Clara, CA, USA) module was used to evaluate the real-time image processing performance of the developed model. The Jetson SUB mini-PC was operated on an Ubuntu 18.04 operating system equipped with a 64-bit Carmel ARM (Carmel v8.2, NVIDIA, Santa Clara, CA, USA) CPU, and a Volta GPU (Volta, NVIDIA, Santa Clara, CA, USA) with 8 GB of RAM.

### 3. Results and Discussion

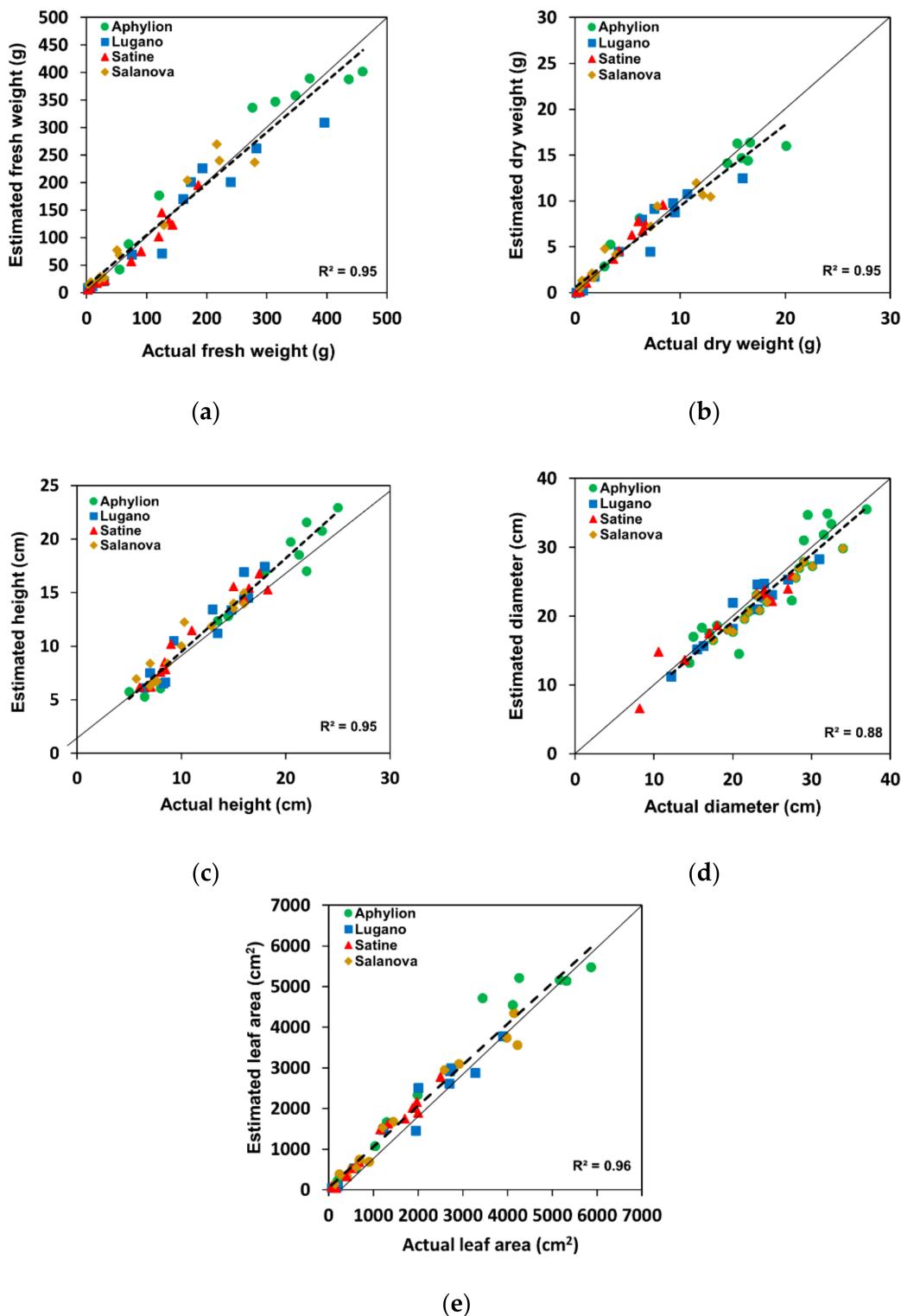
Figure 6 illustrates the changes in training loss and validation loss while training the first and second model stages. In the first and second stages, the validation loss tends to decrease as the training loss decreases. This result revealed that the gap between the training loss and validation loss narrowed. The model weight was not saved or passed on to the next training with another validation fold using the best checkpoint when the validation loss did not converge in every cross-validation cycle to prevent overfitting, where the validation loss diverges and increases.



**Figure 6.** Change in training loss and validation loss during training of the developed model: (a) training loss in the first stage, (b) validation loss in the first stage, (c) training loss in the second stage, and (d) validation loss in the second stage. Repetition of loss reduction depicts five-fold cross-validation. The model weight was not kept when the validation loss did not converge in every fold to avoid overfitting.

In deep learning applications, checkpointing operates similarly to early stopping [50], and early stopping is one of the representative strategies to prevent overfitting, although checkpointing does not necessarily eliminate overfitting [51].

Figure 7 and Table 3 compare the lettuce growth indicators estimated by the developed model using RGB-D and the actual measured values. The results reveal an  $R^2$  greater than 0.88 and a slope greater than 0.90 for all growth indices, implying that the developed CNN model built using RGB-D images could estimate with considerable accuracy.



**Figure 7.** Comparison of growth indicators estimated from the developed model using RGB-D and those measured using the destructive method. The color indices present the values for each lettuce variety: (a) fresh weight, (b) dry weight, (c) height, (d) diameter, and (e) leaf area.

**Table 3.** Growth index estimation of the developed two-stage CNN model using RGB-D images.

	Fresh Weight (g)	Dry Weight (g)	Height (cm)	Diameter (cm)	Leaf Area (cm <sup>2</sup> )
R <sup>2</sup>	0.95	0.95	0.95	0.89	0.96
RMSE <sup>1</sup>	27.85	1.26	1.53	2.28	326.04
NRMSE <sup>2</sup> (%)	6.09	6.30	7.65	7.92	5.62

<sup>1</sup> Root mean square error. <sup>2</sup> Normalized root mean square error. Ratio of RMSE to the difference between the maximum and minimum value for the entire dataset.

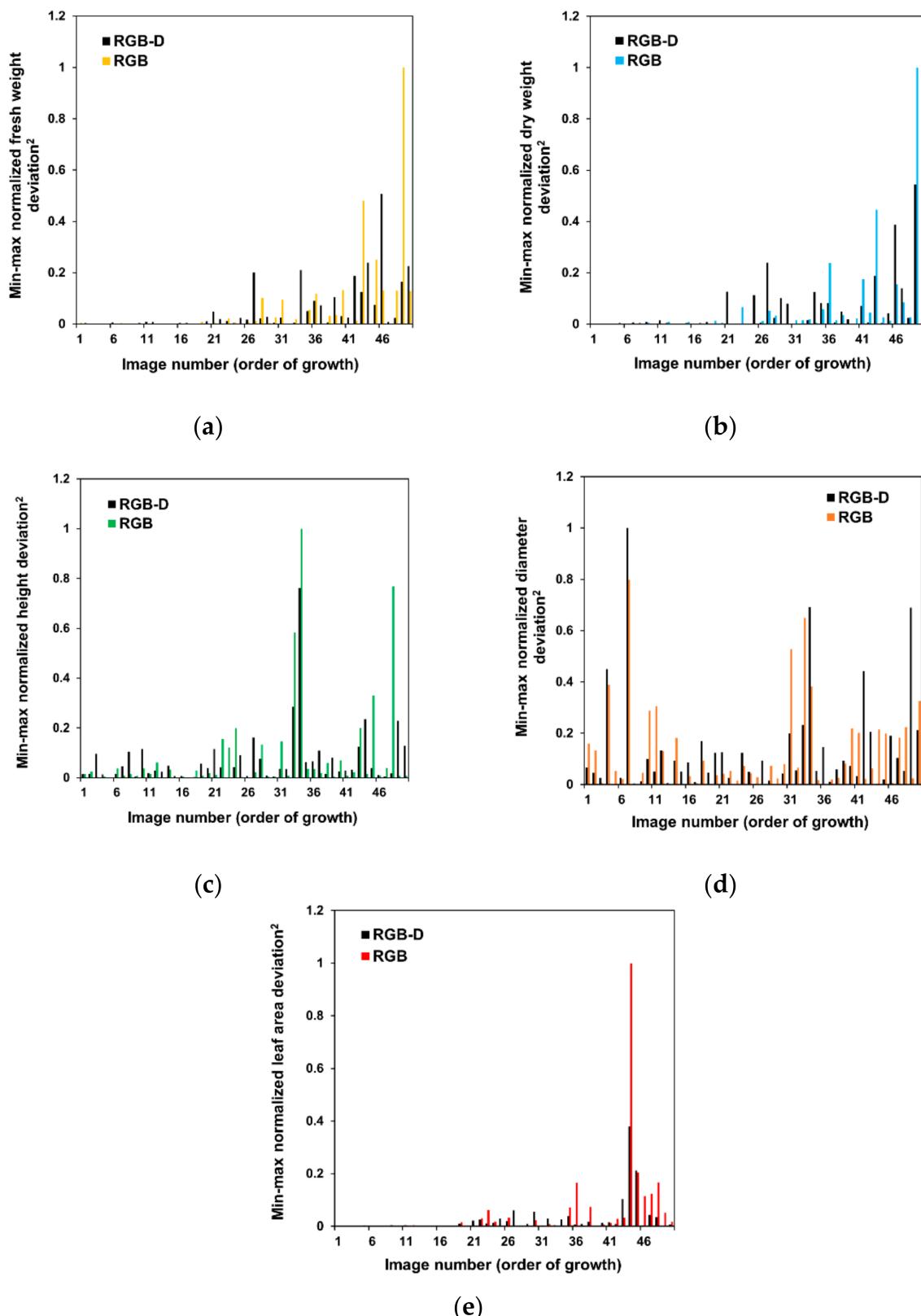
Similar to previous studies, the image interpretation capability of the CNN made it possible to implement the model with a good explanatory power [22,28]. The R<sup>2</sup> and NRMSE values for the leaf area were better than those of the other indices. Depending on the plant growth, the point at which each leaf is attached to the stem and the angle of each leaf may vary. The leaf positions and angles contain information in the vertical direction and can affect the nonprojected leaf area. Fresh and dry weights are also affected when plants grow vertically. The R<sup>2</sup> and NRMSE values of the fresh and dry weights support this. Therefore, an RGB-D image model may be effective for estimating growth indices.

In a previous study [28], Raja et al. proposed a CNN-based estimation model using the same dataset as that in this study [30]. The NMSE value of this study was 0.87, larger than the NMSE values of 0.068 and 0.069 of the MIMO and MISO models to which the DCNN was applied in a previous study. However, the NMSE for this model was better than the values of 0.092 and 0.088 for the MIMO and MISO models using the original CNN. Compared to the RMSE values of the MIMO model using the CNN by Raja et al. (fresh weight: 29.95, dry weight: 1.1, height: 1.75, diameter: 2.3, and leaf area: 344.96), the results (Table 3) were slightly improved except for dry weight. In contrast, the proposed model presented lower accuracy for the fresh weight, dry weight, and leaf area than the results of applying the DCNN to MIMO (fresh weight: 20.54, dry weight: 0.98, height: 1.75, diameter: 2.57, and leaf area: 274.23). The MISO model using the DCNN exhibited the best RMSEs for the growth indices except for fresh weight (fresh weight: 25.53, dry weight: 0.88, height: 1.64, diameter: 2.39, and leaf area: 272.58). The overall results of SIMO and SISO models using CNN and DCNN were similar to the proposed model results, with high and low indices.

However, our study focused on investigating the effectiveness of using RGB-D images and the CNN in the late growth stage rather than improving the accuracy. Securing the real-time applicability of the developed model was also a major concern. Therefore, given this real-time applicability, the multi-output model was developed and investigated, and the single-output model was not considered.

The developed RGB-D model also reduced errors in the late growth stage of lettuce, as displayed in Figure 8, which presents normalized deviations between the actual values and indices estimated from the newly developed model using RGB-D images and only RGB images in 20 samples of the late growth. Depth processing layers were skipped in the two-stage model to obtain the results using only RGB images. There was no significant difference between the performance of the developed model without the depth processing layers and the original ResNetV50 model using RGB images.

In Figure 8, the deviations were improved compared to the RGB-only model in the late growth, except for the diameter. These results indicated that using depth images in the CNN could advance the estimation performance for fresh weight, dry weight, height, diameter, and leaf area in the late growth stage. The increased accuracy for the late growth stage may lead to model applicability over the entire growth cycle. Figure 3 presents the quadratic increase in fresh weight and leaf area by vertical growth of crops. Due to this growth pattern, 3D information, such as depth images, may improve the growth estimation model.



**Figure 8.** Comparison of growth index estimation of the developed two-stage CNN model using RGB-D images and RGB images in 20 samples in the late growth stage: (a) fresh weight, (b) dry weight, (c) height, (d) diameter, and (e) leaf area.

In contrast, model accuracy using only RGB images was relatively higher in the case of diameter. The result may be because the diameter is not significantly affected by the vertical information of the crops, as the diameter was measured in the projected plant area.

In addition to the growth cycle, for the growth indices for each lettuce cultivar, the  $R^2$  values are similar to the estimated results for all cultivars (Table 4). Therefore, it is expected to be widely applicable to various varieties, such as *Lactuca sativa* L., domestic lettuce grown in South Korea.

**Table 4.** Comparison of  $R^2$  values of four lettuce varieties.

	Fresh Weight ( $R^2$ )	Dry Weight ( $R^2$ )	Height ( $R^2$ )	Diameter ( $R^2$ )	Leaf Area ( $R^2$ )
Aphylion	0.96	0.96	0.96	0.92	0.96
Lugano	0.93	0.91	0.91	0.86	0.96
Satine	0.96	0.99	0.95	0.92	0.98
Salanova	0.95	0.95	0.91	0.96	0.97
Total	0.95	0.95	0.95	0.89	0.96

Table 5 lists the average time and standard deviations required to load one image and infer growth indicators on a desktop and embedded board. The inference speed was still reasonable for the RGB-D images and the CNN model. The average loading and inference time per image was 0.83 s on the embedded board, which is a slightly longer time than that on the desktop. An on-the-go image acquisition device presented in a previous study [9] required approximately 2.83 s to capture one image per plant and was driven by a motor to move to the next plant. In an on-the-go image acquisition system, the time for image cropping may not be considered because the camera is located at the center of the plant. The processing time presented in this study corresponds to approximately 31% of the device running time; therefore, it is likely to be sufficiently applicable to real-time image acquisition and estimation via an on-the-go device through multiprocessing.

**Table 5.** Processing time required to load and infer lettuce RGB-D images using an i7-11700 CPU and GeForce RTX-3090 GPU versus a Jetson SUB mini-PC based on Jetson Xavier NX.

	Loading Time per Image (s)	Loading Time per 50 Images (s)	Inference Time per Image (s)	Inference Time per 50 Images (s)
i7-11700 and RTX-3090 PC	$0.03 \pm 0.001$	$0.52 \pm 0.015$	$0.13 \pm 0.044$	$0.55 \pm 0.079$
Jetson SUB mini-PC	$0.034 \pm 0.004$	$2.56 \pm 0.990$	$0.49 \pm 0.190$	$5.59 \pm 0.108$

#### 4. Conclusions

In this study, we developed and evaluated a two-stage CNN model using RGB and depth images to estimate the fresh weight, dry weight, height, diameter, and leaf area of lettuce grown in a greenhouse hydroponics system. The RGB-D images and CNN-based architecture effectively estimate the growth indices of greenhouse lettuces, demonstrating that the CNN-based model using RGB-D images may provide nondestructive, fast, and accurate measurements of growth indices.

The proposed RGB-D model can reduce estimation errors in the late growth phase when the growth indices are greatly increased. This advantage can improve model applicability for estimating growth indices. The model can potentially effectively model the interaction between crop physiology and the environment.

Since the late stage of crop growth is closely related to the optimal harvest time, the proposed model may also help predict the proper harvest time of crops and develop

decision support systems or harvesting robots that operate automatically. The rapid inference speed of the CNN on individual images also enables sensing and harvesting tasks requiring real-time motion and image processing.

The lettuce images used in this study include a limited number of varieties and a small volume of data. In addition, data were collected manually. In the future, we will apply the developed model to an on-the-go embedded system in a greenhouse and collect and learn from a large volume of data from lettuce varieties grown in South Korea to perform real-time growth estimation and improve the accuracy and measurement convenience. Another goal for improvement concerns the automatic measurement of plant center points.

Although this study was conducted only on lettuce, the CNN model is more versatile than traditional image processing methods and can be applied to various environments and crops after slight tuning, such as through transfer learning. The proposed estimation model using RGB-D data can be applied to fruit and vegetable (in addition to leafy vegetable) crops.

The rapid and precise measurement of growth indicators using RGB-D images and CNN models has great potential for agricultural applications. Developing such estimation models will advance precision agriculture and ultimately improve the productivity of horticultural crops.

**Author Contributions:** Conceptualization, H.-J.K. and M.-S.G.; methodology, M.-S.G., H.-J.K. and D.-W.K.; software, M.-S.G.; validation, M.-S.G. and H.-J.K.; formal analysis, M.-S.G., H.-J.K. and D.-W.K.; investigation, M.-S.G.; data curation, M.-S.G. and D.-W.K.; writing—original draft preparation, M.-S.G.; and writing—review and editing, H.-J.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Brain Korea 21 Fostering Outstanding Universities for Research, funded by the Ministry of Education of Korea and the National Research Foundation of Korea. This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry and Korea Smart Farm R&D Foundation through the Smart Farm Innovation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs and the Ministry of Science and ICT, Rural Development Administration (RDA), Republic of Korea (21006031HD030). This study was supported by the “Research Program for Agricultural Science & Technology Development (Project No. PJ014911)”, RDA, Republic of Korea.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This study used the Third Autonomous Greenhouse Challenge: Online Challenge Lettuce Images dataset publicly available at 4TU.ResearchData [30]. The ImageNet dataset was also used, which is available on the ImageNet website [42]. More details about the data are available in Sections 2.1–2.3.

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

## References

1. Smart, R.E. Principles of grapevine canopy microclimate manipulation with implications for yield and quality: A review. *Am. J. Enol. Vitic.* **1985**, *36*, 230–239.
2. Williams, L.E. Growth of ‘Thompson Seedless’ grapevines. I. Leaf area development and dry weight distribution. *J. Am. Soc. Hortic. Sci.* **1987**, *112*, 325–330.
3. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. *Crop Evapotranspiration-Guidelines for Computing Crop Water Requirements-FAO Irrigation and Drainage Paper 56*; Food and Agriculture Organization of the United Nations: Rome, Italy, 1998.
4. Pearcy, R.W.; Ehleringer, J.R.; Mooney, H.; Rundel, P.W. *Plant Physiological Ecology: Field Methods and Instrumentation*; Springer Science & Business Media: Berlin, Germany, 2021.
5. Cho, Y.Y.; Oh, S.; Oh, M.M.; Son, J.E. Estimation of individual leaf area, fresh weight, and dry weight of hydroponically grown cucumbers (*Cucumis sativus* L.) using leaf length, width, and SPAD value. *Sci. Hortic.* **2007**, *111*, 330–334. [[CrossRef](#)]
6. Peksen, E. Non-destructive leaf area estimation model for faba bean (*Vicia faba* L.). *Sci. Hortic.* **2007**, *113*, 322–328. [[CrossRef](#)]
7. Lati, R.N.; Filin, S.; Eizenberg, H. Plant growth parameter estimation from sparse 3D reconstruction based on highly-textured feature points. *Precis. Agric.* **2013**, *14*, 586–605. [[CrossRef](#)]

8. Yeh, Y.-H.F.; Lai, T.-C.; Liu, T.-Y.; Liu, C.-C.; Chung, W.-C.; Lin, T.-T. An automated growth measurement system for leafy vegetables. *Biosyst. Eng.* **2014**, *117*, 43–50. [[CrossRef](#)]
9. Jiang, J.S.; Kim, H.J.; Cho, W.J. On-the-go image processing system for spatial mapping of lettuce fresh weight in plant factory. *IFAC-PapersOnLine* **2018**, *51*, 130–134. [[CrossRef](#)]
10. Jin, X.; Zarco-Tejada, P.J.; Schmidhalter, U.; Reynolds, M.P.; Hawkesford, M.J.; Varshney, R.K.; Yang, T.; Nie, C.; Li, Z.; Ming, B.; et al. High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *IEEE Trans. Geosci. Remote Sens.* **2021**, *9*, 200–231. [[CrossRef](#)]
11. Sze, V.; Chen, Y.H.; Yang, T.J.; Emer, J.S. Efficient processing of deep neural networks: A tutorial and survey. *Proc. Inst. Electr. Electron. Eng.* **2017**, *105*, 2295–2329. [[CrossRef](#)]
12. Dyrmann, M.; Karstoft, H.; Midtiby, H.S. Plant species classification using deep convolutional neural network. *Biosyst. Eng.* **2016**, *151*, 72–80. [[CrossRef](#)]
13. Hasan, A.M.; Sohel, F.; Diepeveen, D.; Laga, H.; Jones, M.G. A survey of deep learning techniques for weed detection from images. *Comput. Electron. Agric.* **2021**, *184*, 106067. [[CrossRef](#)]
14. Jiang, Y.; Li, C. Convolutional neural networks for image-based high-throughput plant phenotyping: A review. *Plant Phenomics* **2020**, *2020*, 4152816. [[CrossRef](#)]
15. Fu, L.; Gao, F.; Wu, J.; Li, R.; Karkee, M.; Zhang, Q. Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review. *Comput. Electron. Agric.* **2020**, *177*, 105687. [[CrossRef](#)]
16. Liu, J.; Wang, X. Plant diseases and pests detection based on deep learning: A review. *Plant Methods* **2021**, *17*, 22. [[CrossRef](#)]
17. Vasanthi, V. Crop growth monitoring and leaf area index estimation using wireless sensor network and CNN. In Proceedings of the 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2–4 September 2021. [[CrossRef](#)]
18. Jin, S.; Su, Y.; Song, S.; Xu, K.; Hu, T.; Yang, Q.; Guo, Q. Non-destructive estimation of field maize biomass using terrestrial lidar: An evaluation from plot level to individual leaf level. *Plant Methods* **2020**, *16*, 69. [[CrossRef](#)]
19. Liu, W.; Li, Y.; Liu, J.; Jiang, J. Estimation of plant height and aboveground biomass of *Toona sinensis* under drought stress using RGB-D imaging. *Forests* **2021**, *12*, 1747. [[CrossRef](#)]
20. Lu, J.Y.; Chang, C.L.; Kuo, Y.F. Monitoring growth rate of lettuce using deep convolutional neural networks. In Proceedings of the 2019 ASABE Annual International Meeting, Boston, MA, USA, 7–10 July 2019. [[CrossRef](#)]
21. Reyes-Yanes, A.; Martinez, P.; Ahmad, R. Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds. *Comput. Electron. Agric.* **2020**, *179*, 105827. [[CrossRef](#)]
22. Zhang, L.; Xu, Z.; Xu, D.; Ma, J.; Chen, Y.; Fu, Z. Growth monitoring of greenhouse lettuce based on a convolutional neural network. *Hortic. Res.* **2020**, *7*, 124. [[CrossRef](#)]
23. Mortensen, A.K.; Bender, A.; Whelan, B.; Barbour, M.M.; Sukkarieh, S.; Karstoft, H.; Gislum, R. Segmentation of lettuce in coloured 3D point clouds for fresh weight estimation. *Comput. Electron. Agric.* **2018**, *154*, 373–381. [[CrossRef](#)]
24. Jiang, Y.; Li, C.; Paterson, A.H.; Sun, S.; Xu, R.; Robertson, J. Quantitative analysis of cotton canopy size in field conditions using a consumer-grade RGB-D camera. *Front. Plant Sci.* **2018**, *8*, 2233. [[CrossRef](#)]
25. Yan, G.; Hu, R.; Luo, J.; Weiss, M.; Jiang, H.; Mu, X.; Zhang, W. Review of indirect optical measurements of leaf area index: Recent advances, challenges, and perspectives. *Agric. For. Meteorol.* **2019**, *265*, 390–411. [[CrossRef](#)]
26. Kolhar, S.; Jagtap, J. Plant trait estimation and classification studies in plant phenotyping using machine vision—A review. *Inf. Process. Agric.* **2021**. [[CrossRef](#)]
27. Quan, L.; Li, H.; Li, H.; Jiang, W.; Lou, Z.; Chen, L. Two-stream dense feature fusion network based on RGB-D data for the real-time prediction of weed aboveground fresh weight in a field environment. *Remote Sens.* **2021**, *13*, 2288. [[CrossRef](#)]
28. Raja, P.V.; Olenskyj, A.; Kamangir, H.; Earles, M. Simultaneously predicting multiple plant traits from multiple sensors using deformable CNN regression. In Proceedings of the 2022 AI for Agriculture and Food Systems (AIAFS), Vancouver, BC, Canada, 28 February–1 March 2022. [[CrossRef](#)]
29. Li, J.; Wang, M. An end-to-end deep RNN based network structure to precisely regress the height of lettuce by single perspective sparse point cloud. In Proceedings of the 2022 North American Plant Phenotyping Network (NAPPN), Athens, GA, USA, 22–25 February 2022. [[CrossRef](#)]
30. Hemming, S.; de Zwart, H.F.; Elings, A.; Bijlaard, M.; van Marrewijk, B.; Petropoulou, A. 3rd Autonomous Greenhouse Challenge: Online Challenge Lettuce Images. Available online: <https://doi.org/10.4121/15023088.v1> (accessed on 2 March 2022).
31. Chang, A.; Jung, J.; Maeda, M.M.; Landivar, J. Crop height monitoring with digital imagery from unmanned aerial system (UAS). *Comput. Electron. Agric.* **2017**, *141*, 232–237. [[CrossRef](#)]
32. Olaniyi, J.O.; Akanbi, W.B.; Adejumo, T.A.; Ak, O.G. Growth, fruit yield and nutritional quality of tomato varieties. *Afr. J. Food Sci.* **2010**, *4*, 398–402.
33. Evans, G.C. *The Quantitative Analysis of Plant Growth*; University of California Press: Berkeley, CA, USA, 1972; Volume 1.
34. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Hartling, S.; Esposito, F.; Fritsch, F.B. Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sens. Environ.* **2020**, *237*, 111599. [[CrossRef](#)]
35. Ahmed, A.K.; Cresswell, G.C.; Haigh, A.M. Comparison of sub-irrigation and overhead irrigation of tomato and lettuce seedlings. *J. Hortic. Sci. Biotechnol.* **2000**, *75*, 350–354. [[CrossRef](#)]

36. Kaselimi, M.; Protopapadakis, E.; Voulodimos, A.; Doulamis, N.; Doulamis, A. Multi-channel recurrent convolutional neural networks for energy disaggregation. *IEEE Access* **2019**, *7*, 81047–81056. [[CrossRef](#)]
37. Bacchi, S.; Zerner, T.; Oakden-Rayner, L.; Kleinig, T.; Patel, S.; Jannes, J. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: A pilot study. *Acad. Radiol.* **2020**, *27*, e19–e23. [[CrossRef](#)] [[PubMed](#)]
38. Zang, H.; Liu, L.; Sun, L.; Cheng, L.; Wei, Z.; Sun, G. Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. *Renew. Energ.* **2020**, *160*, 26–41. [[CrossRef](#)]
39. Yang, H.; Wang, L.; Huang, C.; Luo, X. 3D-CNN-Based sky image feature extraction for short-term global horizontal irradiance forecasting. *Water* **2021**, *13*, 1773. [[CrossRef](#)]
40. Fix, E.; Hodges, J.L. Discriminatory analysis. Nonparametric discrimination: Consistency properties. *Int. Stat. Rev.* **1989**, *57*, 238–247. [[CrossRef](#)]
41. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
42. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009. [[CrossRef](#)]
43. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016. [[CrossRef](#)]
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016. [[CrossRef](#)]
45. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2015**, arXiv:1412.6980. Available online: <https://arxiv.org/abs/1412.6980> (accessed on 2 March 2022).
46. Nair, V.; Hinton, G.E. Rectified linear units improve restricted Boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.
47. Mosteller, F.; Tukey, J.W. *Data Analysis, Including Statistics*; Addison-Wesley: Boston, MA, USA, 1968; pp. 80–203.
48. Golzarian, M.R.; Frick, R.A.; Rajendran, K.; Berger, B.; Roy, S.; Tester, M.; Lun, D.S. Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* **2011**, *7*, 2. [[CrossRef](#)]
49. Feng, H.; Jiang, N.; Huang, C.; Fang, W.; Yang, W.; Chen, G.; Liu, Q. A hyperspectral imaging system for an accurate prediction of the above-ground biomass of individual rice plants. *Rev. Sci. Instrum.* **2013**, *84*, 095107. [[CrossRef](#)]
50. Rice, L.; Wong, E.; Kolter, Z. Overfitting in adversarially robust deep learning. In Proceedings of the 37th International Conference on Machine Learning, Vienna, Austria, 13–18 July 2020.
51. Ying, X. An overview of overfitting and its solutions. In Proceedings of the 2018 International Conference on Computer Information Science and Application Technology, Daqing, China, 7–9 December 2018.