Journal of *Imaging*

MDPI

Article

PlantSR: Super-Resolution Improves Object Detection in Plant Images

Tianyou Jiang [1], Qun Yu [1,2,*], Yang Zhong [1] and Mingshun Shao [1]

[1] College of Information Science and Engineering, Shandong Agricultural University, Tai'an 271018, China; s1729041183@gmail.com (T.J.); inawjsyg@gmail.com (Y.Z.); sms1360169864@gmail.com (M.S.)
[2] Huanghuaihai Key Laboratory of Smart Agricultural Technology, Ministry of Agriculture and Rural Affairs, Tai'an 271018, China
* Correspondence: yuqun@sdau.edu.cn

Abstract: Recent advancements in computer vision, especially deep learning models, have shown considerable promise in tasks related to plant image object detection. However, the efficiency of these deep learning models heavily relies on input image quality, with low-resolution images significantly hindering model performance. Therefore, reconstructing high-quality images through specific techniques will help extract features from plant images, thus improving model performance. In this study, we explored the value of super-resolution technology for improving object detection model performance on plant images. Firstly, we built a comprehensive dataset comprising 1030 high-resolution plant images, named the PlantSR dataset. Subsequently, we developed a super-resolution model using the PlantSR dataset and benchmarked it against several state-of-the-art models designed for general image super-resolution tasks. Our proposed model demonstrated superior performance on the PlantSR dataset, indicating its efficacy in enhancing the super-resolution of plant images. Furthermore, we explored the effect of super-resolution on two specific object detection tasks: apple counting and soybean seed counting. By incorporating super-resolution as a pre-processing step, we observed a significant reduction in mean absolute error. Specifically, with the YOLOv7 model employed for apple counting, the mean absolute error decreased from 13.085 to 5.71. Similarly, with the P2PNet-Soy model utilized for soybean seed counting, the mean absolute error decreased from 19.159 to 15.085. These findings underscore the substantial potential of super-resolution technology in improving the performance of object detection models for accurately detecting and counting specific plants from images. The source codes and associated datasets related to this study are available at Github.

Keywords: super resolution; object detection; image reconstruction; image processing; deep learning; plant phenotypes

check for updates

Citation: Jiang, T.; Yu, Q.; Zhong, Y.; Shao, M. PlantSR: Super-Resolution Improves Object Detection in Plant Images. *J. Imaging* **2024**, *10*, 137. https://doi.org/10.3390/jimaging10060137

Academic Editor: Antonio Fernández-Caballero

Received: 12 May 2024
Revised: 1 June 2024
Accepted: 4 June 2024
Published: 6 June 2024

Copyright: © 2024 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

Plant images play an important role in understanding the intricate structures and characteristics of various plant species [1,2] and are often used for automatic analysis and to extract useful information. However, images of plants often suffer from low resolution and blurriness. Even in cases where the overall image possesses high resolution, certain objects within the image may exhibit low-resolution characteristics. These limitations in image resolution and quality impede the accurate extraction of features from images, consequently restricting the performance of deep learning models created for plant image object detection tasks.

Recent advancements in computer vision, particularly deep learning models, have garnered significant attention and shown promising results in plant image object detection tasks, such as soybean seed counting [3,4] and wheat head detection [5–7]. Despite these successes, the effectiveness of these deep learning models heavily depends on the quality

of input images [8]. Alterations in image resolution notably affect model accuracy, with low-resolution images impairing model performance during both training and inference phases [9–11]. Specifically, low-resolution images offer less informative content during the training phase compared to high-resolution counterparts, potentially undermining model effectiveness. Moreover, during the inference phase, images with lower resolution than training data due to capturing targets at longer distances or camera resolution limitations can also diminish model performance.

Super resolution (SR) is an image processing technique aimed at enhancing spatial resolution by reconstructing high-resolution details from low-resolution images. Deep learning-based methods have emerged as the most effective solution in the field of super-resolution [12–15]. By reconstructing high-resolution details of plant images, super-resolution technology holds the potential to empower deep learning models to effectively learn and extract intricate features from images during the training phase and improve model performance when predicting with low-resolution images during the inference phase.

Several studies have successfully applied SR techniques to plant images, enhancing the performance of deep learning models in specific tasks. For instance, Yamamoto K et al. [16] deployed the super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture. Maqsood M H et al. [17] applied super-resolution generative adversarial networks [18] for upsampling images before using them to train deep learning models for the detection of wheat yellow rust. Cap Q H et al. [19] proposed an effective super-resolution method called LASSR for plant disease diagnosis. Albert P et al. [20] transferred knowledge learned on ground-level images to raw drone images and estimated dry herbage biomass by applying super-resolution technology to raw drone images. These successes demonstrate the significant application value of super resolution in plant image-related tasks.

While numerous datasets and models already exist in the classical field of SR [21–24], specialized SR datasets and models tailored to plant image reconstruction remain absent. In this study, we built a dedicated SR dataset solely comprising plant images (PlantSR dataset), which served as a fundamental dataset for comparing the performance of super-resolution models on plant images. Leveraging the PlantSR dataset, we compared the performance of five established deep learning models (SRCNN, VDSR, EDSR, RCAN, and SwinIR) in the classical field of SR and introduced a novel SR architecture specifically tailored to plant images (PlantSR Model). Our experimental results highlight that our model achieved the highest peak signal-to-noise ratio (PSNR) across all scaling factors (2, 3, and 4) on the PlantSR dataset while maintaining a reasonable model size and inference speed.

Next, we explored the improvement effects of our super-resolution model on apple counting and soybean seed counting tasks. In the apple counting task, applying PlantSR to test images significantly reduced the mean absolute error from 13.085 to 5.71. Similarly, for soybean seed counting, using PlantSR for both training and test images lowered the mean absolute error from 19.159 to 15.085. These findings underscore the potential of super-resolution technology to improve the performance of deep learning models for plant image-related tasks.

2. Materials and Methods

2.1. PlantSR Dataset

The PlantSR dataset (Figure 1) comprises a total of 1030 plant images, categorized into four primary botanical groups: 25 images of Bryophytes [25], 115 images of Ferns [26], 440 images of Gymnosperms [27], and 450 images of Angiosperms [28]. The uneven distribution of plant images in the dataset reflects the biological reality of plant species diversity in nature. Being consistent with biological taxonomy, Angiosperms and Gymnosperms are typically more abundant in natural ecosystems compared to Ferns and Bryophytes. Therefore, the intentional allocation of a larger number of images to Angiosperms and Gymnosperms in the dataset is justified by their biological prevalence and economic and

ecological importance. All the images were meticulously selected to ensure high resolution and sourced from Plants of the World Online [29]. These images were meticulously divided, with 830 images designated for the training dataset and 200 images for the test dataset. This dataset could be used as a fundamental dataset to compare the performance of super-resolution models on plant images in future studies. Sample images from four categories in the dataset are shown in Figure 2.

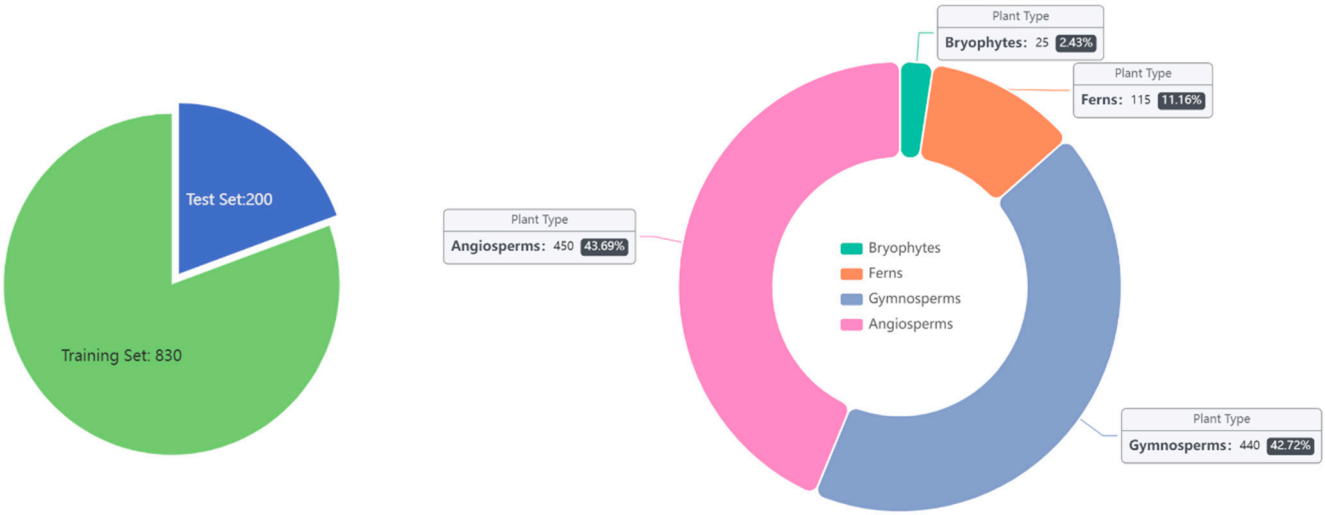


Figure 1. PlantSR dataset. The PlantSR dataset contains 25 images of Bryophytes, 115 images of Ferns, 440 images of Gymnosperms, and 450 images of Angiosperms. These data were divided into 830 images as a training dataset and 200 images as a test dataset.

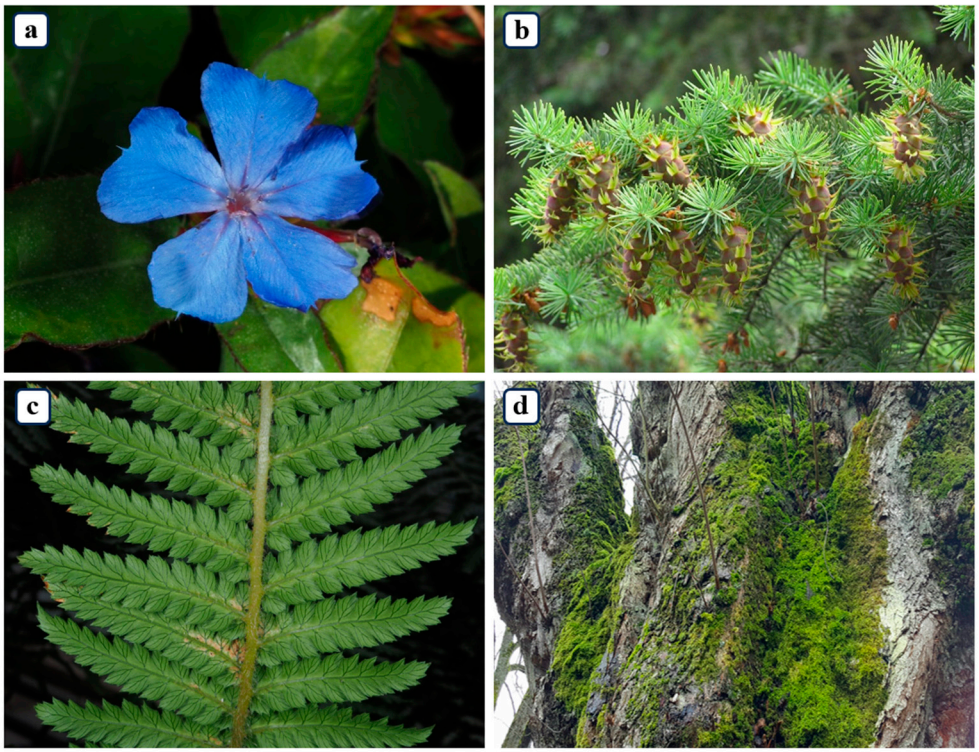


Figure 2. Sample images of four categories in the PlantSR dataset. (**a**) Sample image of Angiosperms. (**b**) Sample image of Gymnosperms. (**c**) Sample image of Ferns. (**d**) Sample image of Bryophytes.

2.2. Architecture of PlantSR Model

As shown in Figure 3, our PlantSR model consisted of three modules: shallow feature extraction, deep feature extraction, and high-resolution (HR) image reconstruction.

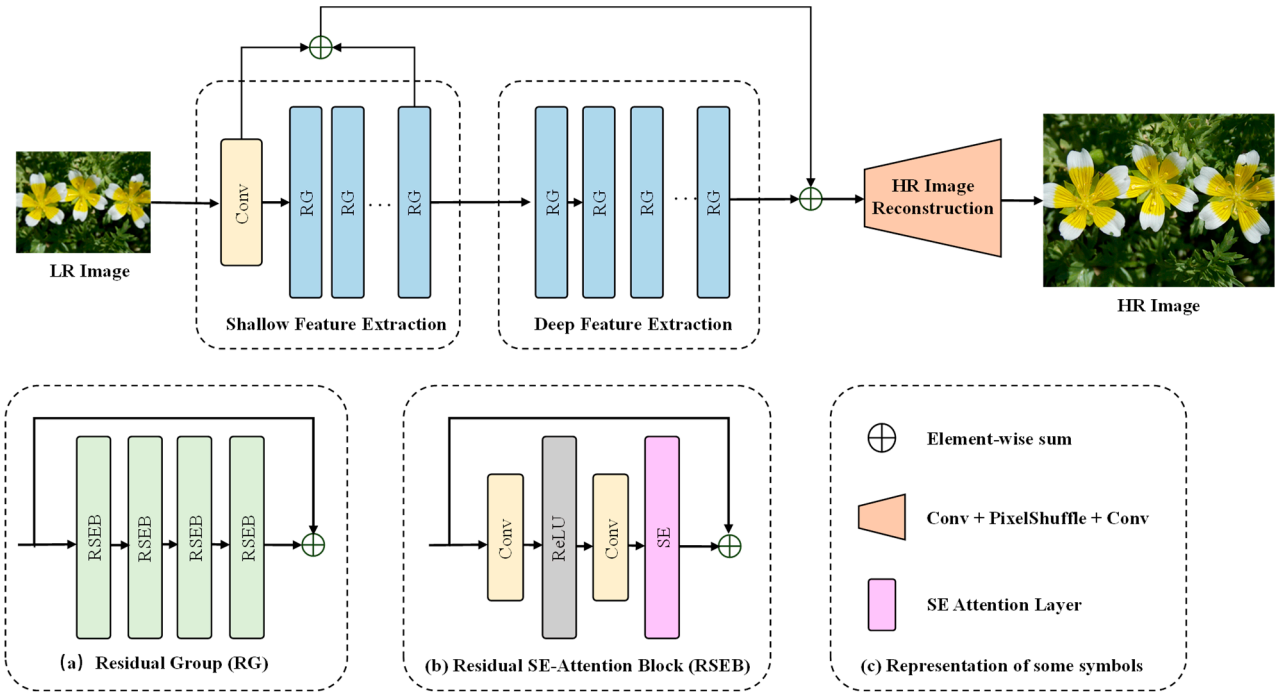


Figure 3. Architecture of PlantSR model.

Given a low-resolution (LR) image I_{LR} as input, we first used a 3×3 convolutional layer to extract feature F_0

$$F_0 = H_{SF_0}(I_{LR}), \tag{1}$$

where H_{SF_0} denotes a 3×3 convolutional layer, which was defined as the first component of the shallow feature extraction module. Then, we used the second component of the shallow feature extraction module to extract shallow feature F_1 from F_0, and used the deep feature extraction module to extract deep feature F_{DF} from F_1

$$F_1 = H_{SF_1}(F_0), \tag{2}$$

$$F_{DF} = H_{DF}(F_1), \tag{3}$$

where H_{SF_1} denotes the second component of the shallow feature extraction module and $H_{DF}(\cdot)$ denotes the deep feature extraction module; these two modules consisted of m and n residual groups (RGs), respectively. Each RG was composed of four Residual SE-attention Blocks (RSEBs) and a residual connection from input to output features. Within each RSEB, an SE-attention layer [30] was applied sequentially after a convolutional layer, a rectified linear unit (ReLU) layer, and another convolutional layer. Once feature maps were obtained from the deep feature extraction module, the HR image reconstruction module was performed to obtain the SR image:

$$I_{SR} = H_{UP}(F_0 + F_1 + F_{DF}), \tag{4}$$

where $H_{UP}(\cdot)$ denotes the HR image reconstruction module, which was initiated by enhancing the feature map channels through a convolutional layer. Following this, the module scaled up the feature maps using a PixelShuffle layer and, in the concluding step, reconstructed the feature maps to the RGB channel through another convolutional layer.

Throughout the training phase, we optimized the parameters by minimizing the L1 pixel loss:

$$L = ||I_{SR} - I_{HR}||_1, \quad (5)$$

where I_{SR} is obtained by taking LR image I_{LR} as the input of PlantSR and I_{HR} is the corresponding ground-truth HR image. We opted for the L1 loss, as it has demonstrated superior performance in image restoration tasks [31]. The loss function was optimized using the Adam optimization algorithm.

2.3. Super-Resolution Effects on Apple Counting Task

As YOLOv7 is one of the most popular models in the field of object detection, we selected it to analyze our super-resolution model effects on counting apples [32]. For training and evaluating the apple counting model, we sourced 1000 images from the internet to form the training dataset. Additionally, we utilized 200 images of harvest-ready apples provided by the 2023 Asia and Pacific Mathematical Contest in Modeling competition as our test dataset. Each image in the test dataset was relatively small, all measuring 270 × 180 pixels.

After training the YOLOv7 model using 1000 images in the training dataset, we conducted tests with and without applying super-resolution to the test images. Super-resolution was achieved using the PlantSR (×3) model. To ensure optimal performance in processing apple images, we fine-tuned a PlantSR model using a curated collection of 100 high-resolution close-up apple images gathered from the internet, leveraging these images to train a model based on the original pretrained model.

2.4. Super-Resolution Effects on Soybean Seed Counting Task

As an improvement to the P2PNet [33], Zhao J et al. proposed an automated soybean seed counting tool called P2PNet-Soy [34]. Thanks to their efforts, the accuracy of soybean seed counting in the field witnessed a significant improvement. P2PNet-Soy incorporates an unsupervised clustering algorithm, the k-d tree [35], as post-processing to find the centers of closely located predictions, thereby enhancing the final prediction accuracy. Additionally, P2PNet-Soy made some adjustments to the architecture of P2PNet to maximize model performance in seed counting and localization. We downloaded their open dataset, which consists of 126 images of soybeans captured from one side as the training dataset and 132 images of soybeans taken from the opposite side as the evaluation dataset [36].

Expanding on the open dataset, we incorporated 60 soybean seed images from SDAU Technology Park and reorganized the dataset as follows: 126 images from the P2PNet-Soy training dataset plus 30 additional images as the training dataset, 50 images from the P2PNet-Soy evaluation dataset plus 10 additional images as the validation dataset, and 82 images from the P2PNet-Soy evaluation dataset as the test dataset.

To explore the effects of super-resolution on the task of soybean seed counting, we trained and evaluated P2PNet-Soy with images in four different situations: (1) Original images across training, validation, and test sets. All images used for training, validation, and testing purposes remained in their original resolution. (2) Original training and validation images with downsampling of test images. The training and validation images remained in their original resolution, while the test images underwent downsampling by a factor of two using bilinear interpolation. (3) Original training and validation images with downsampling and PlantSR (×2) upscaling of test images. Similar to the previous setup, the training and validation images remained in their original resolutions. However, for the test set, images underwent downsampling followed by upscaling using a PlantSR (×2) model. (4) PlantSR (×2) upscaled images across training, validation, and test sets. All images used for training, validation, and testing underwent upscaling by a PlantSR (×2) model.

In the fourth case, we proposed the application of the PlantSR model as a pre-processing step to P2PNet-Soy (Figure 4). Here is a detailed description of this setup: Before employing images for training, validation, and testing in P2PNet-Soy, all images un-

derwent upscaling by a PlantSR (×2) model. To train a specific PlantSR model on soybean seed images, we collected 89 high-resolution close-up images of soybeans from laboratory settings. These images, possessing a higher pixel density than those in the original dataset, were used for transfer learning to train the specific PlantSR (×2) model on soybean seed images. During the training phase, P2PNet-Soy randomly cropped 224 × 224 patches from the original images. Considering the upscale by a factor of two, the resolution of these patches was adjusted to 448 × 448. Moreover, P2PNet-Soy used a k-d tree to filter the prediction results. In the implementation of the k-d tree, the parameter cutoff was set to find all points within the distance "r" from each point. Since the pixel density of the images was tripled, the parameter cutoff was also doubled in our experiment.

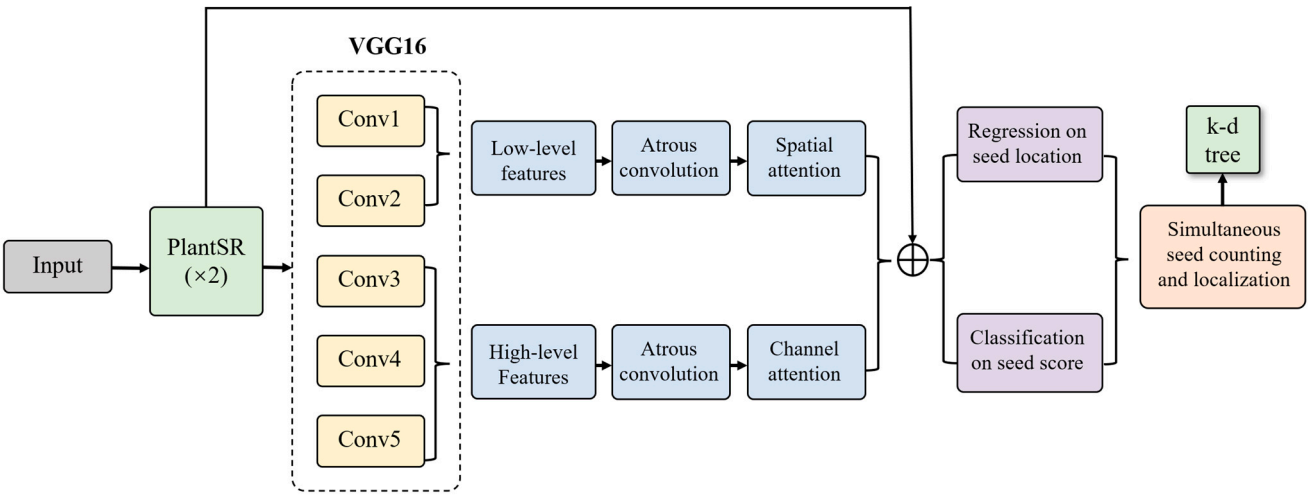


Figure 4. Architecture of P2PNet-Soy after applying the PlantSR model as a pre-processing step.

2.5. Training and Evaluation Settings

To assess the performance of various super-resolution (SR) models, including our PlantSR model, consistent training and evaluation settings were applied. All models were trained and evaluated on the PlantSR dataset and shared identical configurations for training. In each training batch, 32 patches sized 64 × 64 (for scale = 2 and scale = 4) or 63 × 63 (for scale = 3) were extracted from training images, followed by horizontal data augmentation. A total of 38,997 batches constituted one epoch. Each model was trained for multiple epochs, and their evaluation metrics on the test dataset were recorded. The model that performed best in multiple epochs was saved for final comparison. We used Adam [37] as the optimizer and the initial learning rate was set to 10^{-4}. All the SR models were trained on an NVIDIA RTX A5000 24G GPU (NVIDIA, Santa Clara, California, USA).

To explore the effects of the PlantSR model on the task of apple counting, we trained and evaluated YOLOv7 with test images with or without applying super-resolution to the test images. The input size of YOLOv7 was set to 640. A series of data augmentations (mosaic, mixup, random crop, random angle) were applied throughout the training phase. The training of YOLOv7 was conducted on an NVIDIA RTX A5000 24G GPU.

To explore the effects of the PlantSR model on the task of soybean seed counting, we trained and evaluated P2PNet-Soy with images in four different cases. In all four cases, a series of data augmentations (random scale, random crop, random flipping, random change brightness) were applied throughout the training phase, mirroring the approach used for the original P2PNet-Soy. We used Adam as the optimizer and the initial leaning rate was set to 10^{-4} and subsequently halved every 100 epochs. The training of P2PNet-Soy was conducted on an NVIDIA A100 40G Tensor Core GPU (NVIDIA, Santa Clara, CA, USA).

2.6. Evaluation Metrics

To compare the performance of several SR methods, we used two common metrics of image quality, including peak-to-peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [38]. The higher the PSNR and SSIM, the higher the similarity between predicted images and high-resolution images, and the better the model performance. The computation of these metrics was as follows:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right), \tag{6}$$

where MAX represents the maximum possible pixel value in the image, which was 255 here. MSE denotes the mean squared error, indicating the average of squared differences between corresponding pixels in the original and predicted images:

$$SSIM = \frac{(2\mu_x\,\mu_y\, + C_1\,)(2\sigma_{xy}\, + C_2\,)}{(\mu_x{}^2\, + \mu_y{}^2\, + C_1\,)(\sigma_x{}^2\, + \sigma_y{}^2\, + C_2\,)}, \tag{7}$$

where μ_x and μ_y are the means of images x and y, respectively. $\sigma_x{}^2$ and $\sigma_y{}^2$ are the variances of images x and y, respectively. σ_{xy} represents the covariance between x and y. C_1 and C_2 equal $(0.01L)^2$ and $(0.03L)^2$, where L is the dynamic range of pixel values, which was 255 here.

We used mean absolute error (MAE) and root mean square error (RMSE) to evaluate the performance of the soybean seed counting model and apple counting model. The closer the MAE and $RMSE$ were to 0, the smaller the error in counting:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{9}$$

where n represents the number of images used for testing, y_i represents the ground truth number of the target object within each image, and $\hat{y}_i$ represents the number of target objects predicted by the model.

3. Results

Our experimental results are presented in three sections: (1) SR model compression. Based on the PlantSR dataset, we conduct benchmarking of our PlantSR model against five well-established super-resolution (SR) models in the classic field of SR, alongside a traditional method, bicubic interpolation. (2) Super-resolution effects on the apple counting task. We trained and evaluated the YOLOv7 model with test images with or without applying the PlantSR model to achieve super resolution. (3) Super-resolution effects on the soybean seed counting task. We delved into the impact of our PlantSR model on the soybean seed counting task. Here, we meticulously trained and evaluated the P2PNet-Soy model with images in four different cases.

3.1. SR Model Compression

We rigorously benchmarked our proposed PlantSR network against a traditional method, bicubic interpolation, and five state-of-the-art super-resolution (SR) deep learning models: SRCNN [39], VDSR [40], EDSR [41], RCAN [42], and SwinIR [43]. Among the five SR deep learning models, SRCNN and VDSR employed interpolation to the low-resolution (LR) image as a preliminary step, subsequently learning to reconstruct the high-resolution (HR) image from these interpolated images. This approach is commonly referred to as Pre-upsampling SR. Conversely, EDSR, RCAN, and SwinIR first extracted features from LR images and then applied the upsampling module at the end of the network. This

methodology is commonly referred to as Post-upsampling SR. Notably, our PlantSR model also belonged to the Post-upsampling SR category.

Table 1 presents the quantitative results, where the number of parameters is expressed in millions, frames per second (FPS) are denoted in frames, and the input size is specified in the number of pixels. During the training process, patches of sizes 64 × 64 (for scale = 2 and scale = 4) or 63 × 63 (for scale = 3) were utilized as high-resolution (HR) images and were subsequently downsampled by a factor of scale as low-resolution (LR) images. The models learned the transformation process from LR to HR. It is noteworthy that the reported FPS results, although averaged across multiple experimental outcomes, should be considered as reference points due to inherent variations.

Table 1. Compression results of SR methods.

Method	Scale	PSNR	SSIM	Params (M)	FPS	Input Size (Pixels)
Bicubic	×2	38.23	0.9617	/	/	64
SRCNN		39.42	0.9667	0.069	507.0	
VDSR		40.31	0.9716	0.667	231.3	
EDSR		40.26	0.9706	1.370	123.7	
RCAN		40.15	0.9708	15.445	6.6	
SwinIR		40.34	0.9715	16.619	0.8	
PlantSR (Ours)		40.36	0.9716	1.397	37.1	
Bicubic	×3	33.56	0.9128	/	/	63
SRCNN		34.24	0.9190	0.069	475.4	
VDSR		34.79	0.9255	0.667	235.3	
EDSR		35.20	0.9273	1.554	71.9	
RCAN		35.24	0.9290	15.629	6.6	
SwinIR		35.23	0.9290	16.803	0.8	
PlantSR (Ours)		35.26	0.9291	5.760	18.1	
Bicubic	×4	32.13	0.8844	/	/	64
SRCNN		32.64	0.8927	0.069	473.2	
VDSR		33.53	0.9031	0.667	236.6	
EDSR		33.76	0.9063	1.518	57.5	
RCAN		33.78	0.9071	15.888	6.5	
SwinIR		33.83	0.9075	16.766	0.7	
PlantSR (Ours)		33.85	0.9077	13.531	9.8	

The results demonstrate that our PlantSR model achieved exceptional SR performance on plant images while maintaining a reasonable model size and inference speed. Among the compared models, it was observed that SwinIR, which uses Swin Transformer [44] as its backbone network, exhibited satisfactory performance. However, it exhibited comparatively slower inference speeds and higher computational expenses. Visual comparison results of SR (×4) are illustrated in Figure 5; the patches for comparison are marked with yellow boxes in the original images (left).

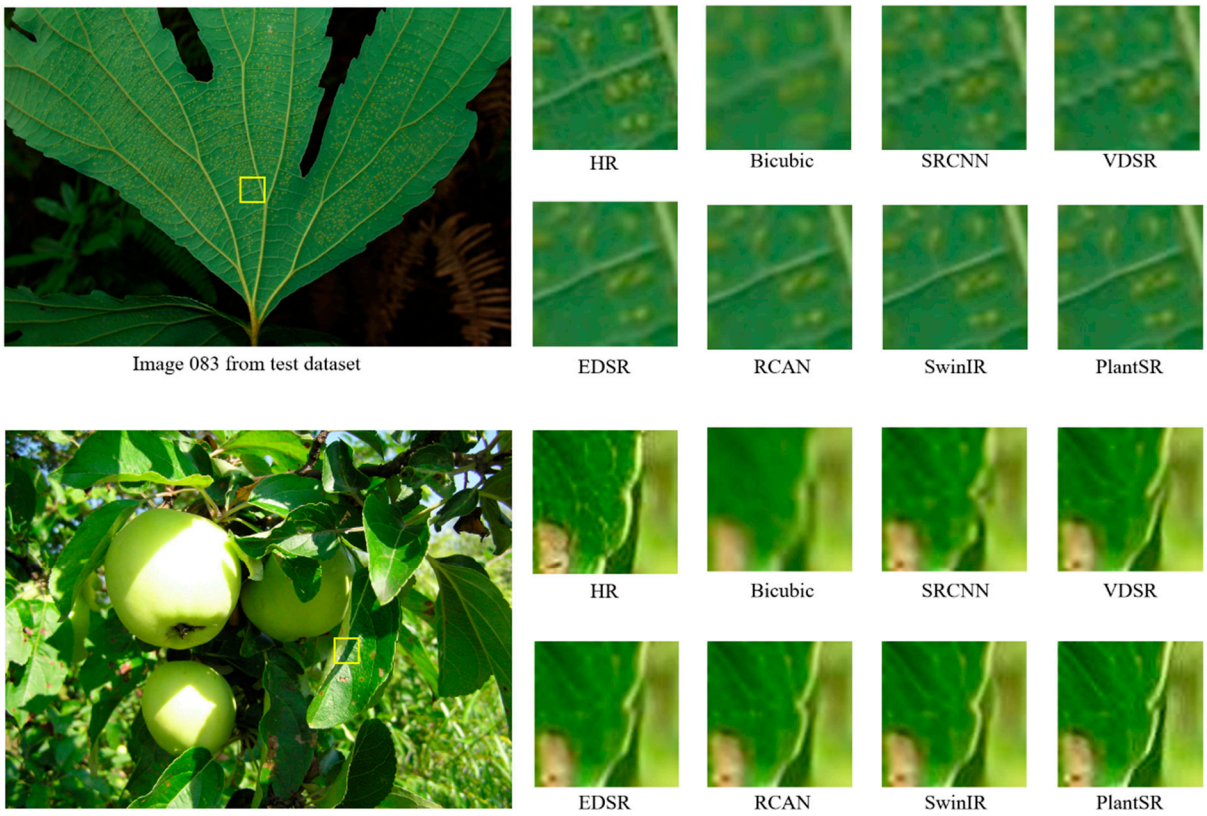


Figure 5. Visual comparison results of image SR (×4) methods.

3.2. Super-Resolution Effects on the Apple Counting Task

We trained a YOLOv7 model using 1000 images collected from the internet and evaluated the model on test images with or without using PlantSR (×3) to achieve super resolution. Before evaluation, we collected some high-resolution close-up images of apples to train a specialized PlantSR (×3) model that had superior performance in upscaling apple images. We used the PlantSR (×3) model pretrained on the PlantSR dataset as the initialization model and conducted further training on these high-resolution apple images, thereby obtaining a specialized PlantSR (×3) model for processing apple images. Before applying super resolution to test images, the resolution of these images was 270 in width and 185 in height. After applying the PlantSR (×3) model, the resolution of these images increased to 810 (270 × 3) in width and 555 (185 × 3) in height. Next, we tested the YOLOv7 model separately under these two cases of test images.

Before applying the PlantSR model to achieve super resolution, the mean absolute error of the apple counting model on test images was 13.085 (Figure 6a). Notably, a significant portion of the images exhibited an under-detection of apples, with predicted values falling below the ground truth values. However, post-application of super resolution to the test images, the mean absolute error notably decreased to 5.71, which showed a great improvement in the model performance (Figure 6b).

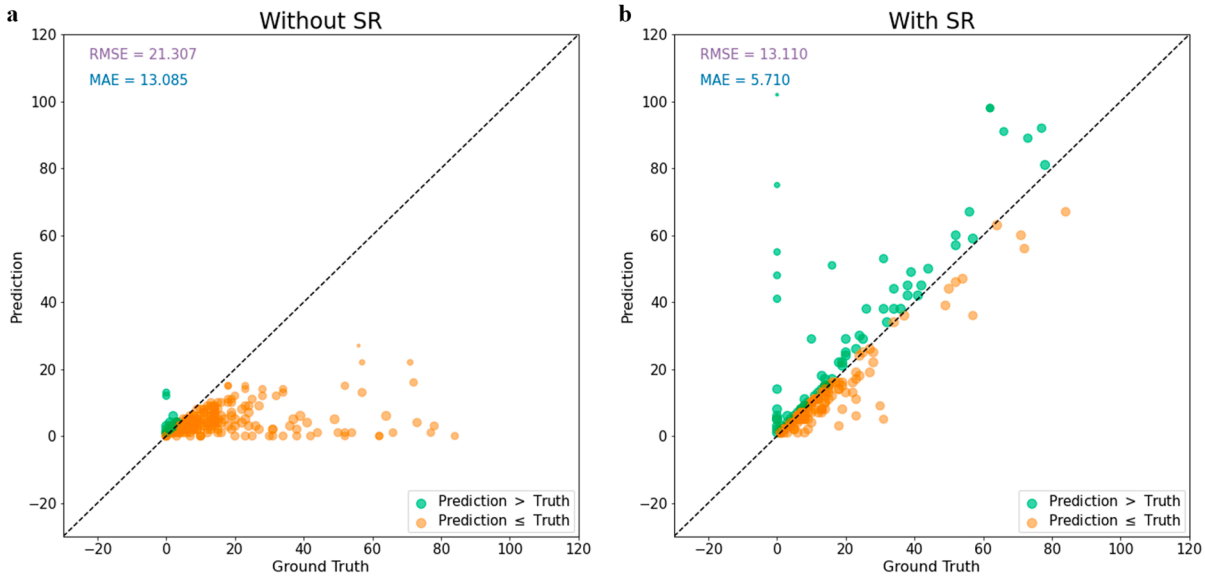


Figure 6. Test results of YOLOv7 with different cases of test images. (**a**) test images without SR. (**b**) test images with SR.

3.3. Super-Resolution Effects on the Soybean Seed Counting Task

We trained and evaluated P2PNet-Soy with images in four different cases (Figure 7): (1) original images across training, validation, and test sets; (2) original training and validation images with downsampling of test images; (3) original training and validation images while test images underwent downsampling followed by upscaling using a PlantSR (×2) model; and (4) PlantSR (×2) upscaled images across training, validation, and test sets.

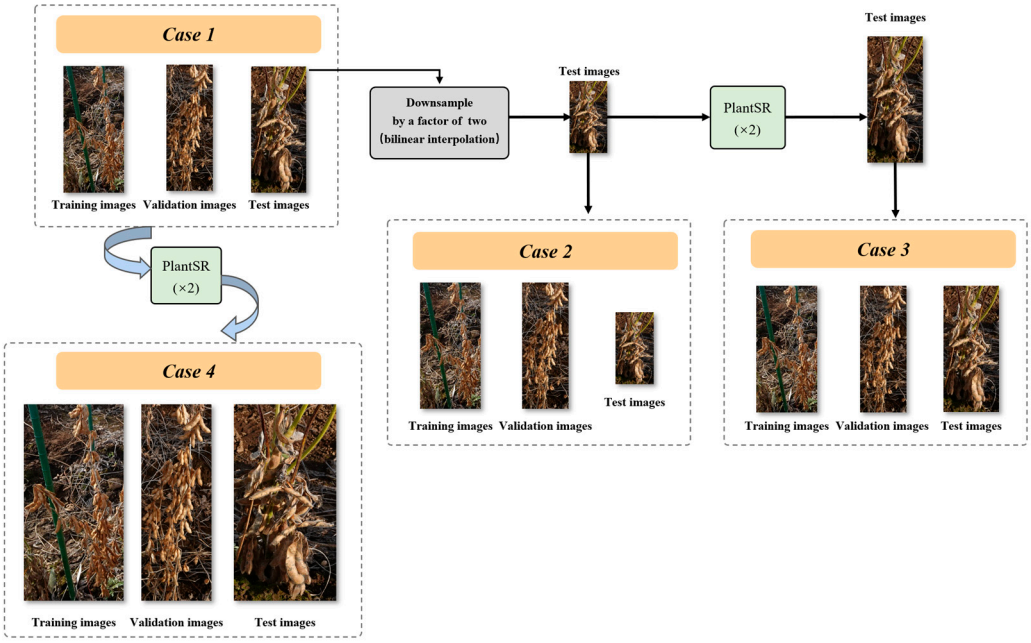


Figure 7. Training, validation, and test images of P2PNet-Soy in four different cases.

For the third and fourth cases, we used a PlantSR (×2) model for processing images. Despite the availability of previously trained models for resolution enhancement, we aimed to train a specialized PlantSR (×2) model with superior performance in upscaling these soybean seed images. To achieve this, we curated a collection of high-resolution close-up images of soybeans from laboratory settings, characterized by a higher pixel density compared to the soybean seed images used for training and evaluating P2PNet-Soy.

We used the PlantSR (×2) model pretrained on the PlantSR dataset as the initialization model and conducted further training on these high-resolution soybean images, obtaining a specialized PlantSR (×2) model for processing soybean seed images.

Table 2 shows the test results of P2PNet-Soy with different cases of images. Comparing the case where all images remained in their original resolution (Case 1) with the case where all images were upscaled using the PlantSR (×2) model (Case 4), we observed a reduction in MAE from 19.16 to 15.09. However, when evaluating the model trained on original images using downscaled test images (Case 2), the MAE increased significantly to 59.23. Notably, implementing the PlantSR (×2) model for these downscaled test images (Case 3) led to a noteworthy reduction in MAE, successfully minimizing it to 19.82, a value close to that achieved in Case 1.

Table 2. Train and evaluate P2PNet-Soy with images in four different cases.

Case	Preprocessing	MAE	RMSE
Case 1	No	19.16	21.49
Case 2	Downsample test images	59.23	61.47
Case 3	Downsample test images and then upscale them using a PlantSR (×2) model	19.82	22.91
Case 4	Upscale all the images	15.09	22.19

4. Discussion

Plant images are often used for automatic analysis and to extract useful information. Due to some limitations, the objects in these images may suffer from low resolution. For instance, when counting soybean seeds through deep neural networks, we usually take images that include the entire soybean plant. Although the entire image is in high resolution, the soybean seeds in the images can still be in low resolution. Another instance is that when using drones to capture images, it is difficult to obtain high-resolution images of the targets on the ground because of the flying altitude of the drone, which results in a longer shooting distance. These limitations in object resolution may lead to insufficient information, resulting in a decrease in the performance of computer vision models.

The foundation of our approach rested on the creation of the PlantSR dataset, comprised exclusively of high-resolution plant images. This dataset not only facilitated the evaluation of SR models but also served as a valuable resource for benchmarking their performance. By ensuring the quality and diversity of images in this dataset, we aimed to provide a robust platform for comparing various SR methodologies.

Through extensive experimentation, we built a PlantSR model with the most effective solution for restoring high-resolution plant images from their low-resolution counterparts. Notably, we deliberately avoided transformer-based architectures in favor of alternative methodologies optimized for efficient inference in plant imagery [45,46]. This decision was driven by the need for practical inference speed and compelling performance gains, as evidenced by achieving the highest peak signal-to-noise ratio (PSNR) across all scaling factors (2, 3, and 4) on the PlantSR dataset.

Furthermore, we conducted experiments to explore the effects of super-resolution on two object detection tasks: apple counting and soybean seed counting. Our results demonstrate significant improvements in counting performance when applying the PlantSR model, particularly when images are initially in low resolution or downsampled. In the soybean seed counting task, compared with the case where all images remained in their original resolution, applying the PlantSR (×2) model to upscale all images successfully reduced the MAE of P2PNet-Soy from 19.16 to 15.09. This enhancement in counting performance can be attributed to several factors. Firstly, the increased pixel density of the soybean seed images facilitated by the SR model enabled the counting model to discern and learn more intricate features. Secondly, by processing all images using the same SR model, a higher level of similarity among the images was achieved, potentially adapting the domain between the training and test images [47]. However, it was observed that downsampling

test images led to a substantial decline in model performance, underlining the significant influence of image resolution on model performance. Notably, implementing the PlantSR (×2) model for these downsampled test images resulted in a remarkable reduction in MAE, minimizing it from 59.23 to 19.89. The apple counting task obtained similar results, and applying the PlantSR (×3) model to the low-resolution test images decreased the MAE from 13.085 to 5.71.

Besides these successes, certain limitations are inherent in our current methodology. Firstly, the PlantSR model, tailored specifically to super-resolution plant images, might not perform optimally for some other tasks in image reconstruction such as artifact reduction and denoising. Secondly, our focus was predominantly on the application of super resolution to RGB images. Nonetheless, beyond RGB images, some other data types captured by various sensors exist, including images from multiple layers of channels taken by multispectral cameras, depth maps obtained by depth cameras, and three-dimensional point clouds acquired through lidar scanning. Super resolution for these diverse data types remains an invaluable and challenging endeavor. Thirdly, enhancing the resolution of input images is likely to result in a slowdown in both the training and inference speed of deep learning models. This is particularly notable given the increased computational cost during the training phase.

In conclusion, our study underscores the potential of super-resolution technology to improve the performance of object detection models in plant image analysis. The development of specialized SR models, exemplified by the PlantSR model, holds immense promise for restoring high-resolution plant images and thereby enhancing the capabilities of computer vision models reliant on such images. Future research directions may involve delving into a generalized model capable of addressing multiple subtasks of image reconstruction for plant images, proposing more extensive models capable of handling a broader range of data types.

Author Contributions: Q.Y. and T.J. conceptualized and designed this study. M.S. and Y.Z. participated in the data collection. T.J. performed the analyses. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (2021YFF1001100).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The source code is available at https://github.com/SkyCol/PlantSR (accessed on 28 November 2023). The PlantSR dataset is available at figshare [48]. The high-resolution soybean seed images are available at figshare [49].

Acknowledgments: We would like to thank Dajian Zhang at Shandong Agricultural University for technical support regarding soybean data.

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

References

1. Walter, A.; Liebisch, F.; Hund, A. Plant phenotyping: From bean weighing to image analysis. *Plant Methods* **2015**, *11*, 14. [CrossRef] [PubMed]
2. Das Choudhury, S.; Samal, A.; Awada, T. Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* **2019**, *10*, 508. [CrossRef] [PubMed]
3. Li, Y.; Jia, J.; Zhang, L.; Khattak, A.M.; Sun, S.; Gao, W.; Wang, M. Soybean seed counting based on pod image using two-column convolution neural network. *IEEE Access* **2019**, *7*, 64177–64185. [CrossRef]
4. Wang, Y.; Li, Y.; Wu, T.; Sun, S. Fast Counting Method of Soybean Seeds Based on Density Estimation and VGG-Two. *Smart Agric.* **2021**, *3*, 111.
5. Khaki, S.; Saeed, N.; Pham, H.; Wang, L. WheatNet: A lightweight convolutional neural network for high-throughput image-based wheat head detection and counting. *Neurocomputing* **2022**, *489*, 78–89. [CrossRef]

6. David, E.; Madec, S.; Sadeghi-Tehran, P.; Aasen, H.; Zheng, B.; Liu, S.; Kirchgessner, N.; Ishikawa, G.; Nagasawa, K.; Badhon, M.A.; et al. Global Wheat Head Detection (GWHD) dataset: A large and diverse dataset of high-resolution RGB-labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics* **2020**, *2020*, 3521852. [CrossRef] [PubMed]
7. David, E.; Serouart, M.; Smith, D.; Madec, S.; Velumani, K.; Liu, S.; Wang, X.; Pinto, F.; Shafiee, S.; Tahir, I.S.A.; et al. Global wheat head detection 2021: An improved dataset for benchmarking wheat head detection methods. *Plant Phenomics* **2021**, *2021*, 9846158. [CrossRef] [PubMed]
8. 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]
9. Luke, J.; Joseph, R.; Balaji, M. Impact of image size on accuracy and generalization of convolutional neural networks. *Int. J. Res. Anal. Rev. (IJRAR)* **2019**, *6*, 70–80.
10. Sabottke, C.F.; Spieler, B.M. The effect of image resolution on deep learning in radiography. *Radiol. Artif. Intell.* **2020**, *2*, e190015. [CrossRef]
11. Shermeyer, J.; Van Etten, A. The effects of super-resolution on object detection performance in satellite imagery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019.
12. Bashir Syed, M.A.; Wang, Y.; Khan, M.; Niu, Y. A comprehensive review of deep learning-based single image super-resolution. *PeerJ Comput. Sci.* **2021**, *7*, e621. [CrossRef]
13. Gendy, G.; He, G.; Sabor, N. Lightweight image super-resolution based on deep learning: State-of-the-art and future directions. *Inf. Fusion* **2023**, *94*, 284–310. [CrossRef]
14. Lepcha, D.C.; Goyal, B.; Dogra, A.; Goyal, V. Image super-resolution: A comprehensive review, recent trends, challenges and applications. *Inf. Fusion* **2023**, *91*, 230–260. [CrossRef]
15. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3365–3387. [CrossRef]
16. Yamamoto, K.; Togami, T.; Yamaguchi, N. Super-Resolution of Plant Disease Images for the Acceleration of Image-based Phenotyping and Vigor Diagnosis in Agriculture. *Sensors* **2017**, *17*, 2557. [CrossRef] [PubMed]
17. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
18. Maqsood, M.H.; Mumtaz, R.; Haq, I.U.; Shafi, U.; Zaidi, S.; Hafeez, M. Super resolution generative adversarial network (Srgans) for wheat stripe rust classification. *Sensors* **2021**, *21*, 7903. [CrossRef] [PubMed]
19. Cap, Q.H.; Tani, H.; Kagiwada, S.; Uga, H.; Iyatomi, H. LASSR: Effective super-resolution method for plant disease diagnosis. *Comput. Electron. Agric.* **2021**, *187*, 106271. [CrossRef]
20. Albert, P.; Saadeldin, M.; Narayanan, B.; Fernandez, J.; Mac Namee, B.; Hennessey, D.; O'Connor, N.E.; McGuinness, K. Unsupervised domain adaptation and super resolution on drone images for autonomous dry herbage biomass estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1636–1646.
21. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 416–423.
22. Timofte, R.; Agustsson, E.; Gool, L.; Yang, M.H.; Zhang, L.; Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M.; et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 114–125.
23. Li, Z.; Liu, Y.; Wang, X.; Liu, X.; Zhang, B.; Liu, J. Blueprint separable residual network for efficient image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 833–843.
24. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
25. Zechmeister, H.G.; Grodzińska, K.; Szarek-Łukaszewska, G. Bryophytes. In *Trace Metals and Other Contaminants in the Environment*; Elsevier: Amsterdam, The Netherlands, 2003; Volume 6, pp. 329–375.
26. Smith, A.R.; Pryer, K.M.; Schuettpelz, E.; Korall, P.; Schneider, H.; Wolf, P.G. A classification for extant ferns. *Taxon* **2006**, *55*, 705–731. [CrossRef]
27. Hutchinson, K.R.S.; House, H. *The Morphology of Gymnosperms*; Scientific Publishers: New York, NY, USA, 2015.
28. Bahadur, B.; Rajam, M.V.; Sahijram, L.; Krishnamurthy, K.V. Angiosperms: An overview. In *Plant Biology and Biotechnology: Volume I: Plant Diversity, Organization, Function and Improvement*; Springer: New Delhi, India, 2015; pp. 361–383.
29. POWO. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. 2024. Available online: http://www.plantsoftheworldonline.org/ (accessed on 11 November 2023).
30. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
31. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging* **2016**, *3*, 47–57. [CrossRef]

32. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
33. Song, Q.; Wang, C.; Jiang, Z.; Wang, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F.; Wu, Y. Rethinking counting and localization in crowds: A purely point-based framework. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3365–3374.
34. Zhao, J.; Kaga, A.; Yamada, T.; Komatsu, K.; Hirata, K.; Kikuchi, A.; Hirafuji, M.; Ninomiya, S.; Guo, W. Improved field-based soybean seed counting and localization with feature level considered. *Plant Phenomics* **2023**, *5*, 0026. [CrossRef]
35. Pharr, M.; Jakob, W.; Humphreys, G. *Physically Based Rendering: From Theory to Implementation*; Morgan Kaufmann: Burlington, MA, USA, 2016.
36. P2PNet-Soy Project. Available online: https://github.com/UTokyo-FieldPhenomics-Lab/P2PNet-Soy (accessed on 21 November 2023).
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
38. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
39. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef] [PubMed]
40. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1646–1654.
41. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
42. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
43. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L.V.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
44. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B.; Wei, Y. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
45. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in Neural Information Processing Systems*; 2017; Volume 30, Available online: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html (accessed on 13 January 2024).
46. Bi, J.; Zhu, Z.; Meng, Q. Transformer in computer vision. In Proceedings of the 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), Fuzhou, China, 24–26 September 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 178–188.
47. Csurka, G. Domain adaptation for visual applications: A comprehensive survey. *arXiv* **2017**, arXiv:1702.05374.
48. Figshare. Dataset. PlantSR Dataset. Available online: https://doi.org/10.6084/m9.figshare.24648150.v1 (accessed on 13 January 2024).
49. Figshare. Dataset. HR_Soybean. Available online: https://doi.org/10.6084/m9.figshare.24994253.v1 (accessed on 28 November 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.