Spatial Super Resolution of Real-World Aerial Images for Image-Based Plant Phenotyping

Masoomeh Aslahishahri 1, Kevin G. Stanley 1, Hema Duddu 2, Steve Sh filiffe 2, Sally Vail 3 and Ian Stavness 1,*

Abstract: Unmanned aerial vehicle (UAV) imaging is a promising data acquisition technique for image-based plant phenotyping. However, UAV images have a lower spatial resolution than similarly equipped in field ground-based vehicle systems, such as carts, because of their distance from the crop canopy, which can be particularly problematic for measuring small-sized plant features. In this study, the performance of three deep learning-based super resolution models, employed as a pre-processing tool to enhance the spatial resolution of low resolution images of three different kinds of crops were evaluated. To train a super resolution model, aerial images employing two separate sensors co-mounted on a UAV flown over lentil, wheat and canola breeding trials were collected. A software workflow to pre-process and align real-world low resolution and high-resolution images and use them as inputs and targets for training super resolution models was created. To demonstrate the effectiveness of real-world images, three different experiments employing synthetic images, manually downsampled high resolution images, or real-world low resolution images as input to the models were conducted. The performance of the super resolution models demonstrates that the models trained with synthetic images cannot generalize to real-world images and fail to reproduce comparable images with the targets. However, the same models trained with real-world datasets can reconstruct higher-fidelity outputs, which are better suited for measuring plant phenotypes.

Keywords: aerial imaging; super resolution; pixel size; raw data pre-processing; radiometric calibration; real-world dataset; plant phenotyping

1. Introduction

The demand for sustainable food production continues to increase due to population growth, particularly in the developing world [1]. Stable global food production is threatened by changing and increasingly volatile environmental conditions in food growing regions [2,3]. Breeding crop varieties that have both high yields and resilience to abiotic stress has the potential to overcome these challenges to global food production and stability [4]. New, data-driven techniques for crop breeding require large amounts of genomic and phenomic information [5]. However, traditional plant phenotyping methods are slow, destructive and laborious [6]. Image-based plant phenotyping has the potential to increase the speed and reliability of phenotyping by using remote sensing technology, such as unmanned aerial vehicles (UAVs), to estimate phenomic information from field crop breeding experiments [7–11]. UAVs can monitor a large field in a relatively short period of time with different imaging payloads including RGB, multispectral, hyperspectral, and thermal cameras [12–15]. UAV images can, in turn, be analyzed to estimate phenotypic traits, such as vegetation, flowering and canopy temperature [16–19].

For applications in crop breeding, detailed phenotypic information is often required, such as counting the number of seedlings or plant organs within a small region of a field...
experiment. In order to adequately capture and estimate these fine plant features, images with high spatial resolution are required. For example, previous work has demonstrated that low spatial image resolution has a substantial negative effect on the accuracy of measuring plant ground cover from images [20]. Extracting textural features from aerial images for vegetation classification also requires high spatial resolution [21]. It has also been shown that the accuracy of detecting and localizing objects in images is highly dependent on image resolution [22,23]. The requirement of high spatial image resolution poses a challenge for UAV imaging, because a number of factors limit the resolution of UAV-acquired images.

The resolution of aerial imaging is directly associated with several factors including the camera sensor’s resolution, the camera lens’ focal length, and the altitude of the UAV flight [24]. A smaller ground sampling distance (GSD), representing a higher spatial resolution, requires lower altitude (shorter distance between the camera and target being imaged), a higher resolution sensor, and/or a longer focal length lens. However, these variables have trade-offs that limit GSD in practice. UAVs are often flown at a high altitude or with a shorter focal length lens, in order to capture the entire field in a short period of time, due to limited battery life [25]. High-resolution sensors are expensive and heavier and must be flown on larger and more expensive UAVs with higher payload capacity. For crop imaging, multispectral sensors are often used, which is the trade-off spatial resolution for spectral resolution [21]. Even if a small GSD is obtained by optimizing sensor and flight parameters, during a flight, atmospheric disturbances and wind can cause image blur and substantially reduce the effective spatial resolution of the acquired images. Managing the tradeoffs between cost and image fidelity forces potentially unpalatable compromises on those monitoring crops.

An alternative to achieving high spatial resolution aerial images is enhancing the details of aerial images algorithmically. Image super resolution is a promising approach for reconstructing a high-resolution (HR) image from its low-resolution (LR) counterpart to improve the image detail. This has an obvious benefit, as finer grained imagery can be derived from lower cost, lighter-weight cameras, allowing smaller, less expensive UAVs to be employed. In an agricultural context, image super resolution could be used to convert LR images to HR images, improving image details for phenotyping tasks such as flower counting. Image super resolution is a challenging problem and an active area of computer vision research. Deep learning-based methods have shown potential to enhance the robustness of traditional image super resolution techniques [26–28]. However, most of the state-of-the-art models were evaluated over public datasets consisting of urban scenes, animals, objects and people [29,30]. While a few studies have applied super resolution methods to plant images, these have been conducted with close-up images of fruit [31] or leaves [32,33]. It is not clear how well super resolution methods would work when applied to aerial images of agricultural fields that include substantial high frequency content and highly self-similar plants within the same spatial region of the field.

Because almost all super resolution methods are tested with synthetic data—datasets where HR images are downsampled with a known kernel (e.g., bicubic interpolation) to obtain LR/HR image pairs—the actual performance of super resolution in practice is also uncertain. A few datasets have been proposed with paired real-world LR and HR images by capturing the same static scene with the same sensor with different focal lengths [34] or with two side-by-side sensors with different focal lengths [35]. Recently, the NTIRE 2020 dataset has been proposed for real-world super resolution with smart phone images [36]. However, a paired LR/HR image dataset of aerial images does not currently exist, and is more challenging to collect, as it requires capturing both LR and HR images of a dynamic scene.

In this paper, we address these shortcomings by providing a paired drone image set of low and high resolution images, and characterize the performance against state of the art super resolution algorithms to establish the utility of super resolution approaches in agriculture, and more generally for other applications from drone images. We provide a
new real-world dataset of LR/HR images collected by co-mounting LR and HR sensors on the same UAV. We evaluate the performance of three deep learning-based super resolution models against the collected data. We evaluate how well a model trained with synthetic data generalizes to real-world LR images. The contributions of this study include:

- A novel dataset of paired LR/HR aerial images of agricultural crops, employing two different sensors co-mounted on a UAV. The dataset consists of paired images of canola, wheat and lentil crops from early to late growing seasons.
- A fully automated image processing pipeline for images captured by co-mounted sensors on a UAV. The algorithm provides image matching with maximum spatial overlap, sensor data radiometric calibration and pixel-wise image alignment to contribute a paired real-world LR and HR dataset, essential for curating other UAV-based datasets.
- Extensive quantitative and qualitative evaluations across three different experiments including training and testing super resolution models with synthetic and real-world images to evaluate the efficacy of real-world dataset in a more accurate analysis of image-based plant phenotyping.

2. Background and Related Work

High spatial resolution images are important for computer vision tasks when fine details about the original scene are required [37]. In remote sensing, the pixel size of aerial images matters because it impacts the accuracy of image analysis [20]. There are usually a variety of ground objects in the sensor’s field of view, where a pixel in a remote sensing image may contain a mixture of ground objects information, called mixed pixel. This can impact information extraction, such as image classification and object detection in mixed pixels [38].

Coarse spatial resolution is likely to increase the number of mixed pixels in aerial images, where the major proportion of the objects existing in a mixed pixel determine the identity of ground objects [39]. To some extent, spectral unmixing can handle the identity of mixed pixels, but it cannot provide spatial distribution information for remote sensing images [40]. The super resolution approach is a process of improving details within an image, predicting finer spatial resolution, especially where the details are essentially unknown to reduce the spatial uncertainty in output images. A finer spatial resolution results in smaller pixel sizes and fewer mixed pixels in an image [41].

Super resolution is a classic problem in computer vision, and generally there are many possible upsampling solutions for a given low resolution image. Machine learning or image processing approaches have suggested different complimentary solutions including traditional methods, and deep learning-based algorithms to increase the visual resolution of an image.

2.1. Traditional Super Resolution Methods

The simplest way to provide a super-resolved image is to apply an interpolation filter on sample image by amplifying existing high-frequency information in the image. However, real-world images do not contain severe high-frequency information, meaning that the algorithm would likely create a blurred output [42].

Generating an HR image from a sequence of LR images taken from the same scene is referred to as multi-image super resolution (MISR). This approach works using iterative back projection, where a set of LR images are registered to reconstruct a single HR image by aggregating LR images. MISR can be compromised by registration errors, if the resolution of the original images is not sufficient to observe fine features, or the number of unknown samples is large, which can result in substantial computational costs [43].

Single image super resolution (SISR) refers to the task of restoring an HR image from a single LR image, requiring image priors. The core idea of learning-based SISR methods is to learn the mapping between the LR and HR images. This approach predicts the missing high frequency information that does not exist in the LR image and which cannot
be reconstructed by simple image interpolation [44]. SISR is more general than MISR because there are fundamental limits to MISR, such as poor registration, low resolution gain, too much noise, and slow processing speed that deter MISR approach to provide higher resolutions [45].

2.2. Example-Based Super Resolution

The most successful alternative to interpolation and MISR is example-based super resolution with machine learning. Example-based approaches include parametric methods [46] and non-parametric methods [47], which typically build upon existing machine learning techniques. Parametric methods create the super resolution image via mapping functions, controlled by a compact number of parameters, learned from examples that do not necessarily come from the input image. One possible solution is to adapt the interpolation algorithm to the local covariances images [45].

The introduction of the latest machine learning approaches made substantial improvements in non-parametric models in terms of accuracy and efficiency. Non-parametric methods can, in turn, be classified as internal or external learning methods [48].

Internal learning models sample patches obtained directly from the input image. These patches are extracted from the self-similarity property of natural images [49], such that the smallest patches in a natural image are likely to be observed in the down-scaled version of the same image [50,51]. External super resolution methods use a common set of images to predict missing information, especially high frequency data from a large and representative set of external pairs of images. Numerous super resolution methods used paired LR-HR images, collected from external databases. External super resolution papers generally used supervised learning methods, including nearest neighbor [44], and sparse coding [52,53,55–55].

2.3. Deep Learning-Based Super Resolution

Deep learning techniques have the potential to improve the robustness of traditional image super resolution methods and reconstruct higher quality images in various scales. Many deep learning-based networks achieve significant performance on SISR, employing a synthetic LR image as input for the super resolution model. Some models use predefined interpolation as the upsampling operator to produce an intermediate image with the desired size and feed that image into the network for further refinement [56]. Residual learning and recursive layers were further employed in this framework to enhance image quality [57,58].

Single upsampling frameworks offer an effective solution to increase the spatial resolution at the end of the learning process while learning the mapping in the LR space without a predefined operator [59,60], where residual learning helps to exploit fine features in the LR space [27,61,62]. Progressive upsampling gradually reconstructs HR images, allowing the network to first discover the actual scale structure of the LR image and then predict finer scale details [26]. An iterative up and downsampling framework increases the learning capacity employing concatenation operator [28].

Several different super resolution strategies have been proposed to enhance the quality of real-world images, because the performance of learning-based super resolution drops when the kernel deviates from the true kernel [63]. Most super resolution models which estimate the true kernel involve complex optimization procedures [64–69]. Deep learning is an alternative strategy for estimating the true kernel for real-world image super resolution [70–72]. Deep learning-based models assume that the degradation kernels are not available and attempt to estimate the optimal kernel, either using self-similarity properties [64,70,71] or adapting an alternating optimization algorithm [63,73]. These models have shown promising results in up to 4× image super resolution.
2.4. Image Pre-Processing for Raw Data

Raw image processing employs raw sensor data to enhance the image quality for computer vision tasks. A maximum a posteriori method was proposed to jointly stitch and super-resolve multi-frame raw sensor data [74]. A deep learning-based model was employed for demosaicing and denoising simultaneously [75]. A similar work was proposed to denoise a sequence of synthetic raw data [76]. An end-to-end deep residual network was proposed to learn mapping between raw images and targets to recover in high-quality images [77]. In [78], a learning-based pipeline was introduced to process low-light raw images with their corresponding reference images in an end-to-end fashion.

3. Materials and Methods

In this section, we introduce our aerial image dataset, detail the image pre-processing steps used to obtain paired LR/HR images, and describe the experiments for evaluating super-resolution models with the dataset.

3.1. Dataset Collection

For training with real-world images, we collected original datasets consisting of raw LR and HR images, captured by two separate sensors co-mounted on a UAV, visualized in Figure 1. As the two sensors have different configurations such as discrepancies in capture rate, different spectral bands, and different fields of view, we first need to match images with the maximum overlap of the same scene, calibrate images radiometrically, and align the calibrated images pixel-wise with a registration algorithm. The created dataset enables training a model in an end-to-end fashion to jointly deblur and super-resolve the real-world LR images. Our dataset will be made publicly available to encourage the development of computer vision methods for agricultural imaging and help enable next generation image-based plant phenotyping for accurate measurement of plants and plant organs from aerial images.

![Figure 1. The two sensors co-mounted on a UAV. The red sensor is the Micasense Rededge (LR images) and the black sensor is IXU-1000 PhaseOne (HR images).](image-url)

To create our dataset, we obtain aerial images of the entirety of three breeding field trials of canola, lentil and wheat crops employing two different sensors. The lentil and wheat fields are located side by side, which are captured together during the same flight. The number of images in the lentil and wheat datasets is low, because the flight period used to capture these two fields is equivalent to the time spent on the canola field due to limited...
battery life. The canola image acquisition experiment was conducted at the Agriculture and Agri-Food Canada Research Farm near Saskatoon (52.181196°N, −106.501494°E), SK, and the lentil and wheat image acquisition experiment was conducted at Nasser location (52.163455°N, −106.515038°E) of Kernen Crop Research Farm, Saskatoon, SK in Canada.

The images were acquired over an entire growing season ranging from early, mid to late growth stages over the three fields in 2018. The canola dataset contains images from both 2018 and 2019 growing seasons. Table 1 shows the diversity of growth stages in each dataset related to the captured date and the number of images in each date. The total number of matched images in the canola and in the lentil and wheat datasets are 1201 and 416 images, respectively. As the lentil and wheat datasets were captured together during a single flight, we keep these two datasets together in Table 1, because the calibrated panel reflection (CRP) required to match the LR and HR images based on their relative timestamp, is merely captured once during a flight. To match the LR and HR images pair-wise, we need to obtain the capture time of CRP’s image via its metadata (Figure 2). The numbers of the lentil and wheat images are individually 210 and 206 images respectively, shown, as in total, 416 images in Table 1. The lentil and wheat datasets will be separated later for further analysis when training and testing the models.

Table 1. The number of images captured by each sensor and the number of matched images for different acquisition dates.

| Date       | Canola Rededge (LR) | IXU-1000 (HR) | Lentil and Wheat Rededge (LR) | IXU-1000 (HR) | Matched Images Canola/Lentil-Wheat |
|------------|---------------------|---------------|-------------------------------|---------------|-----------------------------------|
| 28 June 2018 | 293                 | 287           | 306                           | 303           | 239/159                          |
| 27 July 2018 | -                   | -             | 228                           | 229           | -/127                            |
| 30 July 2018 | 335                 | 337           | -                             | -             | 262/-                            |
| 29 August 2018 | -                   | -            | 228                           | 228           | -/130                            |
| 31 August 2018 | 370                 | 364           | -                             | -             | 283/-                            |
| 10 July 2019 | 474                 | 457           | -                             | -             | 417/-                            |
| Total       | 1472                | 1445          | 762                           | 760           | 1201/416                         |

Each dataset covers a diverse range of genotypes. The canola field has 56 varieties of canola, such that substantial variations exist in appearance in the dataset. The wheat field contains 17 varieties in microplots, while the lentil field has 16 varieties in microplots. The wheat or lentil trial has 47 rows and 12 columns of microplots in three aligned blocks. The spacing between rows within a plot in the canola trial and in the lentil and wheat trials were approximately 0.3 m and 0.6 m, respectively. Low sowing density was observed in the lentil field and normal sowing density in the canola and wheat fields [79,80].

For capturing low resolution images, we used the Micasense Rededge camera, a widely used multispectral (5-band) camera for monitoring crops. For capturing high resolution images, we used the Phase One IXU-1000 camera, a high-end medium format RGB camera generally used for commercial purposes. Both cameras were mounted to an Ronin-MX gimbal on a DJI Matrice 600 UAV and captured images simultaneously during the flight. Canola images were captured with the UAV at 20 m altitude, and the lentil and wheat fields were captured from 30 m above the ground. The two sensors are substantially different in their configurations, summarized in Table 2, when the two sensors are at the same altitude about 20 m. For example, LR images represent a smaller area of the field than HR images, because the Phase One IXU-1000 sensor has a wider field of view. At a 20 m flight altitude, the GSD between the two sensors was 1.7 mm/pixel and 12.21 mm/pixel respectively, which represents a significant difference in spatial resolution, meaning small detailed information is missing in the LR images.

Multispectral imaging devices generally measure the energy and light emissions from the surrounding environment. The Micasense Rededge multispectral sensor uses a downwelling light sensor (DLS) in conjunction with MicaSense Rededge’s CRP. To
compensate for the illumination conditions during the flight, CRP images are usually captured before or after each flight. The panel image provides a precise representation of the amount of light variations during the flight and on the ground at the time of capture. This helps to enhance reflectance estimates in situations where the surrounding illumination conditions are changing. This situation requires a conversion to normalized reflectance [81].

Table 2. Configuration comparison between Phase One IXU-1000 and Micasense Rededge sensors.

|                         | Phase One IXU-1000 | Micasense Rededge |
|-------------------------|--------------------|-------------------|
| Capture rate            | ~1 fps             | ~1 fps            |
| Field of view           | 64°                | 47.2°             |
| Sensor resolution       | 100 MP             | 5 × 1.2 MP        |
| Spectral bands          | RGB                | B, G, R, redge, NIR |
| Canola flight altitude  | 20 m               | 20 m              |
| Canola ground sample distance (GSD) | 1.7 mm/pixel | 12.21 mm/pixel |
| Wheat/lentil flight altitude | 30 m            | 30 m              |
| Wheat/lentil ground sample distance (GSD) | 2.6 mm/pixel | 18.31 mm/pixel |

3.2. Image Pre-Processing

During data collection, the two sensors were not synchronized for every frame, meaning images may not be perfectly matched. In addition, the separate spectral sensors of the LR camera require color calibration to obtain a unitless RGB image. Therefore, the following pre-processing steps were performed to create the paired LR/HR dataset for super resolution model training: (1) match LR and HR images with sufficient spatial overlap, (2) correct LR images using at-surface reflectance conversion; and (3) employ a registration algorithm for pixel-wise alignment between LR and HR images.

3.2.1. Image Matching

As the two sensors were not synchronized during flight, we first need to find closely matching LR and HR pairs based on their relative timestamp. To pair training images, we extract the metadata from images of the CRP captured by the two sensors simultaneously. The metadata Original DateTime of the two images are parsed to acquire the time difference between the two sensors, enabling the computation of a threshold value used for matching images with maximum spatial overlap. Figure 2 formally describes the image matching process between LR and HR images, while calibrating the LR images radiometrically.

3.2.2. Radiometric Calibration

The Micasense Rededge sensor has been equipped with a DLS to measure the incident light, because images do not typically provide a uniform radiometric response due to atmospheric disturbances. To utilize the image information in quantitative analysis, the data need to be radiometrically calibrated.

Reflectance measurement in passive imaging is influenced by illumination condition, scattering light in the atmosphere, surrounding environment, the spectral and directional reflectance of an object, and the object’s topography [82]. We used the proscribed process for the Micasense Rededge [83] to convert raw multispectral pixels to radiance and then to reflectance as unitless images. To convert raw pixels into reflectance, an image of the CRP is obtained before the flight, which is used for the raw LR images calibration radiometrically. To determine the transfer function from raw pixels to reflectance for the $i$th band of the multispectral image:

$$F_i = \frac{\rho_i}{\text{avg}(L_i)}$$  \hspace{1cm} (1)
where $F_i$ denotes the calibrated reflectance factor for band $i$. $\rho_i$ represents the average reflectance of the CRP for the $i$th band. $\text{avg}(L_i)$ represents the average value of the radiance for pixels inside the panel for the $i$th band.

![Figure 2](image_url)

Figure 2. Matching raw LR and HR images while calibrating the LR images radiometrically. ‘sec’ denotes second.

The workflow includes compensations for sensor black-level, sensor sensitivity, sensor gain and exposure settings, and vignette effects of the lens. All of the required parameters were obtained from metadata information inside the raw LR images.

In multispectral captures, each band must be aligned with other bands to represent meaningful data. To do this, all of the min and max values over all five channels are adjusted to the minimum and maximum range of all channels. When the transformation is completed, the final image appears in true color.

3.2.3. Multi-Sensor Image Registration

The LR and HR images require to be aligned, where every pixel of the LR image corresponds to a group of pixels of the HR image. As the field of view of the two sensors is dissimilar, pixel-wise alignment is required between the LR and HR datasets. We employ the scale invariant feature transform (SIFT) registration algorithm to perform this task [84].

For image registration, we use matched and radiometrically calibrated images. First, we register multispectral images over different bands, because multispectral sensors cause spatial misalignment between different bands. We followed a traditional approach for multispectral image registration, in which we set a band as a reference and align the other bands accordingly [85]. Second, we register the HR image with its corresponding aligned RGB LR image. For an accurate pixel-wise registration, we alternate between the LR and HR images. We first resize the LR image to the size of the HR image and then designate the LR image as a reference image and align the HR image accordingly. We chose the
LR image as reference because the field of view of the HR sensor is wider. We perform another registration, resizing the HR image obtained from the first alignment phase to the original size of the LR image and hold the HR images as reference and align the LR image accordingly. Figure 3 demonstrates samples of the raw and aligned images existing in our dataset. Small number of paired images in our dataset have oblique view that have not been discarded, because they do not impact the process of training and testing of super resolution models.

Our crafted dataset consists of paired images where the scale of the LR and HR images is not a fixed factor. The size of the raw LR images per band and raw HR images are respectively (960, 1280) and (8708, 11,608, 3), such that the HR image is approximately 9× larger than the LR counterpart. The proposed image pre-processing pipeline can process LR and HR images in their actual sizes for registration. However, in this work, we employed cropped images for image registration to avoid costly computations and memory overload.

We randomly crop the red, green and blue bands of the LR sensor to 500 × 500 pixels and align all three bands pixel-wise to create an RGB composite. We crop the corresponding region from the HR counterpart and align the LR and HR images pixel-wise. After pixel-wise image registration, the HR images are approximately 8× larger in height and width, because the Phase One IXU-1000 sensor has wider field of view and some features are cut off in co-registered images such that the scale size between the aligned LR and HR images is roughly 8×.

We then resized the HR images using bicubic interpolation to create an exact 8× larger HR image than the LR counterpart to facilitate training. If the number of keypoints between the two images during image registration is less than 10 (the default number used by the SIFT algorithm), our algorithm skips those images and moves to the next pair.

To create our dataset, we cropped only one patch out of each LR and HR image during image registration. The proposed algorithm has the capacity to extract more patches from one image. One patch per image was used to reduce computational time.

We employ our dataset to train and test an 8× super resolution model. After image registration, we obtain a dataset consisting of 1581 paired images, including lentil, wheat and canola crops. We split the lentil and wheat dataset to analyse each dataset individually. There are 198 paired images in the lentil, 197 paired images in the wheat, and 1186 paired images in the canola dataset. The number of total paired images in each dataset is slightly different from matched numbers reported in Table 1, because we used random uniform cropping over the LR and HR images for image registration and the automated algorithm skips images where there are no sufficient keypoints, especially in patches with most ground view. The number of paired images may change between runs of the algorithm due to different cropped regions. We also have a dataset combining the three crops called a three-crop dataset.

3.3. Experiments

We evaluated three deep learning-based super resolution models using our dataset. We focus on three different experiments, including synthetic-synthetic, synthetic-real, and real-real strategies to evaluate the performance of each model across real-world images.

3.3.1. Model Selection

Several different models have been introduced to super resolve real-world LR images [63,70,71]. These models have demonstrated promising results for 4× super resolution. For an 8× image super resolution, we employ three state-of-the-art SISR baseline models, including LapSRN [26], SAN [27], and DBPN [28], along with bicubic interpolation.

LapSRN [26] is built upon pyramid blocks to learn feature mapping progressively at log₂M levels, where M is the scaling factor. A progressive upsampling architecture tries to decompose a difficult task into simple tasks to reduce the learning difficulty, especially for larger scaling factors in image super resolution. At each stage, the images are upsampled to an intermediate higher spatial resolution used as input for the subsequent modules.
Figure 3. Example images from our canola (top two rows), lentil (middle two rows), and wheat (bottom two rows) datasets, showing: (a) unaligned raw LR image, (b) the matched raw HR image, (c) radiometrically calibrated and aligned LR cropped patch, (d) the corresponding aligned HR cropped patch. The LR images are $8 \times$ smaller than the HR counterparts.
SAN [27] is a direct learning model which learns salient representations in LR space and upsamples images into desired size via the last convolutional layer to improve computational efficacy. A second-order attention block was employed to learn the spatial correlations of features and improve feature expression.

DBPN [28] leverages the back projection algorithm to capture the feature dependencies between LR and HR pairs. This framework employs up and downsampling layers iteratively and reconstructs the HR image resulting from all intermediate connections. The iterative back projection algorithm helps to compute the reconstruction error based on the connections between the LR and HR spaces and fuse it back to the network to enhance the quality of reconstructed images.

These models were selected as representatives of their respective architectures. All these models have achieved results that are near the state-of-the-art on benchmark datasets, such as BDS100 [86], Set5 [49], Set14 [55], Urban100 [29], and Manga109 [87], reported in [26–28]. Bicubic interpolation was chosen to provide a baseline for comparison.

3.3.2. Experimental Setup

To test the efficacy of the state-of-the-art super resolution models on aerial crop images, we conduct three experiments to evaluate whether or not using real-world images is needed, or if training with synthetic LR, downsampled HR images, is sufficient.

Experiment 1: Synthetic-Synthetic

In the synthetic-synthetic experiment, we train and test the models on synthetic data, where the LR images are obtained by downsampling HR images via bicubic interpolation. A random $256 \times 256$ patch of the HR image was downsampled $8 \times$ to $32 \times 32$. The downsampled image is used as input for the models. The kernel used in the LR image is therefore known. At inference, we compare the performance of the trained models in the same fashion of training, but with larger patches of size $1024 \times 1024$ downsampled to $128 \times 128$ as input. This experiment is meant to provide an approximate upper bound on the performance of the algorithms, because it solves the simpler known kernel problem.

Experiment 2: Synthetic-Real

In the synthetic-real experiment, we train the models using synthetic data, but then test the models using real-world LR images in our dataset. At inference, we employ a $128 \times 128$ patch of our real-world LR image as input. The output is an image with the size of $1024 \times 1024$. This strategy employs a known kernel problem during training, but the unknown kernel problem at inference, and allows us to evaluate how well models trained on synthetic data generalize to unseen real-world LR images.

Experiment 3: Real-Real

In the real-real experiment, we train and test the models on real-world LR/HR images in our dataset. We employ a $32 \times 32$ random patch of the real-world LR image and its corresponding patch from the real-world HR dataset with the size of $256 \times 256$ as input and target, respectively during training. At inference, a $128 \times 128$ patch of the real-world LR image is used as input to be upsampled $8 \times$ and reaches a size of $1024 \times 1024$. For both training and testing, the kernel is not known, therefore we are solving an unknown kernel problem.

3.3.3. Model Training

We train all models on our dataset from scratch without employing pretrained models. We defined two input pipelines for reading either synthetic or real-world images. For the input pipeline of synthetic images, we only use HR images and their downsampled samples for training or testing the models. For the input pipeline of real-world images, we need to find the corresponding regions between the LR and HR images, where the HR images are approximately $8 \times$ larger than their LR counterpart.
We crop one patch per image during the training and testing of all models. We train the models with stochastic cropped images to enhance the models’ performances, however, we require the same patches over all models at inference to evaluate the performance of the models equivalently.

We use 80% of each dataset for training and 20% as a test set. Here, we again crop one patch per image during training and testing. The input data are normalized to $[0, 1]$. Data augmentation is performed by randomly rotating and horizontally flipping the input. The Adam solver [88] with the default parameters ($\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$) is employed to optimize the network parameters. The learning rate is set at $10^{-4}$ for all models and is decayed linearly. All models are trained for 200 epochs.

3.3.4. Evaluation Measures

To evaluate the performance of the super resolution models, we employ three standard measures, including structural similarity index (SSIM) [89] and peak signal-to-noise ratio (PSNR) [90] used in many super resolution tasks along with the dice similarity coefficient (DSC) [91] to assess segmentation accuracy. SSIM measures the perceived quality of the reconstructed image, based on a distortion-free reference image. The human visual system is likely adapted to image structures [92]. SSIM measures the structural similarity between images based on luminance, contrast, and structures’ comparisons [89]. Higher SSIM value represents how well the high frequency information in an image was preserved. PSNR expresses the ratio between the maximum possible power of the reference image and the power of the reconstructed image that influences the quality of image representation. PSNR measures the pixel-level differences between the reconstructed and reference images instead of visual perception. Consequently, it often indicates lower performance in the reconstructed images of real scenes [90]. Higher PSNR means that more noise was removed from an image. DSC measures the similarity between two images to express the spatial overlap between two segmented regions. Higher DSC represents a larger overlap between the two segmented regions.

To evaluate the image quality in terms of SSIM and PSNR at inference, we transform the reconstructed and target images into YCbCr colorspace. This conversion is more effective in pixel intensity evaluation and object definition compared to RGB, because YCbCr benefits from components separation that enables lower resolution capability of human visual system for color detection with respect to luminosity [93].

In addition to standard image metrics, we also evaluate metrics specifically relevant for plant phenotyping: flower segmentation and excess green index [94] commonly used for agricultural purposes. In plant breeding, flowering is an important phenotype that expresses the potential of yield in plants [95].

For flower segmentation, we transform RGB images to Lab colorspace and employ Otsu’s thresholding [96] to segment flowers in both reconstructed and target images [97]. DSC is applied over the two segmented images. Vegetation identification from outdoor imagery is an important step in evaluating crop’s vigour, vegetation coverage, and growth dynamics [16]. Excess green index [94] provides a grey scale image, outlining a plant region of interest following:

$$ExG = 2g - r - b$$

where $g, r, b$ represent green, red and blue bands respectively. The grey scale image plays a role of a mask to segment the entire image. DSC [91] is used as a segmentation measure between the two segmented images.

4. Results

In this section, we evaluate the performance of the super resolution models trained and tested with each individual experiment. To demonstrate the efficacy of real-world images, we have shown extensive quantitative and qualitative analysis across each individual experiment, employing both standard and domain specific measures.
We evaluate the performance of super resolution models on our dataset, employing the synthetic-synthetic, synthetic-real, and real-real experiments. Tables 3 and 4 represent the performance of the models over the PSNR and SSIM values. In the synthetic-synthetic experiment, deep learning models were able to reconstruct reasonable quality images for the canola and three-crop datasets, but they did not outperform the bicubic interpolation for the lentil and wheat datasets. The results for the synthetic-real experiment indicate that the super resolution models cannot generalize to unseen data. For all crops, models trained on synthetic data resulted in a substantial drop in performance when tested with real-world LR images. The results of the real-real experiment demonstrate that the super resolution models have learnt how to handle unknown blur and noise within the LR images. The models performed well on the canola and three-crop datasets. The sensor’s resolution differences caused indistinct reconstructed images in the lentil and wheat datasets, especially where small representations are missing in the LR input image.

Figure 4 shows the quality of reconstructed images both quantitatively and qualitatively on the synthetic-synthetic experiment. In Figure 5, visual artifacts on the reconstructed images represent overestimation of flowering where the models try to transfer the mapping learned during training into unseen data. Figure 6 illustrates the quantitative and qualitative performances of the models on the real-real experiment, which are substantially improved as compared to the synthetic-real experiment.

We observe that SAN [27] fails to reproduce lentil and wheat images tested on either experiment, as visualized in Figures 4–6, because SAN [27] has a deep architecture and it diverges quickly after few iterations over a small training set with less complex features.

Tables 5 and 6 provide results for the accuracy of flower and vegetation segmentation. The models employing the synthetic-synthetic and synthetic-real experiments outperformed or achieved close results with bicubic interpolation in flowering and vegetation segmentation. The models trained and tested with the real-real experiment (excepted SAN) significantly outperformed bicubic interpolation in vegetation segmentation, even though the standard measures (in Tables 3 and 4) did not present substantial quantitative improvements. This means that the models may not produce sufficiently sharp and clear images to satisfy human perception, but they can reconstruct images which reliably reflect agricultural measures such as flowering and vegetation segmentation.

Figure 7 represents flower segmentation on the canola dataset employing the three different experiments. Figures 8–10 demonstrate vegetation segmentation on the canola, lentil and wheat datasets respectively, for the synthetic-synthetic, synthetic-real, and real-real experiments.
Table 3. Average PSNR for $8 \times$ super resolution with different algorithms for the synthetic-synthetic, synthetic-real, and real-real experiments. The highest values are indicated in red.

| Algorithm   | Scale | Canola       | Lentil       | Wheat        | Three-Crop       |
|-------------|-------|--------------|--------------|--------------|-----------------|
|             |       | Syn-Syn 34.68 | Syn-Syn 29.10 | Syn-Syn 29.10 | Syn-Syn 35.14   |
|             |       | Syn-Real 28.83 | Real-Real 28.83 | Real-Real 28.83 | Real-Real 34.21 |
| Bicubic     | 8     |              |              |              |                 |
| LapSRN [26] | 8     | 35.32        | 28.61        | 34.75        | 28.67           |
| SAN [27]    | 8     | 35.57        | 29.16        | 32.88        | 29.90           |
| DBPN [28]   | 8     | 35.63        | 29.15        | 32.88        | 32.26           |

Table 4. Average SSIM for $8 \times$ super resolution with different algorithms for the synthetic-synthetic, synthetic-real, and real-real experiments. The highest values are indicated in red.

| Algorithm   | Scale | Canola       | Lentil       | Wheat        | Three-Crop       |
|-------------|-------|--------------|--------------|--------------|-----------------|
|             |       | Syn-Syn 0.822 | Syn-Syn 0.743 | Syn-Syn 0.821 | Syn-Syn 0.763   |
|             |       | Syn-Real 0.763 | Real-Real 0.763 | Real-Real 0.804 | Real-Real 0.804 |
| Bicubic     | 8     |              |              |              |                 |
| LapSRN [26] | 8     | 0.85         | 0.688        | 0.772        | 0.728           |
| SAN [27]    | 8     | 0.858        | 0.74         | 0.78         | -               |
| DBPN [28]   | 8     | 0.86         | 0.731        | 0.776        | 0.806           |

Table 5. Average flower segmentation DSC value for $8 \times$ super resolution with different algorithms for the synthetic-synthetic, synthetic-real, and real-real experiments. The highest values are indicated in red.

| Algorithm   | Scale | Canola       | Lentil       | Wheat        | Three-Crop       |
|-------------|-------|--------------|--------------|--------------|-----------------|
|             |       | Syn-Syn 0.854 | Syn-Syn 0.7 | Syn-Syn 0.694 | Syn-Syn 0.763   |
|             |       | Syn-Real 0.887 | Real-Real 0.897 | Real-Real 0.793 | Real-Real 0.793 |
| Bicubic     | 8     |              |              |              |                 |
| LapSRN [26] | 8     | 0.897        | 0.696        | 0.79         | -               |
| SAN [27]    | 8     | 0.894        | 0.699        | 0.79         | 0.89            |
| DBPN [28]   | 8     |              |              |              |                 |
Table 6. Average vegetation segmentation DSC value for 8 × super resolution with different algorithms for the synthetic-synthetic, synthetic-real, and real-real experiments. The highest values are indicated in red.

| Algorithm     | Scale | Canola       | Lentil       | Wheat        | Three-Crop   |
|---------------|-------|--------------|--------------|--------------|--------------|
|               |       | Syn-Syn      | Syn-Real     | Real-Real    | Syn-Syn      | Syn-Real     | Real-Real    | Syn-Syn      | Syn-Real     | Real-Real    |
| Bicubic       | 8     | 0.849        | 0.687        | 0.687        | 0.478        | 0.319        | 0.622        | 0.288        | 0.288        | 0.779        |
| LapSRN [26]   | 8     | 0.86         | 0.663        | 0.833        | 0.469        | 0.437        | 0.433        | 0.588        | 0.481        | 0.801        |
| SAN [27]      | 8     | 0.866        | 0.685        | 0.846        | -            | -            | -            | -            | -            | 0.809        |
| DBPN [28]     | 8     | 0.868        | 0.688        | 0.846        | 0.561        | 0.354        | 0.456        | 0.645        | 0.371        | 0.577        |

Figure 4. Example test images from the synthetic-synthetic experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. PSNR and SSIM values are listed below each image. Black images for SAN denote that the model failed to converge when trained on the lentil and wheat datasets.
**Figure 5.** Example test images from the synthetic-real experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. PSNR and SSIM values are listed below each image. Black images for SAN denote that the model failed to converge when trained on the lentil and wheat datasets.

| Dataset | Ground truth HR | bicubic | LapSRN [26] | SAN [27] | DBPN [28] |
|---------|-----------------|--------|-------------|----------|-----------|
| Canola  | ![Canola HR](image1) | ![Canola bicubic](image2) | ![Canola LapSRN](image3) | ![Canola SAN](image4) | ![Canola DBPN](image5) |
| Lentil  | ![Lentil HR](image6) | ![Lentil bicubic](image7) | ![Lentil LapSRN](image8) | ![Lentil SAN](image9) | ![Lentil DBPN](image10) |
| Wheat   | ![Wheat HR](image11) | ![Wheat bicubic](image12) | ![Wheat LapSRN](image13) | ![Wheat SAN](image14) | ![Wheat DBPN](image15) |

**Figure 6.** Example test images from the synthetic-real experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. PSNR and SSIM values are listed below each image. Black images for SAN denote that the model failed to converge when trained on the lentil and wheat datasets.

| Dataset | Ground truth HR | bicubic | LapSRN [26] | SAN [27] | DBPN [28] |
|---------|-----------------|--------|-------------|----------|-----------|
| Canola  | ![Canola HR](image1) | ![Canola bicubic](image2) | ![Canola LapSRN](image3) | ![Canola SAN](image4) | ![Canola DBPN](image5) |
| Lentil  | ![Lentil HR](image6) | ![Lentil bicubic](image7) | ![Lentil LapSRN](image8) | ![Lentil SAN](image9) | ![Lentil DBPN](image10) |
| Wheat   | ![Wheat HR](image11) | ![Wheat bicubic](image12) | ![Wheat LapSRN](image13) | ![Wheat SAN](image14) | ![Wheat DBPN](image15) |
Figure 7. Example test images for vegetation segmentation for the synthetic-synthetic (top row), synthetic-real (middle row), and real-real (bottom row) experiments on the canola dataset during the flowering stage to segment flowers. Yellow pixels indicate flower segmentation (foreground) and DSC values are listed below each image.

Figure 8. Example test images from the synthetic-synthetic experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. Black pixels indicate vegetation segmentation (foreground) and DSC values are listed below each image.
Figure 9. Example test images from the synthetic-real experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. Black pixels indicate vegetation segmentation (foreground) and DSC values are listed below each image.

|          | Ground truth | Bicubic | LapSRN [26] | SAN [27] | DBPN [28] |
|----------|--------------|---------|-------------|-----------|-----------|
| Canola   | (DSC) (0.931)| (0.93)  | (0.934)     | (0.934)   |           |
| Lentil   | (DSC) (0.782)| (0.8)   |             | (0)       | (0.79)    |
| Wheat    | (DSC) (0.791)| (0.792) |             | (0)       | (0.801)   |
| Real LR  | (DSC) (0.791)| (0.792) |             | (0)       | (0.81)    |

Figure 10. Example test images from the real-real experiment for canola, lentil and wheat datasets (top/middle/bottom rows). Ground truth HR images (left) are shown with the cropped region denoted by a red box. Black pixels indicate vegetation segmentation (foreground) and DSC values are listed below each image.

|          | Ground truth | Bicubic | LapSRN [26] | SAN [27] | DBPN [28] |
|----------|--------------|---------|-------------|-----------|-----------|
| Canola   | (DSC) (0.908)| (0.94)  | (0.955)     | (0.96)    |           |
| Lentil   | (DSC) (0.872)| (0.912) |             | (0)       | (0.918)   |
| Wheat    | (DSC) (0.791)| (0.792) |             | (0)       | (0.81)    |

5. Discussion

In this study, we applied super resolution methods to aerial crop images to evaluate how well these methods work for image-based plant phenotyping. We have flown a UAV with two sensors to capture low and high resolution images of a field simultaneously, which enables us to train super resolution models with real-world images. Once trained, these super resolution models would allow a researcher, breeder or grower to fly a low-cost and low-resolution UAV and then enhance the resolution of their images with our pretrained models. We conducted three different experiments, including the synthetic-synthetic,
synthetic-real, and real-real for training and testing super resolution models to demonstrate the efficacy of employing a real-world dataset in aerial image super resolution.

The synthetic-synthetic super resolution models have achieved comparable results with bicubic interpolation over the canola dataset and underperformed in comparison with bicubic interpolation over the lentil and wheat datasets on PSNR and SSIM values. The positive results are likely because the synthetic LR images, even when $8 \times$ downsampling from HR images, are fairly clear and plant features can be well identified. However, the performance of the synthetic-synthetic experiment cannot guarantee achieving the same performance when the models were tested with real-world low resolution images.

The super resolution models trained in the synthetic-real experiment were not able to generalize the feature mapping function learned from synthetic LR image to unseen real-world LR images. This is likely due to the limited or missing overlap between the feature spaces of the two datasets [98]. Small features learned during training on the real-world HR datasets do not exist in the real-world LR test set, because of the GSD differences between the two sensors. Consequently, the super resolution models were prone to overestimating, instead of reconstructing those small features, visualized in Figure 5. The models also failed to reconstruct higher spatial resolution images with smaller pixel sizes to segment flowering and vegetation more precisely, shown in Figures 7 and 9. Therefore, using synthetic data (downsampled HR images) to train super resolution models may not be a practical technique for agricultural analysis.

The super resolution models trained in the real-real experiment have achieved promising results with the canola dataset on both standard and agricultural measures. The lentil and wheat datasets achieved comparable results with bicubic interpolation on PSNR and SSIM values. The three-crop dataset has demonstrated close quantitative PSNR and SSIM values to the canola dataset, as shown in Tables 3 and 4, meaning the distribution of the three-crop dataset is skewed towards the canola dataset.

The real-real experiment outperformed the synthetic-real experiment in every condition tested, therefore we conclude that creating a paired dataset of real-world low and high resolution images is essential for super resolution in agricultural analysis. The reconstructed images over the real-real experiment seem to be sufficiently reliable for the agricultural analysis tested. Plants like wheat which present visibly small phenotypes require smaller pixel size to accurately measure plant density, especially early in the growing season. In our case, the UAV was flown at a higher altitude over the lentil and wheat trials to capture the entirety of the two trials in a relatively short period of time (Section 3.1). Consequently, from a given focal length at a higher altitude, the ground object scales are smaller and require higher spatial resolution. Figure 10 demonstrates the performance of the super resolution models on the real-real experiment, where the smaller pixel sizes, resulting in higher spatial resolution with smaller GSD, aid in more accurate vegetation segmentation in the canola and lentil datasets, but not for the wheat images. This means that the wheat dataset requires either to be captured with higher resolution sensor or the remote sensing device flies at lower altitude.

Smaller pixel sizes also aid in more accurate flower segmentation on the canola dataset in the real-real experiment, as shown in Figure 7. We did not assess flower segmentation on the lentil and wheat dataset, because the flowers on those crops tend to be too small or occluded in aerial imagery, particularly for our lentil and wheat trials that were captured from a higher UAV altitude. The difference between the real LR and ground truth of the lentil and wheat images is visualized in Figure 6.

The importance of pixel size is related to mixed-pixel effects, where a pixel may contain multiple information about different phenotypes of a plant and image background. In such cases, general thresholding techniques such as Otsu’s method [96] calculate a measure of spread for pixels segmented into foreground or background classes. Larger mixed pixels are likely to increase the error rate of plant trait estimation because the plant’s phenotype may dissolve in the background or foreground, as shown in Figure 9, where foreground features such as flowers blend into the green background of vegetation.
The differential performance of the selected models over three different trials and experiments indicates that no model that performs best in every condition, given our dataset. Each model selected has a unique architecture leading to data dependent differences in performance. For post-upsampling architectures including SAN [27], all computational processing occurs in a low spatial resolution space which decreases the computational cost, but can result in reconstruction compromises, especially with large scaling factors. In progressive learning architectures such as LapSRN [26], a difficult task is decomposed into several simple tasks, which reduces the learning difficulty for larger scaling factors. However, as progressive learning is a stacked model, it relies on limited LR features in each block, which can result in lower performance, especially for small scaling factors. Iterative up and downsampling networks represented by DBPN [28] in this work, benefit from the recurrence of internal patches across scales in input images [71]. DBPN [28] focuses on generating a deeper feature map by alternating between the low and high spatial resolution spaces and computing the reconstruction error in each up and downsampling block. For example, DBPN [28] has outperformed the other models in PSNR values across the canola, lentil and wheat datasets over the real-real experiment, but SAN [99] achieved similar performance to DBPN [28] in SSIM values and flower segmentation. In vegetation detection, DBPN [28] marginally outperformed the other model. The vegetation segmentation value of wheat dataset employing DBPN [28] over the real-real experiment shows 100% improvement compared to bicubic interpolation, while LapSRN [26] achieved 91% improvement and SAN [99] failed to reproduce HR images. Consequently, we can conclude that DBPN [28] generally performs better across a range of images, but could be outperformed by a small margin under specific circumstances. Using iterative up and downsampling networks for agricultural drone super resolution tasks seems like the most promising approach, and should be the default choice, all other things being equal.

One limitation of our dataset is the imbalanced training set. There are almost triple the number of canola images compared to the total number of images in the lentil and wheat datasets. One solution could be collecting additional data from the field. However, collecting such data can be expensive. Data augmentation dealing with different random transformations of the images to increase the samples in a dataset and improve variances, is a well-known technique in deep learning. Although, we employed general augmentation techniques, such as data flipping, or rotating in our case, more sophisticated techniques, such as instance-based augmentation and fusion-based augmentation [100] could prove beneficial. Oversampling that resamples less frequent data to balance their amount in comparison with dominant data is another technique with an imbalanced dataset. In [101], a mathematical characterization of image representation was proposed to oversample data for image super resolution by learning the dynamic of information within the image.

There are other possible solutions, such as synthetic minority oversampling technique (SMOTE) [102] or generative adversarial network (GAN) [103] to address oversampling minority samples in training set prior to fitting a model. These techniques are common in image classification [104], but they might be feasible for a minority of sub-samples, such as different crop growth stages or different genotypes. Of particular interest for future work is to explore oversampling techniques on our dataset to produce highly variant sub-samples in the training set.

6. Conclusions

In this work, we have demonstrated the efficacy of image super resolution on real-world aerial image data for more accurate image-based plant phenotyping. We collected raw sensor data via two separate sensors co-mounted on a UAV. Raw images were processed through a three-step automated image pre-processing pipeline to create a dataset of paired and aligned LR and HR images from fields for three crops (canola, wheat, lentil). We conducted three experiments that systematically trained and tested deep learning-based image super resolution models on different combinations of synthetic and real datasets. We performed extensive quantitative and qualitative assessments considering standard
and domain-specific measures to evaluate the image quality. Our main finding is that the models trained on synthetic downsampled HR agricultural images cannot reproduce plant phenotypes accurately when tested on real-world LR images. On the other hand, the models trained and tested with the real-world dataset achieved higher-fidelity results which, in turn, facilitated better performance in the analysis of aerial images for plant organ and vegetation segmentation. Our work indicates that the current state-of-the-art super resolution models, when trained with paired LR/HR aerial images, have potential for improving the accuracy of image-based plant phenotyping.

**Author Contributions:** Conceptualization, M.A., K.G.S. and I.S.; methodology, M.A.; data collection, S.V., S.S., H.D.; data curation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was conducted thanks in part to funding from the Canada First Research Excellence Fund.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The curated datasets for this study are available upon request.

**Conflicts of Interest:** On behalf of all authors, the corresponding author states that there is no direct or indirect commercial or financial interest in the project that can be construed as a potential of conflicts of interest.

**References**

1. Fröna, D.; Szenderák, J.; Harangi-Rákos, M. The challenge of feeding the world. *Sustainability* 2019, **11**, 5816. [CrossRef]
2. Giovannucci, D.; Scherr, S.J.; Nierenberg, D.; Hebebrand, C.; Shapiro, J.; Milder, J.; Wheeler, K. Food and Agriculture: The future of sustainability. In *The Sustainable Development in the 21st Century (SD21) Report for Rio*; United Nations: New York, NY, USA, 2012; Volume 20.
3. Rising, J.; Devineni, N. Crop switching reduces agricultural losses from climate change in the United States by half under RCP 8.5. *Nat. Commun.* 2020, **11**, 1–7. [CrossRef]
4. Snowdon, R.J.; Wittkop, B.; Chen, T.W.; Stahl, A. Crop adaptation to climate change as a consequence of long-term breeding. *Theor. Appl. Genet.* 2020, 1–11. [CrossRef]
5. Furbank, R.T.; Sirault, X.R.; Stone, E.; Zeigler, R. Plant phenome to genome: A big data challenge. In *Sustaining Global Food Security: The Nexus of Science and Policy*; CSIRO Publishing: Collingwood, VIC, Australia, 2019; p. 203.
6. Furbank, R.T.; Tester, M. Phenomics—technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 2011, **16**, 635–644. [CrossRef] [PubMed]
7. Alvarez-Mendoza, C.I.; Teodoro, A.; Quintana, J.; Tituana, K. Estimation of Nitrogen in the Soil of Balsa Trees in Ecuador Using Unmanned Aerial Vehicles. In Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 26 September–2 October 2020; pp. 4610–4613.
8. Huang, J.; Tichit, M.; Poulot, M.; Darly, S.; Li, S.; Petit, C.; Aubry, C. Comparative review of multifunctionality and ecosystem services in sustainable agriculture. *J. Environ. Manag.* 2015, **149**, 138–147. [CrossRef]
9. Sylvestre, G.; Rambaldi, G.; Guerin, D.; Wisniewski, A.; Khan, N.; Veale, J.; Xiao, M. *E-Agriculture in Action: Drones for Agriculture*; Food and Agriculture Organization of the United Nations and International Telecommunication Union: Bangkok, Thailand, 2018; Volume 19, p. 70.
10. Catuogli, L.; Corniglia, M.; Gaetani, M.; Grossi, N.; Magni, S.; Migliazzi, M.; Angelini, L.; Mazzoncini, M.; Silvestri, N.; Fontanelli, M.; et al. Unmanned aerial vehicle to estimate nitrogen status of turfgrasses. *PLoS ONE* 2016, **11**, e0158268.
11. Wang, H.; Mortensen, A.K.; Mao, P.; Boelt, B.; Gislum, R. Estimating the nitrogen nutrition index in grass seed crops using a UAV-mounted multispectral camera. *Int. J. Remote Sens.* 2019, **40**, 2467–2482. [CrossRef]
12. Corrigan, F. Multispectral Imaging Camera Drones in Farming Yield Big Benefits. *Drone Zone*—Drone Technology. 2019. Available online: https://www.dronezon.com/learn-about-drones-quadcopters/multispectral-sensor-drones-in-farming-yield-big-benefits/ (accessed on 9 June 2021).
16. Xue, J.; Su, B. Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* 2017, 2017. [CrossRef]

17. Farjon, G.; Krikeb, O.; Hillel, A.B.; Alchanatis, V. Detection and counting of flowers on apple trees for better chemical thinning decisions. *Precis. Agric.* 2019, 21, 503–521. [CrossRef]

18. Deery, D.M.; Rebetzk, G.; Jimenez-Berni, J.A.; Bovill, B.; James, R.A.; Condon, A.G.; Furbank, R.; Chapman, S.; Fischer, R. Evaluation of the phenotypic repeatability of canopy temperature in wheat using continuous-terrestrial and airborne measurements. *Front. Plant Sci.* 2019, 10, 875. [CrossRef]

19. Madec, S.; Baret, F.; De Solan, B.; Thomas, S.; Dutartre, D.; Jezquel, S.; Hemmerlé, M.; Colombneau, G.; Comar, A. High-throughput phenotyping of plant height: Comparing unmanned aerial vehicles and ground LiDAR estimates. *Front. Plant Sci.* 2017, 8, 2002. [CrossRef]

20. Hu, P.; Guo, W.; Chapman, S.C.; Guo, Y.; Zheng, B. Pixel size of aerial imagery constrains the applications of unmanned aerial vehicle in crop breeding. *ISPRS J. Photogramm. Remote Sens.* 2019, 154, 1–9. [CrossRef]

21. Liu, M.; Yu, T.; Gu, X.; Sun, Z.; Yang, J.; Zhang, Z.; Mi, X.; Cao, W.; Li, J. The impact of spatial resolution on the classification of vegetation types in highly fragmented planting areas based on unmanned aerial vehicle hyperspectral images. *Remote Sens.* 2020, 12, 146. [CrossRef]

22. Dong, Z.; Wang, M.; Wang, Y.; Zhu, Y.; Zhang, Z. Object detection in high resolution remote sensing imagery based on convolutional neural networks with suitable object scale features. *IEEE Trans. Geosci. Remote Sens.* 2019, 58, 2104–2114. [CrossRef]

23. Taira, H.; Chong, K.T. Object detection in very high-resolution aerial images using one-stage densely connected feature pyramid network. *Sensors* 2018, 18, 3341. [CrossRef]

24. Seifert, E.; Seifert, S.; Vogt, H.; Drew, D.; Van Aardt, J.; Kunneke, A.; Seifert, T. Influence of drone altitude, image overlap, and optical sensor resolution on multi-view reconstruction of forest images. *Remote Sens.* 2019, 11, 1252. [CrossRef]

25. Garrigues, S.; Allard, D.; Baret, F.; Weiss, M. Quantifying spatial heterogeneity at the landscape scale using variogram models. *Remote Sens. Environ.* 2006, 103, 81–96. [CrossRef]

26. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.

27. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.T.; Zhang, L. Second-order attention network for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11065–11074.

28. Haris, M.; Shakhranovich, G.; Ukita, N. Deep back-projection networks for super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1664–1673.

29. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.

30. 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, ICCV 2001, Vancouver, BC, Canada, 7–14 July 2001; Volume 2, pp. 416–423.

31. Liu, G.; Wei, J.; Zhu, Y.; Wei, Y. Super-Resolution Based on Residual Dense Network for Agricultural Image. *J. Phys. Conf. Ser.* 2019, 1345, 022012. [CrossRef]

32. Cap, Q.H.; Tani, H.; Uga, H.; Kagiwada, S.; Iyamotomi, H. LASSR: Effective Super-Resolution Method for Plant Disease Diagnosis. *arXiv* 2020, arXiv:2010.06499.

33. 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]

34. Zhang, X.; Chen, Q.; Ng, R.; Kolun, V. Zoom to learn, learn to zoom. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3762–3770.

35. Cai, J.; Zeng, H.; Yang, H.; Cao, Z.; Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3086–3095.

36. NTIRE Contributors. NTIRE 2020 Real World Super-Resolution Challenge. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Seattle, WA, USA, 14–19 June 2020.

37. Park, S.C.; Park, M.K.; Kang, M.G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* 2003, 20, 21–36. [CrossRef]

38. Zhang, B.; Sun, X.; Gao, L.; Yang, L. Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm. *IEEE Trans. Geosci. Remote Sens.* 2011, 49, 2635–2646. [CrossRef]

39. Liu, J.; Luo, B.; Douté, S.; Chanussot, J. Exploration of planetary hyperspectral images with unsupervised spectral unmixing: A case study of planet Mars. *Remote Sens.* 2018, 10, 737. [CrossRef]

40. Wang, P.; Zhang, G.; Hao, S.; Wang, L. Improving remote sensing image super-resolution mapping based on the spatial attraction model by utilizing the pansharpening technique. *Remote Sens.* 2019, 11, 247. [CrossRef]

41. Suresh, M.; Jain, K. Subpixel level mapping of remotely sensed image using colorimetry. *Egypt. J. Remote Sens. Space Sci.* 2018, 21, 65–72. [CrossRef]

42. Siu, W.C.; Hung, K.W. Review of image interpolation and super-resolution. In Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference, Hollywood, CA, USA, 3–6 December 2012; pp. 1–10.
43. Hardeep, P.; Swadas, P.B.; Joshi, M. A survey on techniques and challenges in image super resolution reconstruction. *Int. J. Comput. Sci. Mob. Comput.* 2013, 2, 317–325.

44. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* 2002, 22, 56–65. [CrossRef]

45. Salvador, J. Example-Based Super Resolution; Academic Press: Cambridge, MA, USA, 2016.

46. Guo, K.; Yang, X.; Lin, W.; Zhang, R.; Yu, S. Learning-based super-resolution method with a combining of both global and local constraints. *IET Image Process.* 2012, 6, 337–344. [CrossRef]

47. Shao, W.Z.; Elad, M. Simple, accurate, and robust nonparametric blind super-resolution. In *International Conference on Image and Graphics*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 333–348.

48. Wang, Z.; Yang, Y.; Wang, Z.; Chang, S.; Yang, J.; Huang, T.S. Learning super-resolution jointly from external and internal examples. *IEEE Trans. Image Process.* 2015, 24, 4359–4371. [CrossRef]

49. Bevilacqua, M.; Roumy, A.; Guillemtot, C.; Alberi-Morel, M.L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the 23rd British Machine Vision Conference (BMVC), Surrey, UK, 3–7 September 2012.

50. Freedman, G.; Fattal, R. Image and Video Upscaling from Local Self-Examples. *ACM Trans. Graph.* 2010, 28, 1–10. [CrossRef]

51. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 349–356.

52. Lee, H.; Battle, A.; Raina, R.; Ng, A.Y. Efficient sparse coding algorithms. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 801–808.

53. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* 2010, 19, 2861–2873. [CrossRef]

54. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008.

55. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In *International Conference on Curves and Surfaces*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 711–730. [CrossRef]

56. 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]

57. Kim, J.; Kwon Lee, J.; Mu Lee, K. 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, 27–30 June 2016; pp. 1646–1654.

58. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3147–3155.

59. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 391–407. [CrossRef]

60. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.

61. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. 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.

62. 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 IEEE European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301. [CrossRef]

63. Gu, J.; Lu, H.; Zuo, W.; Dong, C. Blind super-resolution with iterative kernel correction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1604–1613. [CrossRef]

64. Michaelli, T.; Irani, M. Nonparametric blind super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 945–952. [CrossRef]

65. He, H.; Siu, W.C. Single Image super-resolution using Gaussian process regression. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 449–456. [CrossRef]

66. He, Y.; Yap, K.H.; Chen, L.; Chau, L.P. A soft MAP framework for blind super-resolution image reconstruction. *Image Vis. Comput.* 2009, 27, 364–373. [CrossRef]

67. Wang, Q.; Tang, X.; Shum, H. Patch based blind image super resolution. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05), Beijing, China, 17–21 October 2005; Volume 1, pp. 709–716.

68. Begin, I.; Ferrie, F. Blind super-resolution using a learning-based approach. In Proceedings of the 17th International Conference on Pattern Recognition, ICFPR 2004, Cambridge, UK, 26 August 2004; Volume 2, pp. 85–89. [CrossRef]

69. Joshi, N.; Szeliski, R.; Kriegman, D.J. PSF estimation using sharp edge prediction. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8. [CrossRef]

70. Hussein, S.A.; Tirer, T.; Giryes, R. Correction Filter for Single Image Super-Resolution: Robustifying Off-the-Shelf Deep Super-Resolvers. In Proceedings of the IEEE/ICVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1428–1437. [CrossRef]

71. Bell-Kligler, S.; Shocher, A.; Irani, M. Blind super-resolution kernel estimation using an internal-gan. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 284–293. [CrossRef]

72. Zhou, R.; Sussstrunk, S. Kernel modeling super-resolution on real low-resolution images. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 2433–2443. [CrossRef]
102. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 2002, 16, 321–357. [CrossRef]

103. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* 2014, arXiv:1406.2661.

104. Feng, W.; Huang, W.; Bao, W. Imbalanced hyperspectral image classification with an adaptive ensemble method based on SMOTE and rotation forest with differentiated sampling rates. *IEEE Geosci. Remote Sens. Lett.* 2019, 16, 1879–1883. [CrossRef]