

## Article

# RMSRGAN: A Real Multispectral Imagery Super-Resolution Reconstruction for Enhancing Ginkgo Biloba Yield Prediction

Kaixuan Fan <sup>1</sup>, Min Hu <sup>1</sup>, Maocheng Zhao <sup>1,\*</sup>, Liang Qi <sup>1,2</sup>, Weijun Xie <sup>1</sup>, Hongyan Zou <sup>1</sup>, Bin Wu <sup>1</sup>, Shuaishuai Zhao <sup>1</sup> and Xiwei Wang <sup>1</sup>

<sup>1</sup> College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China; shaunvan@njfu.edu.cn (K.F.)

<sup>2</sup> Jinpu Research Institute, Nanjing Forestry University, Nanjing 210037, China

\* Correspondence: mczhao@njfu.edu.cn

**Abstract:** Multispectral remote sensing data with abundant spectral information can be used to compute vegetation indices to improve the accuracy of Ginkgo biloba yield prediction. The limited spatial resolution of multispectral cameras restricts the detail capture over wide farmland, but super-resolution (SR) reconstruction methods can enhance image quality. However, most existing SR models have been trained on images processed from downsampled high-resolution (HR) images, making them less effective in reconstructing real low-resolution (LR) images. This study proposes a GAN-based super-resolution reconstruction method (RMSRGAN) for multispectral remote sensing images of Ginkgo biloba trees in real scenes. A U-Net-based network is employed instead of the traditional discriminator. Convolutional block attention modules (CBAMs) are incorporated into the Residual-in-Residual Dense Blocks (RRDBs) of the generator and the U-Net of the discriminator to preserve image details and texture features. An unmanned aerial vehicle (UAV) equipped with a multispectral camera was employed to capture field multispectral remote sensing images of Ginkgo biloba trees at different spatial resolutions. Four matching HR and LR datasets were created from these images to train RMSRGAN. The proposed model outperforms the traditional models by achieving superior results in both quantitative evaluation metrics (peak signal-to-noise ratio (PSNR) is 32.490, 31.085, 27.084, 26.819, and structural similarity index (SSIM) is 0.894, 0.881, 0.832, 0.818, respectively) and qualitative evaluation visualization. Furthermore, the efficiency of our proposed method was tested by generating individual vegetation indices (VIs) from images taken before and after reconstruction to predict the yield of Ginkgo biloba. The results show that the SR images exhibit better  $R^2$  and RMSE values than LR images. These findings show that RMSRGAN can improve the spatial resolution of real multispectral images, increasing the accuracy of Ginkgo biloba yield prediction and providing more effective and accurate data support for crop management.

**Keywords:** Ginkgo biloba; remote sensing; unmanned aerial vehicle; super-resolution reconstruction; generative adversarial network; yield prediction



**Citation:** Fan, K.; Hu, M.; Zhao, M.; Qi, L.; Xie, W.; Zou, H.; Wu, B.; Zhao, S.; Wang, X. RMSRGAN: A Real Multispectral Imagery Super-Resolution Reconstruction for Enhancing Ginkgo Biloba Yield Prediction. *Forests* **2024**, *15*, 859. <https://doi.org/10.3390/f15050859>

Academic Editor: Phillip G. Comeau

Received: 2 April 2024

Revised: 8 May 2024

Accepted: 12 May 2024

Published: 14 May 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

High-resolution (HR) remote sensing images, which contain abundant feature and texture information, are widely used in fields such as target detection [1,2], change detection [3,4], semantic segmentation [5,6] and land cover classification [7,8]. However, despite significant advancements in remote sensing technology over the past few decades, challenges still need to be overcome in acquiring HR imagery. Therefore, developing super-resolution (SR) reconstruction technology is crucial in this context. SR techniques improve the resolution of low-resolution (LR) remote sensing images through algorithmic processing, bringing them closer to or meeting HR standards. This technology relies on advanced image processing algorithms, such as deep learning [9,10], and requires extensive image analysis and pattern recognition research. SR technology can effectively enhance

the value of LR remote sensing resources, making them more useful in practical applications and supporting decision-making in critical areas such as surface monitoring and environmental assessment.

Recent HR remote sensing imagery advances have substantially transformed agricultural and forestry practices. Key studies in this domain demonstrate diverse applications and progressive capabilities. Wu et al. [11] explored wheat leaf area index (LAI) prediction using high-resolution unmanned aerial vehicle (UAV) imagery combined with multi-sensor data, underscoring the significance of soil background elimination and data fusion for enhanced accuracy. In contrast, Marzougui et al.'s [12] comparative analysis of satellite and unmanned aerial system (UAS) multispectral imagery for field pea yield estimation revealed the limited impact of texture features on model performance despite the benefits of multi-scale data fusion. Ramin et al. [13] highlighted the potential of century-old biochar in boosting chicory growth, as evidenced through UAV imagery that showed increased canopy cover and leaf length. Advancements in HR remote sensing imagery have greatly enhanced agricultural and forestry monitoring, despite challenges from sensor limitations and environmental factors. SR reconstruction technology offers a cost-effective solution to improve image quality and spatial resolution.

Traditional super-resolution (SR) methods, crucial for image enhancement, face limitations that have led to advanced developments. These methods include interpolation-based algorithms like bicubic [14], known for speed but lacking in detail accuracy, and reconstruction-based approaches like iterative back projection [15], providing better detail but suffering from high computational demands. These challenges prompted the shift towards AI-driven SR models for improved performance. Deep learning advancements have profoundly influenced the evolution of super-resolution (SR) techniques. Traditional single-image SR methods, dominant in remote sensing, have been surpassed by deep learning-based models for their enhanced performance and broader applicability. Pioneering this shift, Dong et al. [16] introduced the Super-Resolution Convolutional Neural Network (SRCNN), utilizing Convolutional Neural Network (CNN) technology for SR image reconstruction. Despite its superiority over traditional methods, SRCNN was constrained by limited image content adaptability and slow convergence speed. Addressing these limitations, Shi et al. [17] proposed the Enhanced Super-Resolution Convolutional Neural Network (ESPCN), implementing sub-pixel convolution layers to improve texture retention and training efficiency. Further progression in SR models included Very Deep Super-Resolution (VDSR) [18], which utilized a residual learning strategy to capture high-frequency details. SRCNN and the Fast Super-Resolution Convolutional Neural Network (FSRCNN) [19] emphasized the balance between processing speed and image quality. While CNN- and transformer-based methods excel in recovery, they occasionally produce overly smooth images. In contrast, generative adversarial network (GAN)-based SR methods, such as SRGAN [20], have proven adept at producing highly detailed textures, exploiting perceptual loss, and advanced discriminators for realism. ESRGAN [21] enhances the discriminator with the Relativistic Average GAN and incorporates the Residual-in-Residual Dense Block (RDB) for improved model performance. In addition, the SR field has made significant progress in modeling real-world degradation. The Real-ESRGAN [22] is notable for its approaches to mimicking complex degradations, thereby increasing the robustness and versatility of SR models under different imaging conditions. These GAN-based models refine image quality and address the texture fidelity and structural features of images, overcoming limitations such as twisted lines and background distortions.

In recent years, the super-resolution reconstruction of remote sensing pictures in agriculture and forestry has emerged as a significant study field. Liang et al. [23] proposed a system for early smoke identification in forest fire prevention. They combined a super-resolution network with smoke segmentation, resulting in a similarity coefficient of 0.742 when compared to high-resolution satellite pictures. Huang et al. [24] investigated using super-resolution reconstruction in UAV remote sensing photos for tree species classification to address challenges in dense forests and varying photography angles. Zhang et al. [25] in-

investigated using multispectral drone photography and generative adversarial networks for the super-resolution reconstruction of Chinese cabbage features, where SRGAN and U-Net models achieved a segmentation accuracy of 94.43% and an  $R^2$  above 0.78. Klapp et al. [26] suggest using low-cost thermal cameras and CNN-based super-resolution techniques to enhance agricultural remote sensing by improving image clarity and measurement accuracy. Zeng et al. [27] propose MALNet, a network enhancing agricultural pest image resolution by integrating residual and dense connections, boosting feature exploitation and computational efficiency for accurate pest identification. In general, advanced technologies like CNNs and GANs have demonstrated the potential to enhance the accuracy and clarity of remote sensing data in these specific areas.

Despite advancements in SR methods, challenges persist, particularly issues with regard to the degradation of remote sensing images by atmospheric interference, noise, and lens. Current SR models often struggle with real-world accuracy due to a lack of high-quality reference data. While GANs can produce detailed textures, they face training stability and visual accuracy issues. Most SR methods use pixel-level loss functions, which can overlook the complex features specific to remote sensing, underscoring the need for SR approaches that utilize comprehensive scene information and adapt well to diverse conditions. To tackle these concerns, we propose an upgraded GAN-based network for super-resolution reconstruction in real multispectral remote sensing images of Ginkgo biloba trees (RMSRGAN). The improved resolution achieved by our approach not only captures finer textural details, but also significantly improves the accuracy of yield predictions. This method provides a cost-effective alternative to expensive HR multispectral sensors, which are essential for accurate, large-scale remote sensing. The primary modifications and contributions encompass the following:

1. A dataset comprising multispectral remote sensing images of Ginkgo biloba trees is produced in real scenes. The dataset consists of HR and LR images captured in the field by a UAV equipped with a multispectral camera at different flight heights and matched individually.
2. The discriminator network is improved by using a U-Net structure. The discriminator's U-Net and the generator's Residual-in-Residual Dense Blocks (RRDBs) both use convolutional block attention modules (CBAMs). These strategies enable GAN-based networks to improve super-resolution reconstruction by focusing on image texture and detail.
3. To further validate the effectiveness of the super-resolution reconstruction model proposed in this study, this study conducted yield prediction using vegetation indices on the reconstructed images. In the evaluation of the prediction results, the reconstructed images surpass the original images.

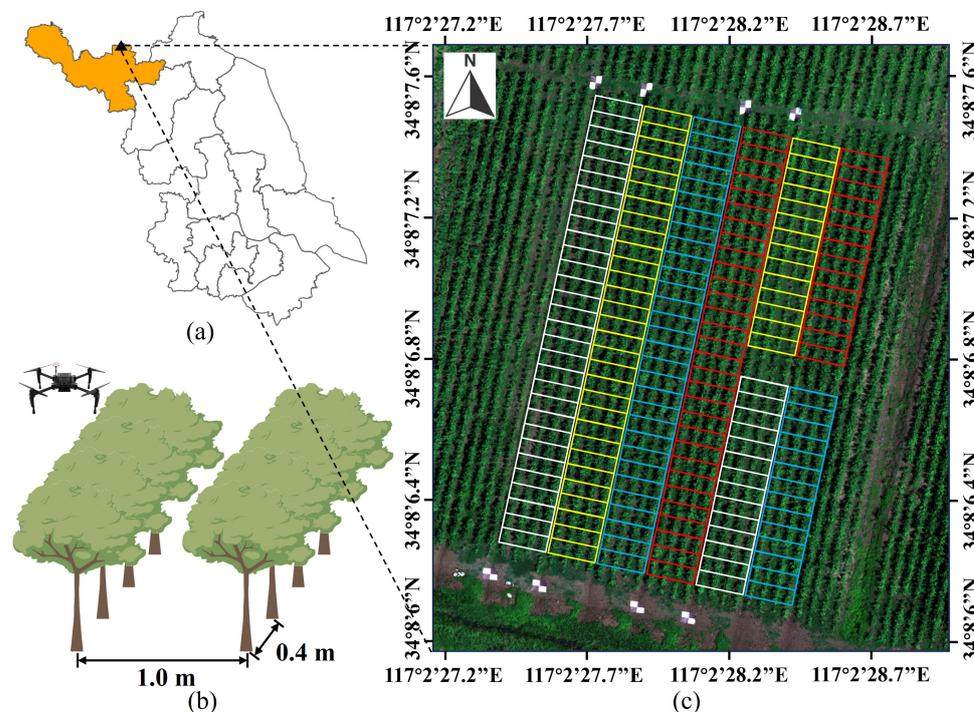
The subsequent sections of this work are structured in the following manner: Section 2 outlines the experimental design, data sources, super-resolution reconstruction models, yield forecast methodologies, and model evaluation criteria. Section 3 presents the experimental findings. Section 4 examines the effectiveness and constraints of RMSRGAN. Section 5 summarizes the contributions of this study and proposes future research areas.

## 2. Materials and Methods

### 2.1. Field Experimental Design

The field experiment was conducted at the Ginkgo biloba planting base of Xuzhou Changrong Agricultural Development Co. in Pizhou, Xuzhou City, Jiangsu Province, China (34°8'42" N, 117°2'45" E), as shown in Figure 1. The investigation was carried out on three-year-old Ginkgo biloba trees planted in February 2020. The trees were planted with a row spacing of 1 m and a column spacing of 0.4 m (Figure 1). The experiment consisted of 176 plots, each measuring  $0.8 \times 4$  m<sup>2</sup>, with eight trees in each plot. Four levels of nitrogen (N) fertilizers (N0–N3: 0, 100, 200 and 300 kg/ha) were applied to the plots. N fertilization was applied in mid-April, early-May and late-May at rates of 40%, 30% and 30%, respectively. Additionally, the same amount of phosphate (P, 120 kg/ha) and potash

(K, 200 kg/ha) were treated in all plots. Field images were acquired between 10:00 and 14:00 local time on 2 August 2023. Field assessments, destructive sampling and UAV flight missions were conducted in clear weather circumstances. Weed and insect pests in the field were controlled manually, and no agricultural chemicals were used.



**Figure 1.** (a) Experimental location of Pizhou, Xuzhou City, Jiangsu Province. (b) Row and column spacing for Ginkgo biloba trees. (c) Experimental plot layout. The white, yellow, blue and red boxes represent the amount of nitrogen fertilizer applied as N0 (0), N1 (100 kg/ha), N2 (200 kg/ha) and N3 (300 kg/ha), respectively.

## 2.2. Data Acquisition and Processing

### 2.2.1. Multispectral Imagery Acquisition

The multispectral (MS) imagery data were acquired using a DJI M100 drone (SZ DJI Technology Co., Shenzhen, China) equipped with a Red-Edge M MS camera (MicaSense Inc., Seattle, WA, USA). The camera features a quintet of spectral lenses—blue (B), green (G), red (R), red-edge (E) and near-infrared (NIR)—and a sunlight sensor that dynamically adjusts to ambient lighting to improve the accuracy of MS images. Each lens captures images at a resolution of  $1280 \times 960$  pixels. Detailed camera information is shown in Table 1. To ensure accurate reflectance data from the digital number (DN) values, radiometric calibration was performed with a standard board before and after each flight session.

**Table 1.** Spectral bands and resolutions of the multispectral camera.

Band	Wavelength	Bandwidth	Image Resolution
Blue (B)	475 nm	20 nm	$1280 \times 960$
Green (G)	560 nm	20 nm	$1280 \times 960$
Red (R)	668 nm	10 nm	$1280 \times 960$
Red-edge (E)	717 nm	10 nm	$1280 \times 960$
Near-infrared (NIR)	842 nm	40 nm	$1280 \times 960$

Utilizing the DJI GS Pro software 2.0.18, users were able to create custom missions and select flight paths for automated flight control. Flight heights were set at 15 m (1×), 30 m (2×) and 60 m (4×) to obtain MS remote sensing imagery at different resolutions. Flight

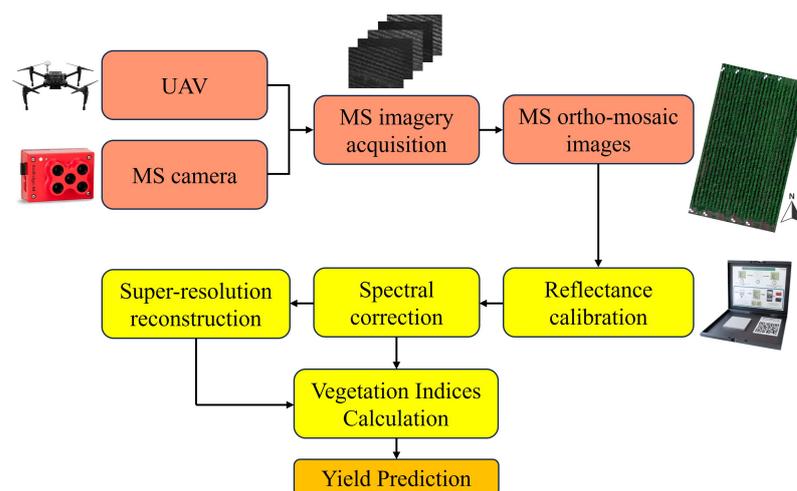
height and speed were taken into account to obtain high-quality images, which resulted in a forward overlap of 85% at 15 m and 90% at higher heights, while maintaining a side overlap of 80%. Table 2 provides detailed flight information. In addition, the aircraft and camera were equipped with a built-in global navigation satellite system (GNSS) module to track the location of the imagery accurately.

**Table 2.** Flight settings at different heights.

Height	Ground Resolution	Speed	Shutter Interval	Forward Overlap	Side Overlap
15 m (1×)	1.0 cm/pixel	1.5 m/s	0.1s	85%	80%
30 m (2×)	2.0 cm/pixel	3.0 m/s	0.1s	90%	80%
60 m (4×)	4.0 cm/pixel	6.0 m/s	0.1s	90%	80%

### 2.2.2. Image Processing

The image processing starts with pre-processing MS camera images in Pix4Dmapper 4.5.6 (Pix4D, Lausanne, Switzerland) to generate ortho-mosaic images and perform radiometric calibration, as shown in Figure 2. The process involves geolocating photos, importing ground control points (GCPs), aligning images, creating dense point clouds and ortho-mosaic images, and calibrating radiometric images. The center points of black and white checkered boards from Figure 1c were used as GCPs, which provide high contrast for precise image alignment. These GCPs are captured with high-precision GPS equipment, ensuring an accuracy of  $\pm 2$  cm in both horizontal and vertical dimensions. The image alignment is performed using the registration tool in the ENVI 5.6 software. Dense point clouds are created by the structure-from-motion (SfM) technique to enhance the photogrammetric process in Pix4Dmapper. Radiometric calibration is used on the MS images to improve the radiometric accuracy of the data and convert the digital number (DN) values to reflectance by comparing them to images with established reflectance values for calibration. Spectral correction further refines this process by adjusting the wavelength calibration and ensuring accurate spectral response. Furthermore, three widely used vegetation indices (VIs) closely associated with crop features were employed in this study for predicting crop production (Table 3).



**Figure 2.** Processing workflow of UAV-based multispectral images. MS represents the multispectral, digital surface model and digital terrain model, respectively.

**Table 3.** Details of selected vegetation indices. NIR, R and B represent the near-infrared, red and blue bands, respectively.

Name	Formula	Reference
Normalized difference vegetation index	$NDVI = \frac{NIR - R}{NIR + R}$	[28]
Ratio vegetation index	$RVI = NIR/R$	[29]
Enhanced vegetation index	$EVI = \frac{2.5 \times (NIR - R)}{NIR + 6 \times R - 7.5 \times B + 1}$	[30]

Plant height extraction involves utilizing ENVI 5.6 software to create a 3D point cloud model from MS images captured on 2 August 2023, to generate the digital surface models (DSMs) and digital terrain models (DTMs) of ground objects. The models are imported into ENVI, and plant surface models are created by subtracting the DTM from the DSM. The plant height obtained from UAV data is verified by comparing it with plant heights measured in the field.

### 2.2.3. Real-MSG Dataset

This study refrained from using the conventional approach of generating LR images through a bicubic interpolation of HR images. This process can result in inconsistencies between artificially created and genuine LR images, which impede the effectiveness of SR models. Our datasets utilized LR and HR aerial photography to train a model that can learn detailed mapping relationships, enhancing the accuracy of SR reconstruction in the real world compared to bicubic interpolation. Therefore, we took UAV-based MS remote sensing images which were acquired at 3 different heights at the same location on 2 August 2023. Detailed information is shown in Table 2. We used 15 m (1×) high images as HR and 30 m (2×) and 60 m (4×) high images as LR, where “1×”, “2×” and “4×” denote the factor by which the image resolution is increased, with 1× representing no change, 2× doubling and 4× quadrupling the original dimensions. By utilizing the registration tool in the ENVI software, the various resolution images were aligned by matching the preset GCPs to ensure that the characteristics in the processed pairs of HR and LR images match as closely as feasible. To simplify the training process, we integrated the single-channel grayscale images of five bands into three-channel color images of RGB (consisting of R, G and B bands) and NER (consisting of NIR, E and R bands). Then, the HR (1×) images were cropped to images measuring 256 × 256 using OpenCV, and the corresponding LR (2×) and LR (4×) images were cropped to dimensions of 128 × 128 and 64 × 64, respectively. Subsequently, we obtained a total of 10,000 image pairs and four datasets, including 2500 pairs of HR (1×)-LR (2×) RGB, 2500 pairs of HR (1×)-LR (4×) RGB, 2500 pairs of HR (1×)-LR (2×) NER and 2500 pairs of HR (1×)-LR (4×) NER. We named the datasets RGB (2×), RGB (4×), NER (2×) and NER (4×) successively and together referred to them as real MS images of Ginkgo biloba trees (Real-MSG). We partitioned the data into training, validation and testing sets with a ratio of 8:1:1 and implemented 5-fold cross-validation.

### 2.2.4. Leaf Yield Measurement

The leaf yield (kg) of each Ginkgo biloba tree was harvested individually by manual means at the maturity stage (3 August 2023) and counted in pre-determined plots. Harvested leaves were weighed for fresh weight using an electronic balance. The harvested leaves were dried at 105 °C for 0.5 h and then at 70 °C until constant weight. Subsequently, the dry weight was weighed and recorded. Finally, the dry-to-fresh weight ratio was found to be 1:3 by comparison.

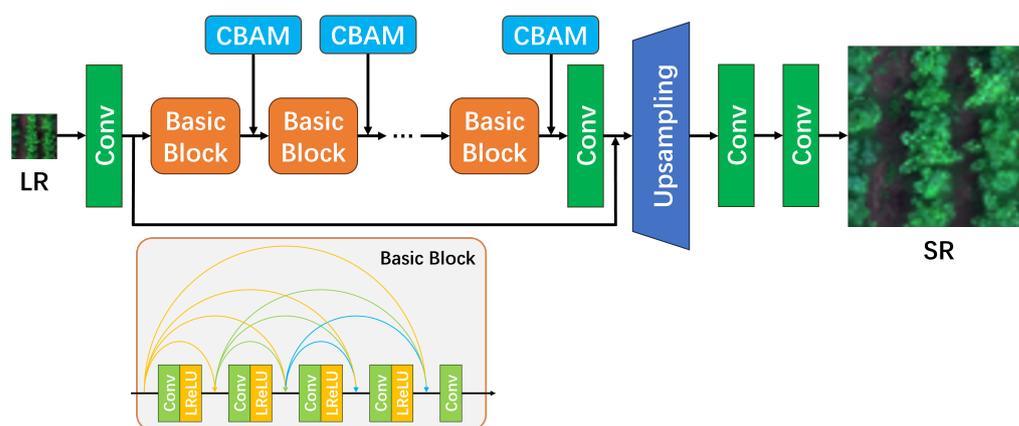
## 2.3. Super-Resolution Reconstruction Model

There are two primary components of RMSRGAN: a discriminator and a generator. Using computations like convolution and upsampling in its network structure, the gen-

erator uses the super-resolution reconstruction method to take the input LR image and produce the SR image. The discriminator then establishes whether or not the SR image is genuine by comparing it to the actual HR image. The discriminator aims to accurately discriminate between created SR photos and real HR images, whereas the generator aims to create SR images convincing enough to trick the discriminator. Through an adversarial process of continual training, the discriminator and generator learn from each other and enhance their performance repeatedly. This iterative training process makes the SR images more realistic and detailed.

### 2.3.1. Generator Network

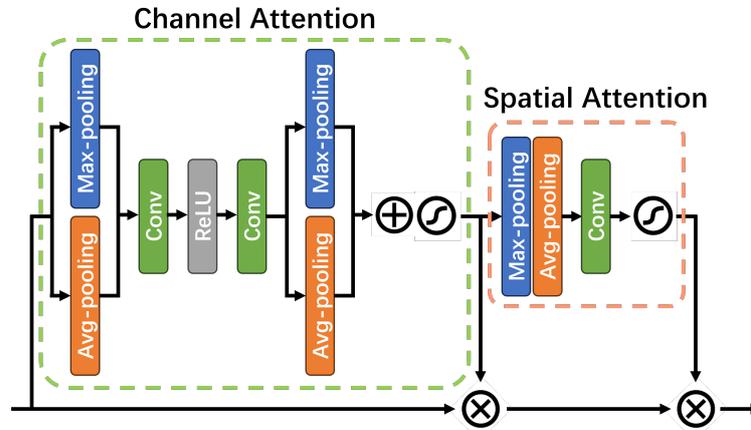
The generator network architecture is shown in Figure 3. First, a  $3 \times 3$  convolutional layer is used to extract the original features in the LR image to obtain the original feature map. Subsequently, the 23 basic block is used to retain the existing image features and discover new ones. The architecture of the basic block is composed of interconnected Residual-in-Residual Dense Block (RRDB), augmented by a convolutional layer with residual connections in the last. The RRDB framework is adept at integrating the strengths of residual and dense networks, as delineated in Figure 3. Residual networks excel at identifying and learning the disparities, often minimal, between inputs and outputs, a characteristic indicated by the potential of most residuals to approach zero. Dense connections efficiently concatenate the feature maps from preceding layers, enriching the input for subsequent layers. While residual networks capitalize on feature reutilization, they are less inclined to unearth new features; dense networks, in contrast, are more explorative but can suffer from increased redundancy. The RRDB design unites these contrasting approaches, enabling the model to navigate complex data landscapes and discern intricate patterns more effectively. This fusion bolsters the model's accuracy and enhances its performance, as evidenced by [21], which demonstrates RRDB's superior handling of diverse and intricate data distributions. And then, the image size is enlarged by the upsampling module ( $2\times$  or  $4\times$ ). Finally, the enlarged image features are further learned after two  $3 \times 3$  convolutional layers. The SR image corresponding to the LR image is reconstructed using the above generator network structure.



**Figure 3.** Generator network structure of RMSRGAN. Conv, LReLU and CBAM represent the convolution layer, leaky rectified linear unit and convolutional block attention module, respectively. The size of the convolutional layer is  $3 \times 3$ , the stride is 1, and the number of filters is 64.

In addition, a convolutional block attention module (CBAM), as shown in Figure 4, is included following each basic block. Implementing CBAM following the RRDB module in GANs significantly enhances the model's performance. The capacity of CBAM is utilized in this combination to concentrate on more informative features by simultaneously employing spatial and channel attention [31]. This strategic integration enhances the efficiency of the generative model in collecting delicate features and textures, resulting in the creation of more realistic images. Furthermore, the combination of RRDB's dense connectivity

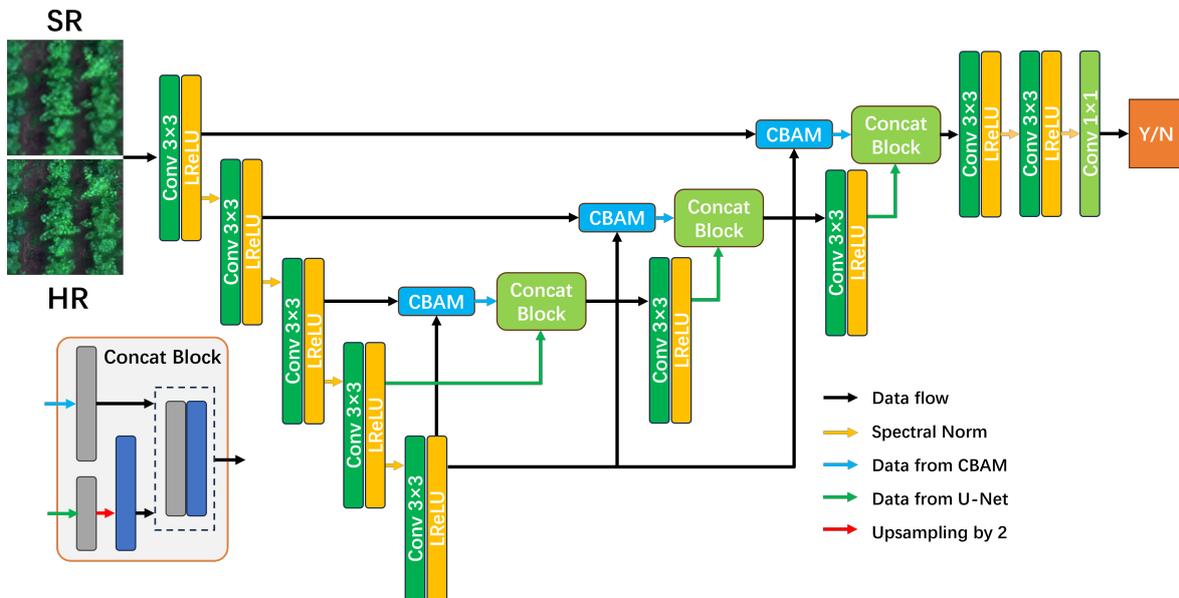
with CBAM’s attention mechanism enhances feature representation, resulting in high-quality output with improved perceptual clarity and detail fidelity [32]. This approach not only refines the visual quality of the generated images, but also contributes to the overall robustness and adaptability of the model for various image synthesis tasks.



**Figure 4.** Convolutional block attention module. ReLU, Max-pooling and Avg-pooling represent the leaky rectified linear unit, max-pooling layer and average-pooling layer, respectively. The size of the convolutional layer is  $3 \times 3$ , the stride is 1, and the number of filters is 64.

### 2.3.2. Discriminator Network

Inspired by the authors of [22], instead of using a traditional discriminator structure, in this study, we chose a discriminator network based on the U-Net structure, as shown in Figure 5. This discriminator network structure consists of a downsampling process (encoder) and an upsampling process (decoder). The encoder performs the extraction of the image features. In contrast, the decoder operates inverse convolution by upsampling, and the image features extracted during encoding are also added to the decoding process.



**Figure 5.** Discriminator network structure of RMSRGAN. Concat Block, Conv and LReLU represent the concatenate block, convolutional layer and leaky rectified linear unit, respectively. The size of the convolutional layer is  $3 \times 3$ , the stride is 1, and the number of filters is 64.

The modification of the U-Net structure for GAN discriminators has been a novel method, especially in the field of image generation. The U-Net design incorporates skip

connections to address gradient attenuation by combining early-stage and advanced features, enhancing the discriminator's ability to distinguish semantics at different scales. The architecture's focus on integrating information from various scales is essential for accurately representing small or localized details, which is necessary for creating images with high-quality structural and textural components. Throughout decoding, U-Net gradually merges features by utilizing deconvolution layers to improve the discriminator's understanding of context. This integration of characteristics significantly enhances the effectiveness of GANs in producing realistic images, highlighting the importance of U-Net in modern image generation applications.

Furthermore, spectral normalization regularization is used for the encoder after it moves from the initial convolution layer in this network design to improve the stability of the discriminator network training. Spectral normalization is a method used to control the Lipschitz constant of functions in deep learning models. It ensures stable training by limiting the spectral norm (the most significant singular value) of weight matrices, denoted as

$$\sigma(W) = \max \frac{\|Wv\|_2}{\|v\|_2} \quad (1)$$

where  $W$  is the weight matrix. Spectral normalization improves model generalization and reduces overfitting by normalizing weights, which leads to a smoother gradient flow and an increased model robustness. This technique adjusts weight matrix scales to reduce training instabilities and enhance the effectiveness of the generative adversarial network. It ensures more consistent training dynamics, resulting in an improved production of high-quality images by strengthening the stability of the discriminator's performance.

### 2.3.3. Loss Function

The generator network's loss function includes perceptual loss and pixel loss. The discriminator network utilizes the binary cross entropy loss function (BCELoss).

Perceptual loss ( $L_{Pct}$ ) [21] in GANs improves the visual quality of generated images by prioritizing perceptual similarity rather than pixel-wise precision. This method integrates content loss with adversarial loss by leveraging deep features extracted from pretrained networks such as VGG-19 to compare generated and target images comprehensively. The total perceptual loss is a combination of the content and adversarial losses. Perceptual loss merges content and style fidelity with human visual standards for realistic, high-quality image generation.

Pixel loss ( $L_{Pxl}$ ) measures the difference between created and target images at the pixel level, typically utilizing the  $L1$  or  $L2$  norm. The formula is as follows:

$$L_{Pxl} = \frac{1}{N} \sum |y - \hat{y}| \quad (2)$$

for  $L1$ , where  $y$  represents the target and  $\hat{y}$  represents the generated picture. This loss highlights the direct resemblance, improving image clarity and detail precision. The final objective function of the generator can be defined as

$$L_G = L_{Pct} + L_{Pxl} \quad (3)$$

BCELoss plays a crucial role in GAN training by effectively penalizing incorrect predictions and transforming them into probabilities using the sigmoid function. This enhances the numerical stability and effectiveness of the discriminator. This procedure directs the generator to generate images that are more authentic, thus enhancing the quality of image generation in the GAN. The effectiveness of a discriminator in GANs depends

on its ability to reliably differentiate genuine images from generated ones. The formula is expressed as

$$L_{BCE} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (4)$$

where  $n$  is the sample size,  $y_i$  indicates the actual label of each image (1 for real and 0 for fake), and  $\hat{y}_i$  reflects the discriminator's predicted probability that the  $i$ th image is real.

#### 2.3.4. Trainig Details

The study utilized experimental hardware consisting of a GeForce RTX 4090 GPU, an Intel Core i9-13900KF @ 3.00 GHz CPU for computational efficiency, and 64 GB of RAM. The code development utilized Python 3.9 as the programming language, PyCharm as the compiler, Pytorch framework for algorithm modeling and implementation, and Windows 10 as the operating system. CUDA11.2 and the related CuDNN were used to speed up model training. In addition, Table 4 provides further information regarding the training of the RMSRGAN model, where  $\beta_1$  and  $\beta_2$  represent the decay rates of the learning rate.

**Table 4.** Experimental details.

Item	RGB (2×)	RGB (4×)	NER (2×)	NER (4×)
Input size	128 × 128	64 × 64	128 × 128	64 × 64
Scaling factor	2	4	2	4
Batch size	4	8	4	8
Epoch	300	300	300	300
Learning rate	$1 \times 10^{-4}$	$1 \times 10^{-4}$	$1 \times 10^{-4}$	$1 \times 10^{-4}$
Optimizer	Adam	Adam	Adam	Adam
$\beta_1, \beta_2$	0.9, 0.999	0.9, 0.999	0.9, 0.999	0.9, 0.999

#### 2.4. Regression Method

In this study, the linear regression model was used to predict the yield of LR, SR and HR remote sensing images. This statistical method establishes a relationship between a single independent variable and a dependent variable by fitting a linear equation to the data. The linear equation is expressed as  $Y = a + bX$ , where  $Y$  represents the yield,  $X$  denotes the resolution of the VI,  $a$  is the y-intercept, and  $b$  is the slope. The slope indicates the expected change in yield for each unit increase in resolution, offering insights into the impact of image quality on agricultural output. This model's simplicity makes it effective for examining the direct influence of image resolution on yields, providing valuable information for enhancing agricultural productivity through optimized remote sensing applications.

#### 2.5. Model Performance Estimation

In assessing SR models, two key measures are typically used: peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). PSNR is a metric used to quantify the quality of reconstruction, denoted as

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX_I^2}{MSE} \right) \quad (5)$$

where  $MAX_I$  is the highest achievable pixel value in the picture, and  $MSE$  is the average squared difference between the original and reconstructed images. This statistic emphasizes the negative correlation between image error and quality. SSIM evaluates the visual quality

and similarity of two photos by assessing changes in structural information, brightness, and contrast. The formula is expressed as

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (6)$$

where  $x$  and  $y$  denote segments of the original and the reconstructed image. The formula computes similarity using the mean ( $\mu_x, \mu_y$ ), variance ( $\sigma_x^2, \sigma_y^2$ ) and covariance ( $\sigma_{xy}$ ) of the photos' brightness. Constants  $c_1$  and  $c_2$  are used to avoid division by zero when the denominator is close to zero. SSIM values vary from  $-1$  to  $1$ , with  $1$  representing complete similarity.

Two evaluation metrics commonly utilized for examining regression models are the coefficient of determination ( $R^2$ ) and root-mean-squared error ( $RMSE$ ).  $R^2$  measures the proportion of the variance in the dependent variable that can be explained by the independent variables, where values nearing  $1$  suggest a more robust model fit.  $RMSE$  evaluates the average size of the forecasting inaccuracies. They can be denoted as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

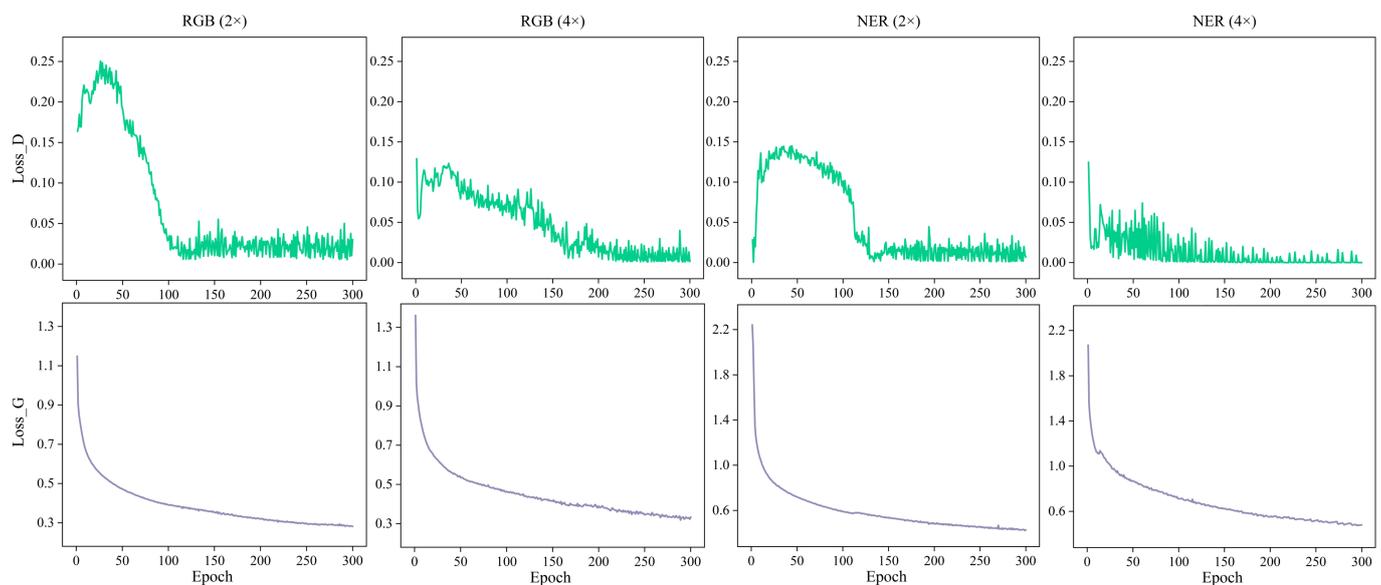
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

where  $y_i$  are actual values,  $\hat{y}_i$  are predicted values,  $\bar{y}$  is the mean of actual values and  $n$  is the number of data points. These measurements offer a thorough insight into a model's ability to predict accurately and the extent of its errors.

### 3. Results

#### 3.1. Loss Curves

The losses of the generator and the discriminator both tend to converge during the training procedure, as shown in Figure 6.



**Figure 6.** Variation curves of generator loss (Loss\_G) and discriminator loss (Loss\_D).

#### 3.2. Quantitative and Qualitative Metrics' Analysis of SR Images

This study conducted trials on the Real-MSG dataset using various models, including Bicubic, SRCNN [16], FSRCNN [19], VDSR [18], SRGAN [20], Real-ESRGAN [22] and RMSRGAN, ensuring that they were trained under the same software and hardware

environments and parameter settings. SRCNN, FSRCNN and VDSR are CNN-based SR models, whereas SRGAN, Real-ESRGAN and RMSRGAN are GAN-based SR models.

This study utilizes the PSNR and SSIM measurements for quantitative evaluation. The PSNR metrics for four datasets (Table 5) show that the double cubic interpolation had the lowest values and the RMSRGAN model (ours) had the highest values. The PSNR values of the three CNN-based models surpass those of the double-cubic interpolation approach. SRGAN has a more excellent PSNR value than the double-cubic interpolation approach, but a lower PSNR value than the CNN-based model among GAN-based models. The PSNR value of Real-ESRGAN is the second highest, only surpassed by our proposed approach. Similarly, as shown in Table 6, the same conclusion as Table 5 can be obtained from comparing SSIM metrics. Table 6 provides a comparable result to Table 5 regarding comparing SSIM measures. A higher SSIM value indicates a better preservation of properties, including brightness, contrast and structural similarity, during super-resolution reconstruction of the image. In addition, when comparing image datasets of the same type, the metrics for RGB (2×) are consistently more significant than those for RGB (4×); the metrics for NER (2×) and NER (4×) vary in either higher or lower, however. And when comparing image datasets of the same scale, the metrics for RGB (2× and 4×) are superior to those for NER (2× and 4×).

**Table 5.** Mean values of PSNR for different methods. Values in bold indicate the best performance.

Dataset	Bicubic	SRCNN	FSRCNN	VDSR	SRGAN	Real-ESRGAN	RMSRGAN
RGB (2×)	17.823	24.681	24.075	24.339	22.572	29.536	<b>32.490</b>
RGB (4×)	16.711	24.250	23.482	23.248	21.287	28.375	<b>31.085</b>
NER (2×)	15.257	20.589	19.608	20.258	17.229	24.166	<b>27.084</b>
NER (4×)	16.491	20.886	19.636	20.139	17.827	23.575	<b>26.819</b>

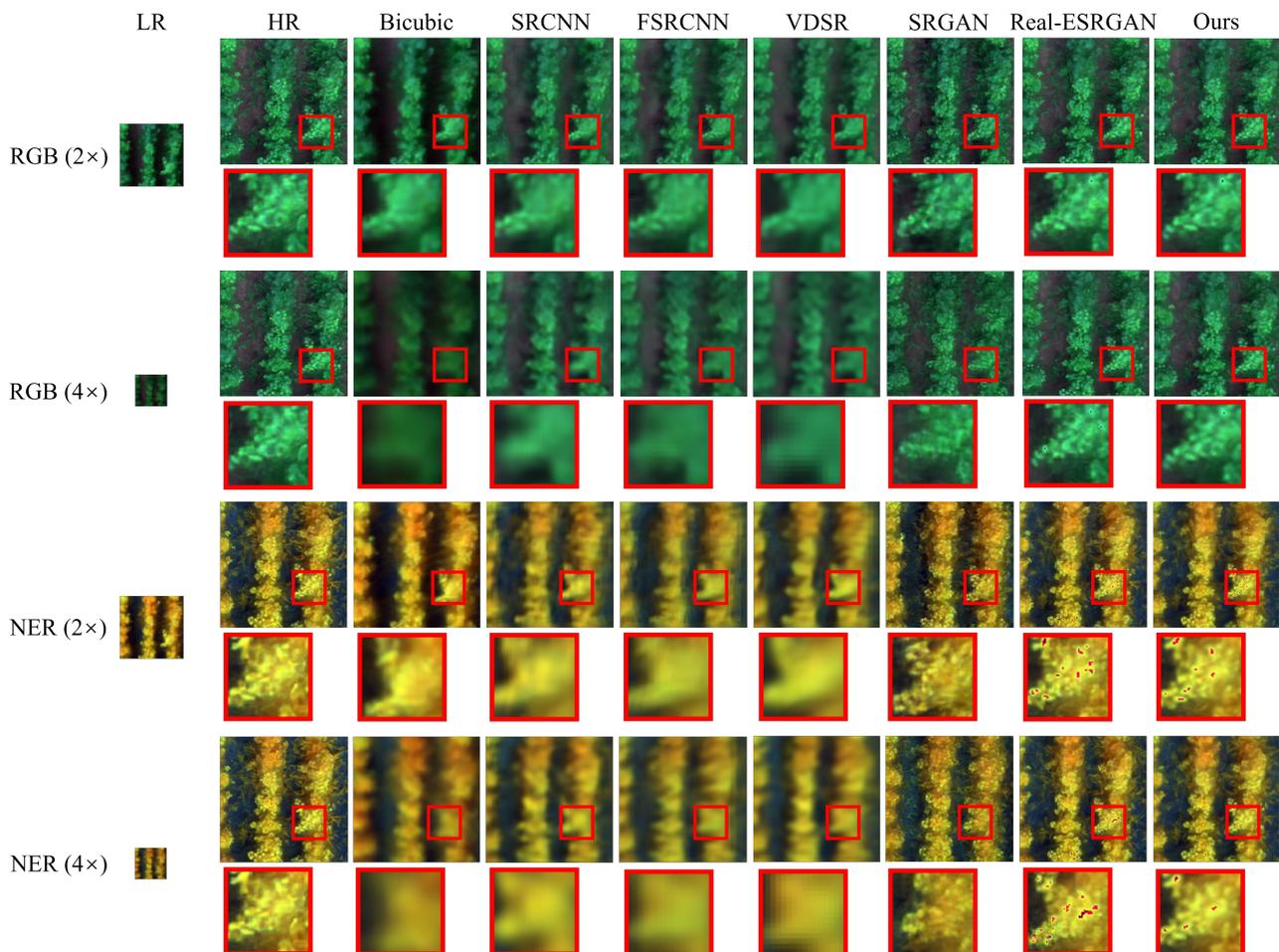
**Table 6.** Mean values of SSIM for different methods. Values in bold indicate the best performance.

Dataset	Bicubic	SRCNN	FSRCNN	VDSR	SRGAN	Real-ESRGAN	RMSRGAN
RGB (2×)	0.489	0.654	0.640	0.581	0.528	0.826	<b>0.908</b>
RGB (4×)	0.457	0.616	0.598	0.529	0.467	0.797	<b>0.884</b>
NER (2×)	0.362	0.533	0.510	0.524	0.360	0.740	<b>0.836</b>
NER (4×)	0.427	0.543	0.517	0.518	0.381	0.699	<b>0.841</b>

We conducted a qualitative assessment of the reconstructed images in addition to the quantitative evaluation, as depicted in Figure 7. The first column displays the LR image, the second column shows the HR image, and the subsequent columns exhibit the LR image rebuilt by each of the seven models. We have captured the same locations in these photos for zooming-in to aid in evaluating image details. These sites are marked by red boxes in Figure 7. The three CNN-based reconstruction models can recover more image details than the bicubic interpolation method on RGB images, but the contours are blurred; however, they cannot produce better results than the bicubic interpolation method regarding the reconstruction effect on NER images. The three GAN-based reconstruction models may significantly enhance the visual texture and features of the image during reconstruction. Real-ESRGAN and RMSRGAN (ours) can better recreate the texture aspects of images and produce more impressive visual effects than SRGAN. However, a small amount of unwanted artifacts appear in the reconstruction results of the former two models, especially in the brighter regions of the NER image.

RMSRGAN (ours) is determined to be the top performer based on a combination of quantitative and qualitative evaluation outcomes, excelling in super-resolution reconstruction metrics and image quality. This demonstrates our success in enhancing the model network's structure, which allows for improved super-resolution reconstruction with enhanced image texture and detail qualities. This also validates the authenticity and precision

of the real-based multispectral image dataset (Real-RMS) we developed. The Real-ESRGAN model achieved high metrics and image reconstruction quality, demonstrating the effectiveness of GAN-based reconstruction in handling the task of reconstructing real images. SRGAN's quantitative evaluation criteria are inferior to CNN-based reconstruction approaches, yet it surpasses them significantly in terms of picture reconstruction effectiveness. This indicates that the GAN-based reconstruction model may more effectively recover picture details, resulting in outcomes that closely resemble high-resolution photos.



**Figure 7.** Visual effects of reconstructed images by different methods.

### 3.3. Ablation Study

Three sets of ablation experiments were designed to evaluate the effectiveness of the enhancements made by each proposed component. The baseline model selected was from ESRGAN [21], utilizing RRDB modules in its generator network, and its discriminator network is structured based on a CNN. We incrementally incorporate CBAM and U-Net architectures into the baseline model. All models were trained with identical settings and assessed on a test dataset. Table 7 provides comparative data on quantitative indicators, showing a clear improvement in model performance throughout the trial.

**Table 7.** Ablation experiments results. Values in bold indicate the best performance. The symbol “+” indicates the incremental addition of the module.

Metric	Dataset	Baseline	+ CBAM in Generator	+ U-Net	+ CBAM in U-Net
PSNR	RGB (2×)	27.932	28.467	31.049	<b>32.490</b>
	RGB (4×)	27.103	27.976	30.164	<b>31.085</b>
	NER (2×)	23.532	24.164	26.261	<b>27.084</b>
	NER (4×)	22.850	23.093	26.011	<b>26.819</b>
SSIM	RGB (2×)	0.712	0.748	0.841	<b>0.894</b>
	RGB (4×)	0.684	0.727	0.833	<b>0.881</b>
	NER (2×)	0.673	0.701	0.801	<b>0.832</b>
	NER (4×)	0.646	0.688	0.781	<b>0.818</b>

Firstly, by incorporating CBAM after the RRDB module of the generator, the model can enhance important features and suppress unimportant ones, leading to improved image feature representation. This enhancement allows the model to respond more accurately to high-frequency details and texture information, ultimately enhancing quantitative evaluation metrics. Subsequently, the discriminator structure is optimized from a CNN-based structure to a U-Net-based one. The U-Net model efficiently captures an image’s intricate details and surrounding context by utilizing its encoder–decoder architecture along with skip connections. This allows the discriminator to better assess nuanced distinctions, such as texture and edges, in the generated image, enhancing the quality of super-resolution reconstruction. Furthermore, U-Net excels in preserving spatial information within images, a crucial aspect for accurately retrieving detailed information in high-resolution photos. Improvements were achieved in the PSNR and SSIM values. Finally, CBAM is incorporated into the U-Net architecture to increase the discriminator’s sensitivity to high-frequency details, improving the generator’s ability to restore high-quality high-frequency information more effectively. Moreover, the discriminator’s enhanced ability to detect anomalies in the created image allows the generator to minimize artifacts while learning, creating a more authentic and lifelike image. The final experiment produced the most favorable results when using quantitative evaluation indicators. The ablation experiments demonstrated that our suggested technique improved the key parameters, confirming the effectiveness of each optimization decision.

### 3.4. Estimation of Yield

In this section, we used VIs from MS remote sensing images to predict yield. We also researched their regression effects to make sure the super-resolution reconstruction algorithm works. In our previous research, we trained on the Real-MSG dataset to obtain the training weights for RMSRGAN. We still merged three single-channel images to create four remote sensing images: RGB (2× and 4×) and NER (2× and 4×). We divided the 2× and 4× images into 16 and 32 equally sized sub-images, respectively, and fed them into the trained RMSRGAN model for super-resolution reconstruction. The input and output sizes of the model were adjusted according to the input photographs. Finally, we obtained the multispectral image data after super-resolution reconstruction. Then, we calculated the VIs (Table 3) based on the MS images for flight heights of 15 m, 30 m and 60 m, respectively. In addition, VIs were calculated using data obtained after super-resolution reconstruction at 30 m and 60 m. To provide a comprehensive analysis, we performed linear regression analysis on all the multispectral image data, including both original and reconstructed images, to assess the effectiveness of our super-resolution model. For the yield estimation from these VIs, we utilized the linear regression model to calculate the  $R^2$  and  $RMSE$ , which are detailed in Table 8.

**Table 8.** Linear regression models using vegetation indices.

Metric	VI	HR	LR 2×	SR 2×	LR 4×	SR 4×
$R^2$	NDVI	0.653	0.598	0.638 ↑	0.543	0.587 ↑
	RVI	0.722	0.669	0.703 ↑	0.643	0.649 ↑
	EVI	0.565	0.534	0.543 ↑	0.502	0.530 ↑
RMSE	NDVI	0.414	0.441	0.423 ↓	0.470	0.451 ↓
	RVI	0.370	0.702	0.405 ↓	0.927	0.412 ↓
	EVI	0.463	0.476	0.474 ↓	0.490	0.481 ↓

HR, LR 2× and LR 4× represent the multispectral remote sensing images captured at flight heights of 15 m, 30 m and 60 m, respectively. SR 2× and SR 4× represent the multispectral remote sensing images taken at 30 m and 60 m flight heights after super-resolution reconstruction, respectively. ↑ and ↓ represent the improved metrics of SR compared to LR.

The  $R^2$  value indicates the goodness of fit of the model to the data. A high  $R^2$  value indicates that the model effectively accounts for the variability in the dependent variable. A lower RMSE value indicates reduced error and more accuracy in the model's predictions. For NDVI prediction results, the highest  $R^2$  value is HR ( $R^2 = 0.653$ ), the lowest is LR 4× ( $R^2 = 0.543$ ), and LR 2× is in between the two above ( $R^2 = 0.598$ ); the lowest RMSE value is HR (RMSE = 0.414 kg·m<sup>-2</sup>), and the lowest is LR 4× (RMSE = 0.470 kg·m<sup>-2</sup>), and LR 2× is in between the above two (RMSE = 0.441 kg·m<sup>-2</sup>). The above results are closely related to the flight height of the MS remote sensing imagery. The lower the flight height, the higher the ground resolution of the remote sensing images and the more accurate the preservation of image details. When comparing the super-resolution reconstruction results with the original pictures, the performance of SR 2× ( $R^2 = 0.638$ , RMSE = 0.423 kg·m<sup>-2</sup>) surpasses that of LR 2×. Similarly, the performance of SR 4× ( $R^2 = 0.587$ , RMSE = 0.451 kg·m<sup>-2</sup>) is superior to that of LR 4×. This result suggests that the MS image after RMSRGAN super-resolution reconstruction may restore more image information, resulting in improved performance in yield prediction. These conclusions are the same for both prediction outcomes based on RVI and EVI.

The results shown in Table 8 indicate significant enhancements in yield prediction from super-resolution (SR) reconstructed images compared to low-resolution (LR) images across various vegetation indices (VIs). Notably, for NDVI, RVI and EVI, the SR 2× images not only show improved  $R^2$  values but also exhibit reduced RMSE, highlighting a more precise yield prediction. For example, the NDVI in SR 2× achieves an  $R^2$  value of 0.638, compared to 0.598 in LR 2×, and reduces RMSE to 0.423 kg·m<sup>-2</sup> from 0.441 kg·m<sup>-2</sup>. This improvement suggests that SR techniques effectively restore critical image details, thus enhancing the reliability of remote sensing data for agricultural yield estimation.

#### 4. Discussion

Multispectral (MS) remote sensing imagery offers enhanced spectral resolution compared to visible light-based (RGB) remote sensing imagery. These supplementary bands can gather crucial data about the plant's physiological condition that is not discernible in RGB photos. Moreover, MS images enable the computation of many vegetation indices (NDVI, RVI and EVI in this study), which are strongly linked to biophysical factors like vegetation growth status, enhancing the precision of yield prediction. With the abundant and intricate features in multispectral remote sensing photos, many current algorithms need help to reconstruct these details reliably. Because high-resolution multispectral cameras are often prohibitively expensive and not widely available, there is a need to enhance the resolution of multispectral cameras through super-resolution algorithms for widespread use in precision agriculture research. To overcome this challenge, we propose a super-resolution reconstruction based on the generative adversarial network for real multispectral remote sensing images (RMSRGAN). RMSRGAN comprises a generator network and a discriminator network. The generator network utilizes the RRDB module with CBAM to recreate texture features of remote sensing images while maintaining maximum global detail. The

discriminator network is improved by using a U-Net-based network with skip connections to directly merge shallow and deep features. Furthermore, CBAM is incorporated into the U-Net model. These enhancements are crucial for creating high-resolution images, particularly for building intricate features and textures.

Most current super-resolution reconstruction algorithms rely on low-resolution images created by reducing the resolution of high-resolution images. The image pairings produced using this method must include essential characteristics of the real image, leading to unreliable model validity and robustness. We obtained MS remote sensing images from three different flight heights (15 m, 30 m and 60 m) to evaluate the model on real MS remote sensing data. The initial remote sensing images underwent processing techniques such as stitching and radiometric correction to provide MS ortho-images with three distinct resolutions (1×, 2× and 4×). Five single-band images were merged into two sets of three-band images, RGB and NER, to simplify the training process. The images above are split into sub-images of sizes  $256 \times 256$  (1×),  $128 \times 128$  (2×) and  $64 \times 64$  (4×) based on various resolutions. These sub-images are all directly correlated in geographical position. Finally, we obtained the Ginkgo biloba tree dataset based on real MS remote sensing images (Real-MSG).

This study initially evaluated the effect of super-resolution reconstruction by analyzing the generated images through quantitative and qualitative analyses. The RMSRGAN model performed exceptionally well in both quantitative and qualitative assessments. Then, we computed the correlation values between the reconstructed and original images for single-variable yield prediction. The findings indicate that images enhanced with super-resolution using the RMSRGAN model yield superior regression outcomes compared to the original images. Therefore, our proposed approach can effectively perform super-resolution reconstruction of MS remote sensing images obtained in real situations. The reconstructed images can exceed the original ones in yield prediction, demonstrating the improved reconstruction capabilities of the model. Nevertheless, certain limitations in this study should be acknowledged. The first is that images created by the GAN-based super-resolution reconstruction model exhibit artifacts. More artifacts are present in the NER images compared to the RGB images, with most appearing in regions of higher brightness. This may be due to the increased complexity of the spectrum, which may be causing artifacts to arise due to the model's excessive attention to local features. Furthermore, while this study focused on univariate yield prediction models, future research will aim to develop more robust multivariate and multimodal yield prediction models that incorporate a combination of vegetation indices and other relevant parameters. This will allow us to address the observed limitations and increase the applicability of super-resolution techniques in precision agriculture, potentially improving the accuracy and reliability of crop yield predictions under varying environmental conditions.

## 5. Conclusions

Most current super-resolution reconstruction models rely on downsampling high-resolution photos to create image pairs. However, they do not have practical uses in reconstructing real HR and LR image pairs. This study developed an excellent GAN-based super-resolution reconstruction model using MS UAV remote sensing images of Ginkgo biloba trees from real scenes. The results show that integrating the CBAM in the RMSRGAN model and improving the discriminator with a U-Net-based structure can significantly improve the effectiveness in the super-resolution reconstruction of real MS images. The images generated by RMSRGAN performed the best in both quantitative metrics and qualitative evaluations across the four datasets (RGB 2×, RGB 4×, NER 2× and NER 4×). The PSNR values are 32.490, 31.085, 27.084 and 26.819, respectively, and the SSIM values are 0.894, 0.881, 0.832 and 0.818, respectively. On the other hand, RMSRGAN produces images with the best level of image detail recovery in visual perception. These findings indicate that our proposed model improvement strategy is more suitable for the super-resolution reconstruction of MS images in real scenes. Furthermore, the usefulness of RMSRGAN was

assessed by using images of *Ginkgo biloba* before and after reconstruction to predict yield. The results indicate that the images reconstructed by RMSRGAN have superior  $R^2$  and  $RMSE$  values compared to the original low-resolution image in yield prediction. Overall, RMSRGAN can enhance the precision of LR multispectral remote sensing imagery for yield prediction of *Ginkgo biloba* trees. The proposed method can effectively reconstruct low-resolution MS images of *Ginkgo biloba* trees, enhancing the accuracy of *Ginkgo biloba* tree yield prediction and facilitating precise management of *Ginkgo* trees in the field. However, the reconstructed multispectral images produced some visual artifacts, which is a problem that needs to be studied and solved in the future.

**Author Contributions:** Conceptualization, K.F. and M.Z.; Data curation, K.F., M.H. and S.Z.; Funding acquisition, M.Z., L.Q. and X.W.; Methodology, K.F. and M.H.; Project administration, K.F. and M.H.; Software, K.F., M.H. and S.Z.; Supervision, M.Z., L.Q., W.X., H.Z. and B.W.; Validation, K.F.; Writing—original draft, K.F.; Writing—review and editing, K.F., M.Z., L.Q., W.X., H.Z. and B.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** We extend our heartfelt gratitude to the following funding sources: Jiangsu Agriculture Science and Technology Innovation Fund [No. CX (23) 1027], STI2030—Major Projects [No. 2023ZD0405605], Jinpu Research Institute Research Special Funds Project [No. 319610001], and *Metasequoia* Faculty Research Initiation Fee Project [No. 163040193, No. 163040194].

**Data Availability Statement:** The datasets analyzed during the current study and the data of experimental images used to support the findings of this research are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors sincerely thank the academic editors and reviewers for their useful comments and constructive suggestions.

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

## References

- Jha, S.S.; Nidamanuri, R.R. Gudalur Spectral Target Detection (GST-D): A New Benchmark Dataset and Engineered Material Target Detection in Multi-Platform Remote Sensing Data. *Remote Sens.* **2020**, *12*, 2145. [[CrossRef](#)] [[CrossRef](#)]
- Jha, S.S.; Nidamanuri, R.R.; Ientilucci, E.J. Influence of atmospheric modeling on spectral target detection through forward modeling approach in multi-platform remote sensing data. *ISPRS J. Photogramm.* **2022**, *183*, 286–306. [[CrossRef](#)] [[CrossRef](#)]
- Padrón-Hidalgo, J.A.; Laparra, V.; Longbotham, N.; Camps-Valls, G. Kernel Anomalous Change Detection for Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 7743–7755. [[CrossRef](#)] [[CrossRef](#)]
- Shangguan, Y.; Li, J.; Chang, L. Dual-Attention Cross Fusion Context Network for Remote Sensing Change Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2023**, *16*, 8943–8959. [[CrossRef](#)] [[CrossRef](#)]
- Kan, X.; Lu, Z.; Zhang, Y.; Zhu, L.; Sian, K.; Wang, J.; Liu, X.; Zhou, Z.; Cao, H. DSRSS-Net: Improved-Resolution Snow Cover Mapping from FY-4A Satellite Images Using the Dual-Branch Super-Resolution Semantic Segmentation Network. *Remote Sens.* **2023**, *15*, 4431. [[CrossRef](#)] [[CrossRef](#)]
- Kemker, R.; Luu, R.; Kanan, C. Low-Shot Learning for the Semantic Segmentation of Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6214–6223. [[CrossRef](#)] [[CrossRef](#)]
- Xue, Z.; Yu, X.; Yu, A.; Liu, B.; Zhang, P.; Wu, S. Self-Supervised Feature Learning for Multimodal Remote Sensing Image Land Cover Classification. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–15. [[CrossRef](#)] [[CrossRef](#)]
- Moharrami, M.; Attarchi, S.; Gloaguen, R.; Alavipanah, S.K. Integration of Sentinel-1 and Sentinel-2 Data for Ground Truth Sample Migration for Multi-Temporal Land Cover Mapping. *IEEE Trans. Remote Sens.* **2024**, *16*, 1566. [[CrossRef](#)] [[CrossRef](#)]
- Burak, E.; Elif, S. Deep neural network ensembles for remote sensing land cover and land use classification. *Int. J. Digit. Earth* **2021**, *14*, 1868–1881. [[CrossRef](#)]
- Lu, T.; Wan, L.; Qi, S.; Gao, M. Land Cover Classification of UAV Remote Sensing Based on Transformer–CNN Hybrid Architecture. *Sensors* **2023**, *23*, 5288. [[CrossRef](#)] [[CrossRef](#)]
- Wu, S.; Deng, L.; Guo, L.; Wu, Y. Wheat leaf area index prediction using data fusion based on high-resolution unmanned aerial vehicle imagery. *Plant Method* **2022**, *18*, 68. [[CrossRef](#)] [[CrossRef](#)] [[PubMed](#)]
- Marzougui, A.; McGee, R.; Van, V.; Sankaran, S. Remote sensing for field pea yield estimation: A study of multi-scale data fusion approaches in phenomics. *Front. Plant Sci.* **2023**, *14*, 1111575. [[CrossRef](#)] [[CrossRef](#)] [[PubMed](#)]
- Ramin, H.D.; Antoine, D.; Julien, F.; Victor, B.; Jean, T.; Bernard, T.; Edmundo, P.G.; Jeroen, M. Remotely-sensed assessment of the impact of century-old biochar on chicory crop growth using high-resolution UAV-based imagery. *Int. J. Appl. Earth Obs.* **2020**, *91*, 102147. [[CrossRef](#)]

14. Zhang, X.-G. A New Kind of Super-Resolution Reconstruction Algorithm Based on the ICM and the Bicubic Interpolation. In Proceedings of the International Symposium on Intelligent Information Technology Application Workshops, Washington, DC, USA, 21–22 December 2008; pp. 817–820. [[CrossRef](#)]
15. Rasti, P.; Demirel, H.; Anbarjafari, G. Image Resolution Enhancement by Using Interpolation Followed by Iterative Back Projection. In Proceedings of the 21st Signal Processing and Communications Applications Conference (SIU), Haspolat, Turkey, 24–26 April 2013; pp. 1–4. [[CrossRef](#)]
16. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [[CrossRef](#)] [[CrossRef](#)] [[PubMed](#)]
17. 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 (CVPR), Las Vegas, NV, USA, 26 June– 1 July 2016; pp. 1874–1883. [[CrossRef](#)]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
19. 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; Springer: Munich, Germany, 2018; pp. 294–310. [[CrossRef](#)]
20. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690. [[CrossRef](#)]
21. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018. [[CrossRef](#)]
22. Wang, X.; Xie, L.; Dong, C.; Shan, Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Nashville, TN, USA, 20–25 June 2021; pp. 1905–1914. [[CrossRef](#)]
23. Liang, H.; Zheng, C.; Liu, X.; Tian, Y.; Zhang, J.; Cui, W. Super-Resolution Reconstruction of Remote Sensing Data Based on Multiple Satellite Sources for Forest Fire Smoke Segmentation. *Remote Sens.* **2023**, *15*, 4180. [[CrossRef](#)] [[CrossRef](#)]
24. Huang, Y.; Wen, X.; Gao, Y.; Zhang, Y.; Lin, G. Tree Species Classification in UAV Remote Sensing Images Based on Super-Resolution Reconstruction and Deep Learning. *Remote Sens.* **2023**, *15*, 2942. [[CrossRef](#)] [[CrossRef](#)]
25. Zhang, J.; Wang, X.; Liu, J.; Zhang, D.; Lu, Y.; Zhou, Y.; Sun, L.; Hou, S.; Fan, X.; Shen, S.; et al. Multispectral Drone Imagery and SRGAN for Rapid Phenotypic Mapping of Individual Chinese Cabbage Plants. *Plant Phenomics* **2022**, *2022*, 7. [[CrossRef](#)] [[CrossRef](#)] [[PubMed](#)]
26. Klapp, I.; Yafin, P.; Oz, N.; Brand, O.; Bahat, I.; Goldshtein, E.; Cohen, Y.; Alchanatis, V.; Sochen, N. Computational end-to-end and super-resolution methods to improve thermal infrared remote sensing for agriculture. *Precision Agric.* **2021**, *22*, 452–474. [[CrossRef](#)] [[CrossRef](#)]
27. Zeng, Q.; Chang, S.; Wang, S.; Ni, W. Multi-scale adaptive learning network with double connection mechanism for super-resolution on agricultural pest images. *Vis. Comput.* **2024**, *40*, 153–167. [[CrossRef](#)] [[CrossRef](#)]
28. Myneni, R.B.; Hall, F.G.; Sellers, P.J.; Marshak, A.L. The interpretation of spectral vegetation indexes. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 481–486. [[CrossRef](#)] [[CrossRef](#)]
29. Wei, C.; Huang, J.; Wang, X.; Blackburn, G.A.; Zhang, Y.; Wang, S.; Mansaray, L.R. Hyperspectral characterization of freezing injury and its biochemical impacts in oilseed rape leaves. *Remote Sens. Environ.* **2017**, *195*, 56–66. [[CrossRef](#)] [[CrossRef](#)]
30. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.; Gao, X.; Ferreira, L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)] [[CrossRef](#)]
31. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[CrossRef](#)]
32. Liu, Z.; Yu, S.; Deng, H.; Jiang, G.; Wang, R.; Yang, X.; Song, J.; Chen, J.; Mao, X. 3D mineral prospectivity modeling in the Sanshandao goldfield, China using the convolutional neural network with attention mechanism. *Ore Geol. Rev.* **2024**, *164*, 105861. [[CrossRef](#)] [[CrossRef](#)]

**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.