Performance evaluation of super-resolution techniques for characterization and phenotyping of commodity crops using multispectral imagery

Abhiram Karukayil[1], *João F. C. Mota*[1], *Mateo Vásquez-Guevara*[2], *Tito Arevalo-Ramirez*[2] *and Fernando Auat Cheein*[3,4]

[1] School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK, ak5001@hw.ac.uk, j.mota@hw.ac.uk
[2] Department of Electrical Engineering, Department of Mechanical Engineering and Metallurgy, School of Engineering, Faculty of Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile; mvasquezg9@estudiante.uc.cl, tito.arevalo@uc.cl
[3] School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK
[4] Harper Adams University, England, UK, fauat@harper-adams.ac.uk

Correspondence: João F. C. Mota (j.mota@hw.ac.uk)

Keywords: *Vegetation indices, single-image super-resolution, remote sensing, multispectral images.*

Multispectral imagery provides spectral information beyond the visible range for vegetation monitoring, but it typically suffers from limited spatial resolution, which can obscure fine crop structures and affect vegetation-index (VI) estimates. This study evaluates whether single-image super-resolution (SISR) methods developed for red, green, and blue (RGB) imagery can enhance the spatial resolution of red, green, and near-infrared (RGN) crop images while preserving spectral fidelity. Eight pretrained methods are applied directly to RGN images without fine-tuning: bicubic and nearest-neighbor interpolation, Real-ESRGAN, HAT, HiT-SR, SRFormer, ResShift, and StableSR. Reconstruction at 4× and 16× is evaluated for whole images and manually segmented leaf regions. Reconstruction quality is assessed using mean absolute error (MAE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), spectral angle mapper (SAM), and erreur relative globale adimensionnelle de synthèse (ERGAS). Eight VIs, computed from reference and super-resolved images, are evaluated using MAE, PSNR, and SSIM. Transformer-based methods significantly outperform GAN- and diffusion-based approaches in both evaluation settings ($p < 0.001$), with HAT, HiT-SR, and SRFormer forming the highest-performing statistical tier. These findings indicate that RGB-trained SISR can enhance spatial detail in RGN crop imagery while maintaining spectral consistency and VI reliability for high-resolution crop phenotyping.

1 Introduction

The increasing global demand for food, coupled with the overuse of arable land and climate instability, has accelerated the adoption of precision agriculture. A key technology facilitating this transition is multispectral remote sensing, which allows non-destructive monitoring of crop canopies over large areas, either from unmanned aerial vehicles (UAVs) or from satellite platforms [1]. Vegetation indices (VIs) such as the normalized difference vegetation index (NDVI) and the optimized soil-adjusted vegetation index (OSAVI) are commonly derived from multispectral bands to quantify canopy vigor, biomass, and structural characteristics [2]. However, the relatively low spatial resolution of multispectral sensors, which prioritize spectral fidelity over spatial detail, limits their ability to resolve fine-scale structures such as individual leaves. As a result, pixels capturing both healthy and stressed tissues can distort spectral responses and reduce the reliability of VIs for diagnostic applications [3].

Image super-resolution (SR) has emerged as a promising approach to address these spatial limitations by reconstructing high-resolution images from low-resolution observations. Super-resolution methods are broadly categorized into single-image super-resolution (SISR) and multi-frame super-resolution (MFSR) [4]. SISR methods reconstruct high-resolution images from a single input by exploiting internal image statistics, whereas MFSR methods combine information from multiple observations of the same scene. Traditional SISR approaches include classical techniques such as nearest-neighbor upsampling and bicubic upsampling, while recent advances in deep learning have led to substantial improvements using convolutional neural networks (CNNs) [5], generative adversarial networks (GANs) [6], transformers [7], and diffusion models [8]. These models learn from large datasets strong image priors and long-range dependencies, achieving state-of-the-art performance on red, green, and blue (RGB) image reconstruction tasks. Previous studies demonstrated the potential of SR in agricultural contexts, such as disease monitoring and UAV-based crop analysis [9–11]. However, most existing SR architectures are designed for RGB images and optimized for perceptual quality rather than pixel fidelity [12]. This raises concerns about their suitability for multispectral data, where spectral accuracy is a key factor for reliable vegetation analysis.

Multispectral and hyperspectral imaging captures spatial and spectral information simultaneously, providing detailed spectral measurements for remote-sensing applications [13]. Building on these capabilities, super-resolution methods have been developed to enhance the spatial resolution of both types of imagery. Recent unsupervised methods, including M2U-Net [14], EDIP-Net [15], and XINet [16], investigate hyperspectral reconstruction using image observations, spectral information, and image-specific priors. Related work has also explored spectral-band modelling for SR and large-scale representation learning for remote sensing, as illustrated by MambaX [17] and SpectralGPT [18]. These studies are designed for hyperspectral or multispectral data, whereas the present work examines the direct transfer of RGB-trained SISR methods to RGN crop imagery. In agricultural applications, UAV-based multispectral imaging can acquire data from much shorter distances than satellite platforms, but payload capacity, swath width, sensor optics, and operational constraints can still limit the spatial resolution obtainable for fine-scale crop analysis. High-resolution multispectral sensors may also impose additional cost, weight, and power requirements [19]. Consequently, SR may offer a practical alternative for enhancing spatial detail in multispectral imagery, provided that spectral integrity and vegetation-index reliability are preserved. Since this study focuses exclusively on single-image super-resolution, the terms SR and SISR are used interchangeably here.

Motivated by these challenges, we investigate the applicability of RGB-trained SISR models for multispectral crop images, with a focus on improving spatial resolution. To assess their ability to transfer beyond the RGB domain, we apply the methods directly to multispectral RGN images. All super-resolution methods are evaluated using their publicly available pretrained implementations and default inference configurations, without adapting the models to the acquired RGN imagery, allowing us to compare their performance when directly applied to RGN multispectral imagery. In the absence of UAV deployment, and to emulate low-altitude observations, we acquired RGN imagery at ground level using a MAPIR Survey3W multispectral camera, resulting in a horizontal or near-horizontal viewing perspective. We consider eight SISR methods, spanning four methodological categories: classical filtering approaches (bicubic and nearest neighbor [20]), GAN-based (Real-ESRGAN [6]), transformer-based (hybrid attention transformer (HAT) [21], hierarchical transformer for efficient image super-resolution (HiT-SR) [22], SRFormer [7]), and diffusion-based (ResShift [8], StableSR [23]). To provide a comprehensive benchmark, we also consider two scale factors, 4× and 16×, and analyze two different scenarios: whole-image reconstruction, where metrics are computed across the entire image, and leaf-level reconstruction, where the same images are evaluated only within manually segmented leaf regions using masked analysis. For the 4× experiments, each image was processed once using the corresponding 4× model. For 16× reconstruction, the same 4× inference procedure was applied sequentially twice. This protocol was consistently applied across all methods and acquired images. Reconstruction performance on RGN images is assessed using mean absolute error (MAE) [24], peak signal-to-noise ratio (PSNR) [25], structural similarity index (SSIM) [25], spectral angle mapper (SAM), and erreur relative globale adimensionnelle de synthèse (ERGAS). And VI evaluation is performed using eight indices computed from both reference and super-resolved images, where performance is measured on the resulting VI maps using MAE, PSNR, and SSIM. Our main contribution is a systematic benchmark of RGB-trained SR methods for multispectral crop imagery, examining how different SR architectures influence spatial reconstruction quality, spectral consistency, and VI behavior for precision crop phenotyping.

2 Materials and Methods

This section describes how we prepared the dataset, implemented the super-resolution (SR) methods, and used vegetation indices (VI) in the experiments.

2.1 Data Acquisition

We collected the data between 11:00 and 13:00 UTC−4 on 7 March 2025 in a commercial corn field located at 34.198140° S, 70.701147° W, in Graneros commune, Cachapoal Province, Libertador General Bernardo O'Higgins Region, Chile; see Fig. 2a. The study area is characterized by a Mediterranean climate, with cold,

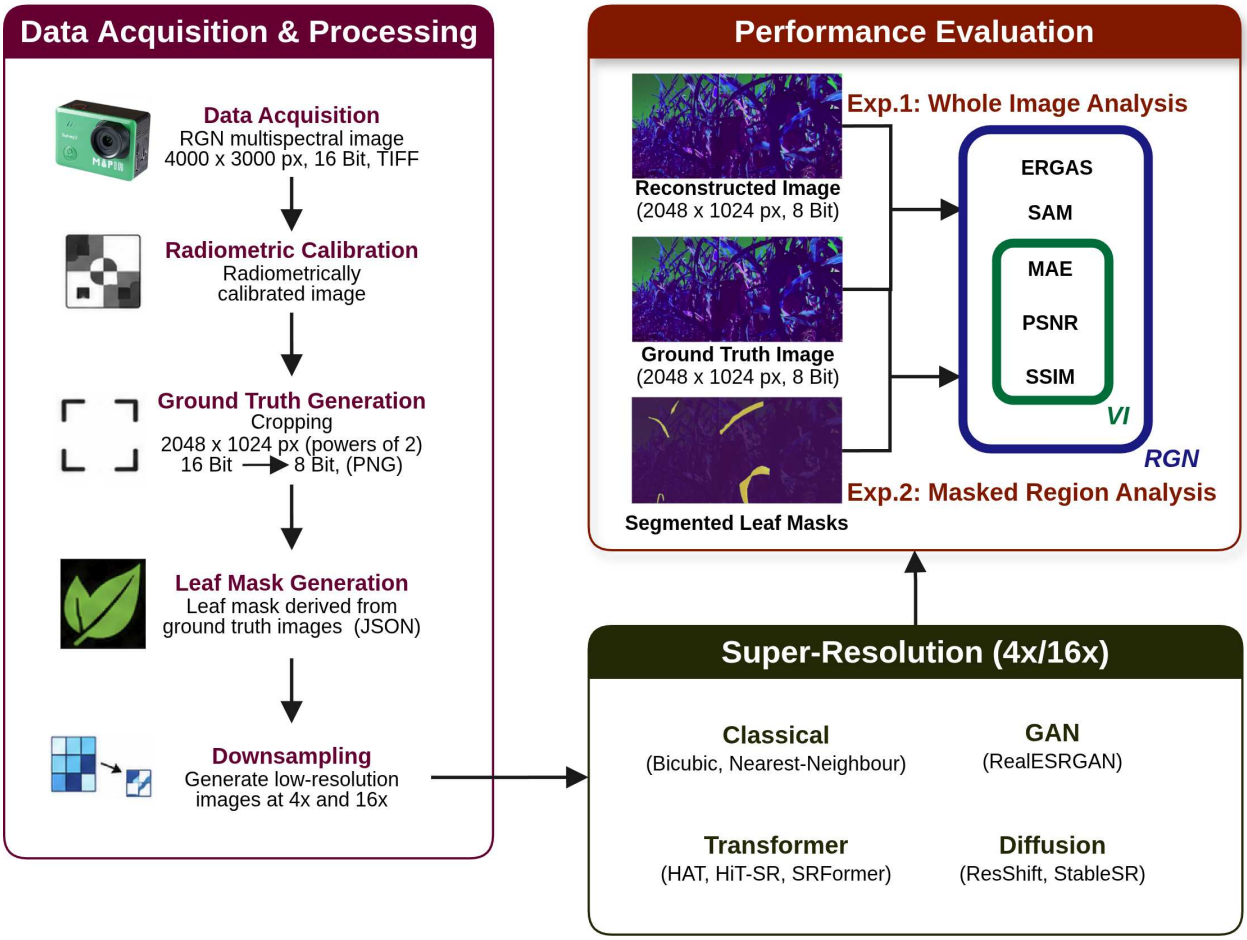
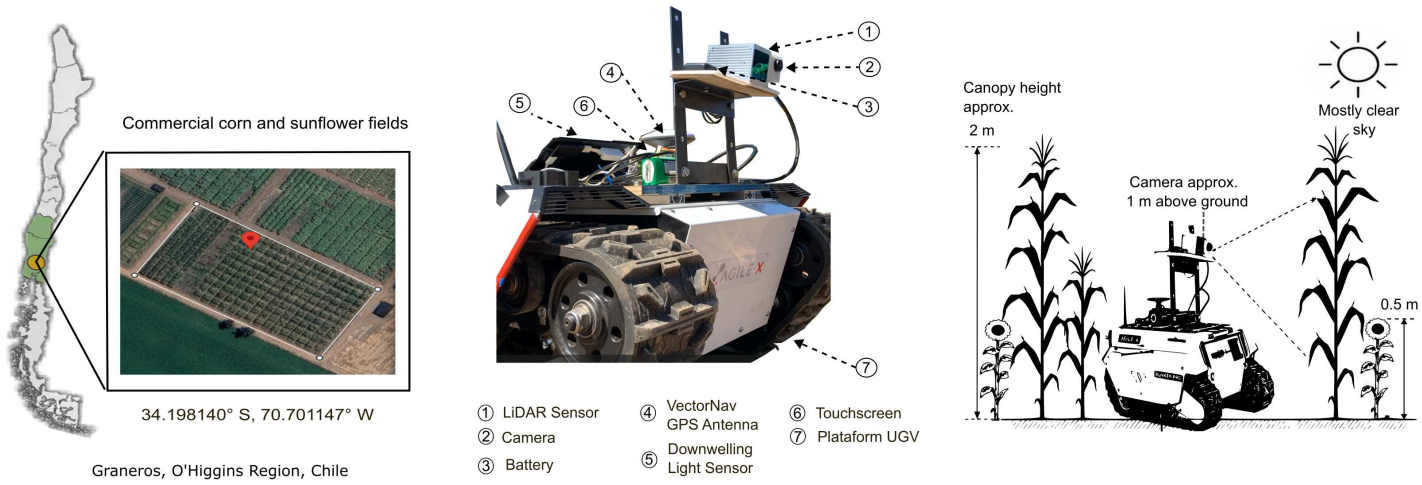


Figure 1: Our experimental workflow for multispectral image acquisition, super-resolution processing, and performance assessment through whole-image (Experiment 1, Sec. 3.3) and leaf-level (Experiment 2, Sec. 3.4) evaluation. Groundtruth images are generated during data acquisition & processing by extracting a centered 2048×1024 image patch and converting it from 16-bit to an 8-bit portable network graphics (PNG) format. The images are subsequently downsampled by factors of $\times 4$ and $\times 16$, reconstructed using different SR methods, and evaluated at the whole-image and masked-region levels. The RGN reconstruction metrics are indicated in the blue box, while VI reconstruction metrics are indicated in the green box.

partly cloudy winters and warm, dry summers, and reported annual temperatures ranging approximately from 4°C to 29°C [26]. During the acquisition period, weather conditions were mostly clear, with air temperatures ranging from 22.2 to 24.5 (°)C, relative humidity between 37% and 46%, wind speeds of 5.0–11.4 km h^{-1}, cloud cover of 1–7%, and shortwave solar radiation between 731 and 866 W m^{-2} [27].

At the time of acquisition, the maize crop was at the V14 phenological stage, approximately ten days before flowering, with an average canopy height of 2 m. In contrast, the sunflower crop was at approximately 10% flowering, with an average canopy height of 0.5 m. The field corresponded to experimental maize and sunflower material sown on 30 October 2024, with a row spacing of 0.70 m and an approximate plant density of 70,000 plants ha^{-1}. See Fig. 3 for illustrative RGB images. We acquired multispectral data using a MAPIR Survey3W multispectral camera mounted on a Bunker Agilex unmanned ground vehicle (UGV), positioned approximately 1 m above the ground, as illustrated in Figs 2b-2c. The camera captures red, green, and near-infrared bands centered at 660, 550, and 850 nm, respectively, enabling the computation of vegetation indices such as the normalized difference vegetation index (NDVI). The system is equipped with a Sony Exmor R IMX117 12 MP CMOS sensor and produces images with a spatial resolution of 4000×3000 pixels. In addition, the camera incorporates a non-fisheye wide-angle lens, which reduces geometric distortion and preserves spatial image quality for subsequent analysis. Images were acquired with ISO fixed at 50, aperture set to $f/2.8$, and a focal length of 3.37 mm, while exposure time was automatically adjusted according to illumination conditions, ranging approximately from 1/447 to 1/133 s during field acquisition. RAW and JPEG image pairs were recorded simultaneously at intervals



(a) Study area and field location (b) UGV platform (c) Multispectral acquisition setup

Figure 2: Ground-based multispectral data acquisition framework. (a) Study area and field location. (b) Bunker Agilex UGV platform configuration. (c) Ground-based multispectral acquisition setup.



(a) Maize (b) Sunflower

Figure 3: RGB images of (a) maize and (b) sunflowers during data collection under identical imaging conditions.

of approximately 2–3 s. While RAW files preserve 12-bit per-channel radiometric information required for quantitative analysis, JPEG files provide 8-bit visual references for rapid inspection.

2.2 Data Preparation

Figure 1 shows our experimental workflow, from data acquisition to evaluation. Following the manufacturer's recommendation, we performed radiometric calibration using MAPIR camera control software [28] and used reflectance calibration targets acquired during the same experimental session. Specifically, we used the MAPIR diffuse reflectance standard calibration target package T4-R125, which contains four diffuse reflectance standards of known reflectance, each measuring 125×125 mm, together with a QR/fiducial marker for automatic target detection and panel sampling. According to the manufacturer, the reflectance standards are characterized over the 250–2500 nm spectral range using spectrophotometers equipped with an integrating sphere. We used 17 calibration-target images within the MAPIR Camera Control workflow: 13 (resp. 4) images acquired before (resp. after) the field sequence. The RAW images were imported into the software, the T4-R125 calibration target was automatically identified through the QR/fiducial marker, and the four reflectance panels were sampled to estimate the relationship between pixel intensity and surface reflectance. MAPIR Camera Control then applied the radiometric calibration to the RAW image sequence and exported the results as reflectance-calibrated, 16-bit, three-band tagged image file format (TIFF) images. This workflow reduces variability associated with illumination changes, sensor

spectral response, and optical effects such as vignetting, thereby improving the consistency of reflectance measurements across the dataset.

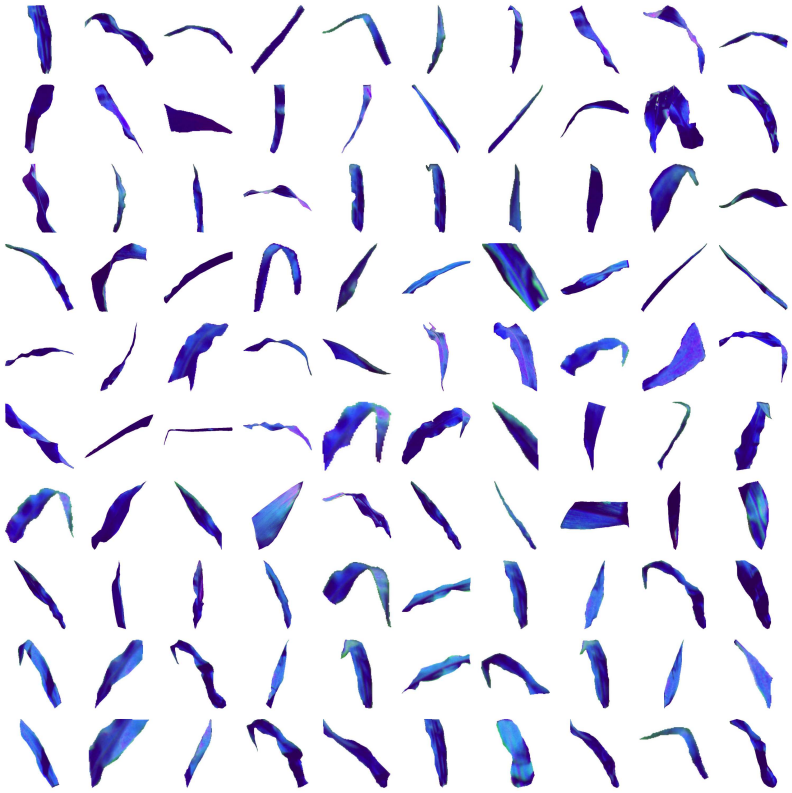


Figure 4: Visualization of the 100 manually segmented maize leaves used for masked-region analysis (Sec. 3.4). Note the morphological variability within the dataset.

To evaluate the performance of SISR techniques, we acquired a dataset of 75 radiometrically calibrated multispectral images comprising red, green, and near-infrared (RGN) bands. The dataset included 35 maize field images and 40 sunflower field images, which served as the high-resolution reference data for all experiments. Groundtruth images were generated by cropping the original 4000×3000 pixel, 16-bit TIFF images to 2048×1024 pixels, ensuring compatibility with the 4× and 16× downsampling and reconstruction procedures. Low-resolution inputs were generated by downsampling the reference images using the INTER_AREA method implemented in OpenCV [29] in Python. The cropped images were subsequently converted to 8-bit PNG format for compatibility with all the evaluated SR methods. We used the full dataset to analyze whole-image reconstruction (Sec. 3.3), and 100 maize leaves for leaf-level reconstruction (Sec. 3.4). For the latter, we selected 20 non-adjacent images, from which we manually segmented 5 leaves using labelme-ytu [30]. The resulting annotations were stored in javascript object notation (JSON) format and used for the masked-region evaluation, as shown in Figure 4.

2.3 Super-Resolution Methods

Table 1 summarizes the SR methods considered in this study. They include classical interpolation, GAN-based, transformer-based, and diffusion-based, providing a representative spectrum of modern SR approaches. Bicubic interpolation [31] estimates new pixel values using cubic interpolation in a local neighborhood, based on a 4×4 grid of neighboring pixels. Nearest-neighbor (NN) interpolation [39] assigns to each output pixel the value of the closest input pixel without introducing additional spatial information. Although NN interpolation does not improve image detail, it preserves the original pixel arrangement, although this often results in blocky artifacts and reduced visual smoothness. Transformer-based methods, such as hybrid attention transformer (HAT) [21], hierarchical transformer for efficient image super-resolution (HiT-SR) [22], and SRFormer [7], use self-attention mechanisms to capture long-range spatial dependencies with the aim of recovering fine structural and textural details. They overcome limitations

Table 1: Super-resolution algorithms evaluated in this study. They are classified by category and, for the learning-based methods, the training datasets are also indicated.

SR Technique	Category	Training Datasets
Bicubic [31]	Classical	-
NN [20]	Classical	-
HAT [21]	Transformer	ImageNet [32] and DF2K(DIV2K + Flickr2K [33])
HiT-SR [22]	Transformer	DIV2K [34]
SR Former [7]	Transformer	DIV2K [34] and DF2K(DIV2K + Flickr2K [33])
Real ESRGAN [6]	GAN	DIV2K [34], Flickr2K [33] and OutdoorSceneTraining (OST) [35]
ResShift [8, 36]	Diffusion Model	ImageNet [32]
Stable SR [23]	Diffusion Model	DIV2K [34], DIV8K [37], Flickr2K [33], OST [35] and FFHQ [38]

of early vision transformers by improving cross-window and multi-scale feature interactions. For example, HAT [21] enhances information exchange via hybrid attention, which combines channel attention with overlapping spatial attention. HiT-SR [22], on the other hand, employs a hierarchical multi-scale windowing strategy with efficient spatial–channel interaction modules to capture broader contextual dependencies with reduced computational complexity. SRFormer implicitly expands the receptive field by using a permuted self-attention mechanism. Real-ESRGAN [6] builds on the generative adversarial network (GAN) model ESRGAN [40] by incorporating a high-order degradation model to simulate realistic complex distortions. These methods are trained using adversarial and perceptual losses, which encourage visually sharp and perceptually realistic textures, albeit at the cost of reduced pixel-level accuracy. Training stability is improved by using a U-Net-based [41] discriminator with spectral normalization, which enhances the discrimination of structural artifacts and stabilizes adversarial learning. Diffusion-based methods, such as ResShift [8, 36] and StableSR [23], formulate SR as an iterative denoising process. These models progressively refine low-resolution inputs through stochastic sampling steps. However, reliance on iterative refinement may introduce deviations in spectral accuracy in remote sensing applications where radiometric consistency must be maintained.

Overall, the methods assessed here mainly differ in how they reconstruct high-resolution information from low-resolution inputs. Traditional interpolation techniques estimate pixel values using information from neighboring pixels without learning from data, making them straightforward and computationally efficient. Transformer-based approaches learn spatial features and use attention mechanisms to model dependencies across different regions of an image, supporting the reconstruction of complex structures. GAN-based methods use adversarial training and perceptual losses to produce more visually realistic outputs, although the synthesized details may not closely correspond to the ground-truth pixel values. Diffusion-based techniques formulate reconstruction as an iterative denoising process, enabling the generation of fine details but with higher computational cost and the potential to introduce artifacts.

2.4 Vegetation Indices

Vegetation indices (VIs), which enable assessing crop condition, biomass, canopy vigor, and physiological stress, can be computed from reflectance values obtained from multispectral images. Table 2 summarizes the VIs we used, all of which computed from the red, green, and near-infrared (NIR) bands. The normalized difference vegetation index (NDVI) and difference vegetation index (DVI) were used to characterize vegetation presence and canopy condition. Biomass-related characteristics were assessed using the ratio vegetation index (RVI) and modified simple ratio (MSR), both based on the relationship between NIR and red reflectance. The MSR can reduce saturation effects in dense vegetation. To account for soil background effects, we also evaluated the soil adjusted vegetation index (SAVI) and optimized soil adjusted vegetation index (OSAVI). Physiological stress and chlorophyll-related responses were assessed using the green normalized difference vegetation index (GNDVI), which uses green reflectance instead of red reflectance and is more sensitive to variations in chlorophyll content.

Table 2: Vegetation indices used in this study to assess super-resolution reconstruction. R_{NIR}, R_{Red}, and R_{Green} denote reflectance values of the near-infrared (850 nm), red (550 nm), and green (660 nm) bands.

Index	**Description**	**Formula**	**Ref.**
DVI	Difference Vegetation Index	$R_{\text{NIR}} - R_{\text{Red}}$	[42]
MSAVI	Modified Soil Adjusted Vegetation Index	$2R_{\text{NIR}} + \frac{1}{2} - \frac{1}{2}\sqrt{(2R_{\text{NIR}}+1)^2 - 8(R_{\text{NIR}} - R_{\text{Red}})}$	[43]
NDVI	Normalized Difference Vegetation Index	$(R_{\text{NIR}} - R_{\text{Red}})/(R_{\text{NIR}} + R_{\text{Red}})$	[44]
OSAVI	Optimized Soil Adjusted Vegetation Index	$1.16(R_{\text{NIR}} - R_{\text{Red}})/(R_{\text{NIR}} + R_{\text{Red}} + 0.16)$	[45]
RVI	Ratio Vegetation Index	$R_{\text{NIR}}/R_{\text{Red}}$	[46]
SAVI	Soil Adjusted Vegetation Index	$1.5(R_{\text{NIR}} - R_{\text{Red}})/(R_{\text{NIR}} + R_{\text{Red}} + 0.5)$	[47]
GNDVI	Green Normalized Difference Vegetation Index	$(R_{\text{NIR}} - R_{\text{Green}})/(R_{\text{NIR}} + R_{\text{Green}})$	[48]
MSR	Modified Simple Ratio	$(R_{\text{NIR}}/R_{\text{Red}} - 1)/\sqrt{R_{\text{NIR}}/R_{\text{Red}} + 1}$	[49]

3 Experimental Results

This section presents the evaluation metrics and assesses the performance of the super-resolution methods described in Section 2.3. Their ability to preserve spatial detail and spectral consistency is evaluated using the complementary metrics introduced in Section 3.1. Since multispectral super-resolution (SR) requires the preservation of both spatial detail and spectral information, reconstruction quality is evaluated from these two perspectives. For readability, all MAE values are scaled by 100, and the input RGB images are normalized to the range $[0, 1]$.

We conducted two complementary experiments to evaluate reconstruction performance at different spatial scales. Experiment 1, reported in Section 3.3, evaluates whole-image reconstruction using the complete dataset of 35 maize-field images and 40 sunflower-field images. Experiment 2, reported in Section 3.4, evaluates leaf-level reconstruction within manually segmented mask regions using 100 maize leaves extracted from the same dataset. Together, these experiments allow us to assess the performance of the competing approaches both at the whole-image level and within individual leaf regions. In both experiments, reconstruction quality for RGN images is measured using MAE, PSNR, SSIM, SAM, and ERGAS, while VI reconstruction quality is evaluated using MAE, PSNR, and SSIM. Numerical values reported below in the text or in tables are average values computed across all the data samples. In contrast, boxplots visualize the distribution of errors and display the median as their central line. The evaluation metrics used in these experiments are described below.

3.1 Evaluation Metrics

Unlike RGB or grayscale images, multispectral SR requires preserving both spatial detail and spectral information. To evaluate reconstruction performance, we thus selected a range of metrics, each of which captures a different aspect of image quality. Given an image of size $H \times W \times C$, where H and W denote spatial dimensions and C the number of channels/bands, we represent the value of each pixel of the reconstructed (resp. groundtruth) image by $x_i \in \mathbb{R}$ (resp. $y_i \in \mathbb{R}$), for $i = 1, \ldots, N := H \cdot W \cdot C$. We consider $C = 1$ for single-channel VI analysis and $C = 3$ for RGN images.

To assess reconstruction quality, we used the mean absolute error (MAE) [24], $(1/N)\sum_{i=1}^{N} |x_i - y_i|$, with lower values indicating smaller reconstruction errors and hence better performance. We also considered the standard peak signal-to-noise ratio (PSNR) [25], $10\log_{10}\left(\text{MAX}^2/\text{MSE}\right)$, which compares the maximum possible pixel value, MAX, with the mean-squared error (MSE), $(1/N)\sum_{i=1}^{N}(x_i - y_i)^2$. Note that PSNR is expressed on a log scale (dB), and higher values indicate lower reconstruction error. The structural similarity index measure (SSIM) [25], defined as

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{1}$$

evaluates similarity between reconstructed and groundtruth images in terms of luminance, contrast, and structural characteristics. In (1), μ_x and μ_y represent mean intensities, σ_x^2 and σ_y^2 variances, and σ_{xy} is the covariance between x and y. The constants $C_1, C_2 > 0$ are used for numerical stability. The higher the SSIM, the more similar the images are, with the value of 1 indicating perfect similarity. We computed SSIM values separately for each spectral band and report the average across all the bands.

To assess spectral reconstruction accuracy, we used spectral angle mapper (SAM) [50] and erreur relative globale adimensionnelle de synthèse (ERGAS) [51]. SAM is defined as

$$\mathrm{SAM}(x, y) = \cos^{-1}\left(\frac{\sum_{i=1}^{C} x_i y_i}{\sqrt{\sum_{i=1}^{C} x_i^2}\sqrt{\sum_{i=1}^{C} y_i^2}}\right),$$

where C denotes the number of spectral bands. It views each image as a vector in a high-dimensional space and measures their angular similarity, ignoring magnitude factors. In turn, ERGAS is defined as

$$\mathrm{ERGAS} = \frac{100}{r}\sqrt{\frac{1}{C}\sum_{c=1}^{C}\left(\frac{\mathrm{RMSE}^{(c)}(x, y)}{\mu^{(c)}}\right)^2},$$

where r represents the ratio between the spatial resolutions of the high- and low-resolution images, taking the values 4 for 4× upsampling and 16 for 16× upsampling. Again, C denotes the number of channels. The root mean squared error (RMSE) in the cth spectral band is defined as $\mathrm{RMSE}^{(c)}(x, y) = \sqrt{(1/N_c)\sum_{i=1}^{N_c}\left(x_i^{(c)} - y_i^{(c)}\right)^2}$, where $x_i^{(c)}$ and $y_i^{(c)}$ denote the pixel values at location i of the reconstructed and reference images in channel c, and N_c is the total number of pixels in that band. Also, $\mu^{(c)} = (1/N_c)\sum_{i=1}^{N_c} y_i^{(c)}$ denotes the mean value of $y^{(c)}$. By normalizing the reconstruction error within each band, ERGAS allows comparing different methods in terms of their ability to preserve spectral information.

To statistically validate performance differences across SR models, we performed non-parametric Kruskal–Wallis tests followed by Dunn's post-hoc tests with Bonferroni correction [52] for all metrics under each experimental setup. We first applied the Kruskal–Wallis test to assess whether at least one model differed from the others. The degrees of freedom for this test are $\mathrm{df} = k - 1 = 7$, where $k = 8$ is the number of SR models compared. When the Kruskal–Wallis test indicated a significant difference, we performed pairwise comparisons between SR models using Dunn's post-hoc test. To reduce the risk of false positives due to multiple comparisons, we applied a Bonferroni correction.

3.2 Computational Complexity of Super-Resolution Methods

To assess the practical computational requirements of the evaluated methods, Table 3 reports the number of trainable parameters, floating-point operations (FLOPs), and measured inference time. Learning-based methods were evaluated using publicly released pretrained checkpoints without retraining or fine-tuning; therefore, training time is not applicable to the present study. FLOPs and inference time were measured for a 512×256-pixel low-resolution RGN input at 4× upscaling on an NVIDIA GeForce RTX 4090 Laptop GPU with 16 GB memory. The reported inference times correspond to a single 4× inference pass. Table 3 shows substantial differences in computational requirements among the evaluated SISR methods. Bicubic and nearest-neighbor interpolation required negligible processing time. Among the learning-based approaches, HiT-SR had the lowest parameter count (0.87 M) and FLOPs (0.23 T), with an inference time of 0.31 s per image. Real-ESRGAN has the least inferencing time (0.24 s per image) even with more parameters and FLOPs than HiT-SR. HAT and SRFormer required 1.74 s and 1.42 s per image, respectively. The diffusion-based methods had the highest computational requirements. StableSR required 1554.64 M parameters and 371.46 T FLOPs, with a mean inference time of 32.42 s per image. These results indicate that both reconstruction quality and inference cost should be considered when selecting an SR method for large-scale agricultural imaging.

Table 3: Computational characteristics of the evaluated super-resolution methods. Learning-based methods were evaluated using publicly released pretrained checkpoints without retraining or fine-tuning. FLOPs and inference time were measured for an LR RGN input of 512×256 pixels at 4× upscaling using an NVIDIA GeForce RTX 4090 (16 GB, mobile) GPU.

SR Technique	Parameters (M)	FLOPs (T)	Inference time (s/image)
Bicubic [31]	–	–	3.79×10^{-4}
Nearest neighbor [20]	–	–	3.39×10^{-4}
HAT [21]	20.77	5.48	1.74
HiT-SR [22]	0.87	0.23	0.31
SRFormer [7]	10.52	3.04	1.42
Real-ESRGAN [6]	16.70	4.70	0.24
ResShift [8, 36]	173.92	43.21	4.36
Stable SR [23]	1554.64	371.46	32.42

Checkpoint details. For reproducibility, the publicly released pretrained weights used for inference are: HAT:`-HAT_SRx4_ImageNet-pretrain.pth`; HiT-SR:`-HiT-SRF-4x.pth`; SRFormer:`-SRFormer_SRx4_DF2K.pth`; RealESRGAN:`-RealESRGAN_x4plus.pth`; ResShift–`resshift_realsrx4_s4_v3.pth` with `autoencoder_vq_f4.pth`; StableSR–`stablesr_turbo.ckpt` with `vqgan_cfw_00011.ckpt`. Bicubic and NN do not require pretrained weights.

3.3 Experiment 1: Whole-Image Reconstruction

Figure 5 shows the reconstruction at a 4× factor by all the methods of two representative whole-image RGN examples, one for a maize crop (Fig. 5a), another for a sunflower crop (Fig. 5b). In both examples, it can be observed that the transformer- and attention-based algorithms (HAT, HiT-SR, and SRFormer) obtained the best reconstruction performance, followed by the classical methods (bicubic and NN). The worst performance was obtained by generative methods (RealESRGAN, ResShift, and StableSR) which, despite producing visually-pleasing reconstructions, introduced artifacts inexistent in the groundtruth images. Figure 6a reports the results for each metric in the entire dataset for both scaling factors 4× and 16×. The results are visualized using box plots, where the central horizontal line within each box denotes the median performance. The box boundaries represent the first and third quartiles (Q1–Q3), corresponding to the interquartile range (IQR). The whiskers extend to the most extreme data points within 1.5 times the IQR, while any samples outside this range are treated as outliers and plotted individually. To statistically validate performance differences given the skewed error distributions and outliers, we performed non-parametric Kruskal–Wallis tests followed by post-hoc Dunn's tests with Bonferroni correction. The omnibus tests confirmed significant differences among models across all metrics for both 4× ($H(7) \in [324.14, 516.55]$, $p < 0.001$) and 16× ($H(7) \in [76.00, 383.70]$, $p < 0.001$) setups.

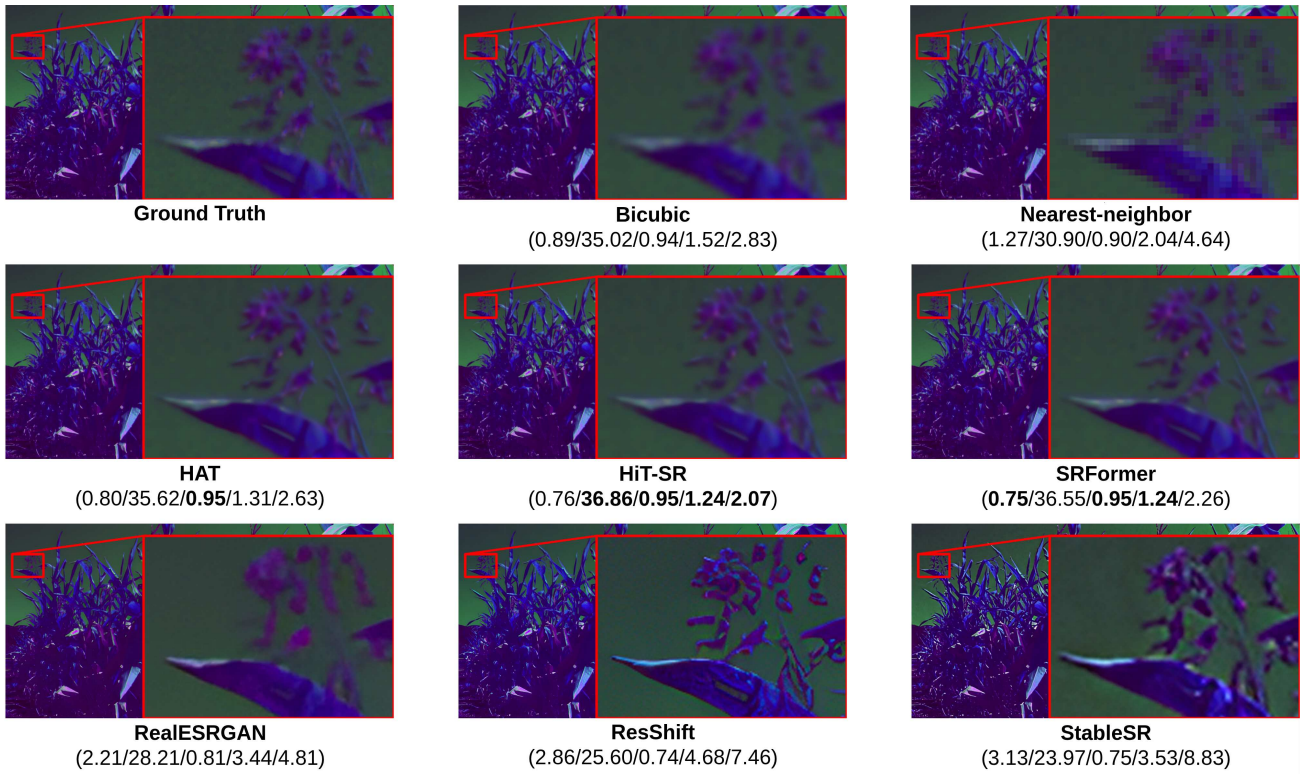
Fig. 5 further validates this trend: for 4× upsampling (blue boxes), transformer- and attention-based methods consistently achieve the strongest results. Specifically, SRFormer obtained the lowest MAE (0.98) and lowest SAM (1.32°), while sharing the highest SSIM (0.94) with HAT (1.00 MAE, 1.34° SAM). HiT-SR achieved the highest PSNR (34.95 dB) and the lowest ERGAS (2.09), with only marginal differences relative to SRFormer and HAT. Post-hoc pairwise tests confirmed that all three Transformer models significantly outperformed RealESRGAN, ResShift, and StableSR across all metrics ($p < 0.001$, Bonferroni-adjusted), while differences among HAT, HiT-SR, and SRFormer were not statistically significant. Despite its simplicity, bicubic interpolation remained a competitive baseline, achieving a PSNR of 33.58 dB and an SSIM of 0.92. In contrast, the generative approaches exhibited substantially larger reconstruction errors. For a 16× factor (orange boxes), reconstruction quality deteriorated across all methods and exhibited larger variations. Transformer based approaches remained robust; HAT attained the lowest MAE (3.20), highest SSIM (0.73), and lowest SAM (3.24°), closely followed by SRFormer (3.24 MAE, 0.72 SSIM, 3.29° SAM) and HiT-SR (3.26 MAE, 0.72 SSIM, 3.38° SAM), with no statistically significant differences within the Transformer family. Pairwise tests confirmed that Transformers maintained statistically superior MAE, PSNR, SSIM, and SAM over generative models ($p \leq 0.008$). Bicubic interpolation attained the highest PSNR (24.27 dB) and matched HiT-SR's lowest ERGAS (1.98), while maintaining competitive fidelity relative to deep learning models at this extreme scaling factor. Generative and diffusion models continued

to yield the lowest quantitative fidelity, with ResShift and StableSR recording the highest spectral distortion and reconstruction errors.

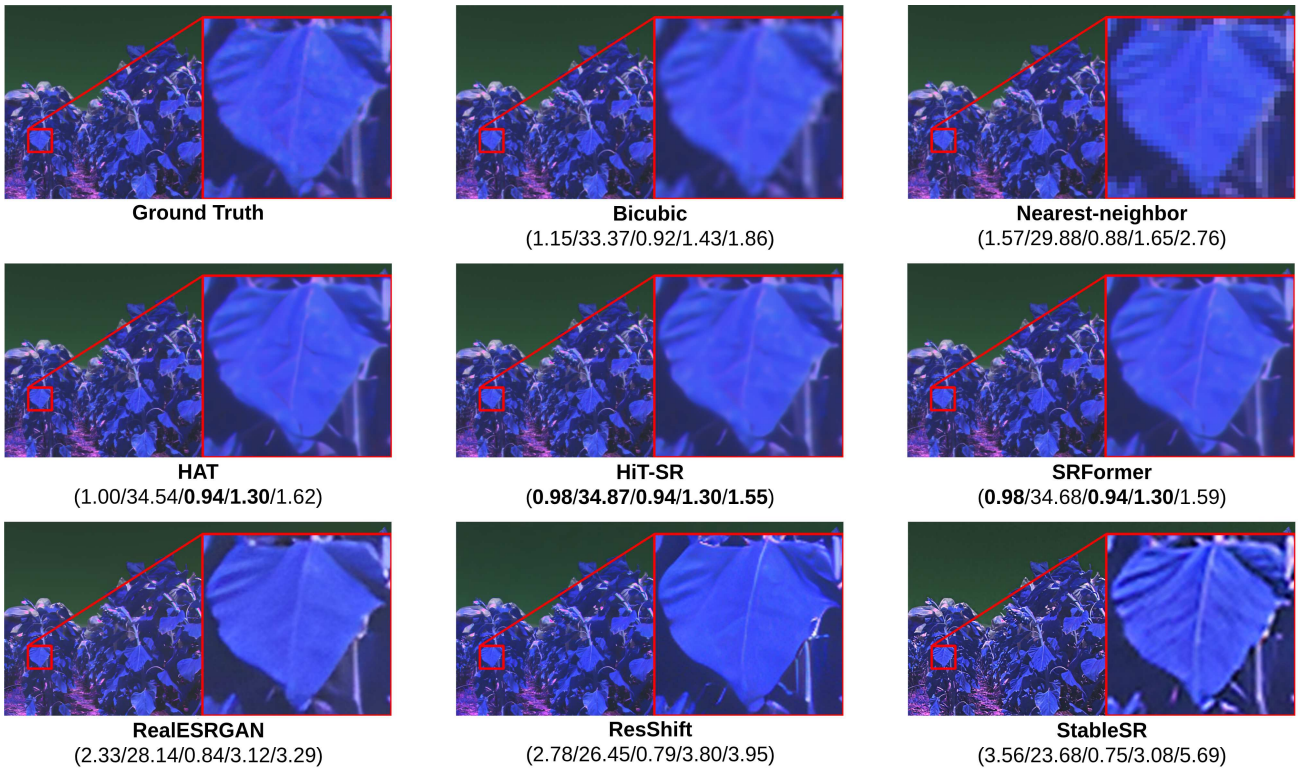
Table 4a reports VI reconstruction results, with each column displaying the MAE, PSNR, and SSIM values for each VI index of Table 2. Non-parametric Kruskal–Wallis tests confirmed statistically significant performance differences across models for all eight vegetation indices at both 4× ($H(7) \in [395.95, 499.24]$, $p < 0.001$) and 16× ($H(7) \in [76.00, 455.25]$, $p < 0.001$) scale factors. Post-hoc Dunn's tests with Bonferroni correction verified that Transformer architectures (HAT, HiT-SR, and SRFormer) significantly outperformed GAN- and diffusion-based baselines across all VI metrics ($p < 0.001$), while pairwise differences within the Transformer family were not statistically significant. It can be seen that HAT, SRFormer, and HiT-SR consistently achieve the strongest performance across both scaling factors, with HAT yielding the highest overall performance on most indices. At 4×, SAVI and OSAVI were reconstructed most accurately, where HAT achieved 41.46 dB PSNR and 0.95 SSIM for SAVI (MAE of 1.78) and 39.41 dB PSNR and 0.93 SSIM for OSAVI (MAE of 1.92), with SRFormer and HiT-SR producing closely comparable results. The DVI and NDVI remain within a relatively high-quality reconstruction range, whereas ratio-based indices such as MSR and RVI are reconstructed with poorer accuracy across all evaluated methods. At 16×, performance degradation becomes more noticeable, especially for the ratio-based indices. HAT maintained the highest accuracy across these indices, achieving 34.85 dB PSNR (MAE of 3.64, SSIM of 0.85) for SAVI and 32.90 dB PSNR (MAE of 3.86, SSIM of 0.80) for OSAVI, compared to 34.67 dB and 32.70 dB for SRFormer, and 34.52 dB and 32.54 dB for HiT-SR. In contrast, GAN and diffusion-based methods exhibited substantially lower reconstruction quality, with ResShift producing the lowest RVI reconstruction at 16×, achieving an MAE of 64.72, a PSNR of 12.37 dB, and an SSIM of 0.44.

3.4 Experiment 2: Leaf-Level Reconstruction Analysis Within Masked Regions

Figure 6b and Table 4b show the results of leaf-level SR across 100 manually segmented leaf regions. Unlike Experiment 1, where reconstruction metrics were computed over the entire image, here metrics are computed exclusively in leaf regions. Non-parametric Kruskal–Wallis tests confirmed statistically significant performance variations across models across all raw bands and vegetation indices at both 4× ($H(7) \in [338.31, 593.72]$, $p < 0.001$) and 16× ($H(7) \in [101.41, 420.34]$, $p < 0.001$) scale factors. Comparing Figs. 6a and 6b for 4×, we observe that reconstruction performance improves in terms of PSNR, SSIM, and ERGAS for the top-performing methods. SRFormer obtained the best metrics: MAE of 1.06, PSNR of 36.63 dB, SSIM of 0.94, SAM of 1.26°, and ERGAS of 2.00. Relative to Experiment 1, PSNR increases from 34.86 to 36.63 dB for SRFormer and from 34.95 to 36.60 dB for HiT-SR, and from 34.64 to 36.12 dB for HAT. Bicubic interpolation also shows improved performance, with PSNR increasing from 33.58 dB to 34.69 dB while maintaining an SSIM of 0.92. Diffusion-based and GAN-based methods obtained larger errors and lower spectral consistency for most metrics. At 16×, reconstruction performance decreases for all methods. SRFormer achieved the lowest MAE (4.73) and the highest SSIM (0.63), HAT obtained the lowest SAM (4.94°), and HiT-SR achieved the highest PSNR (23.80 dB) alongside the lowest ERGAS (2.40) among the learning-based methods. Unlike Experiment 1 where bicubic interpolation achieved the top PSNR at 16×, here the transformer based models (HiT-SR at 23.80 dB, SRFormer at 23.76 dB, and HAT at 23.55 dB) outperform bicubic interpolation (23.16 dB). Compared with Experiment 1, all methods exhibit reduced performance at 16×, and differences between learning-based methods and bicubic interpolation are less pronounced under this scaling factor. As shown in Fig. 6b, generative models exhibit higher variability and more outliers.

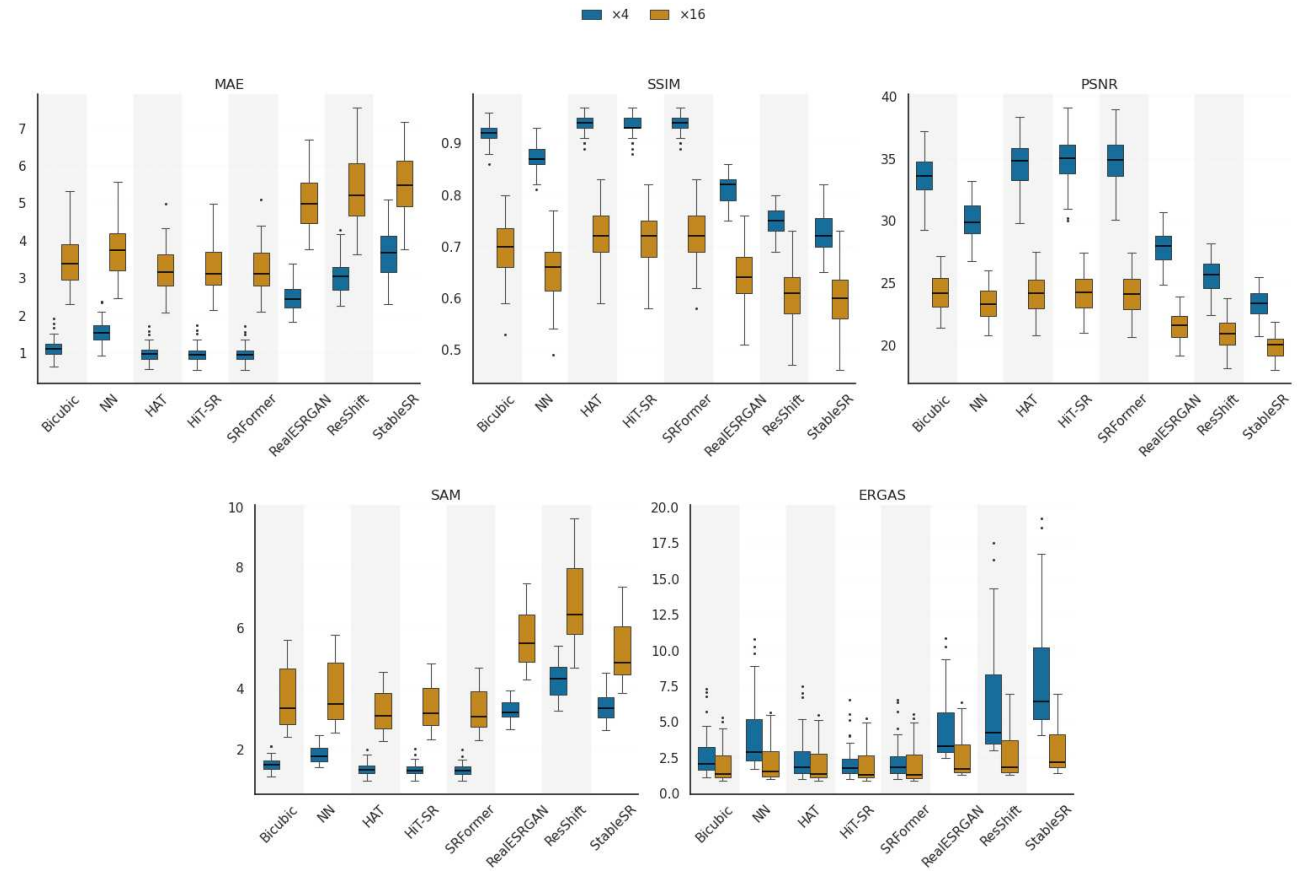


(a) Maize

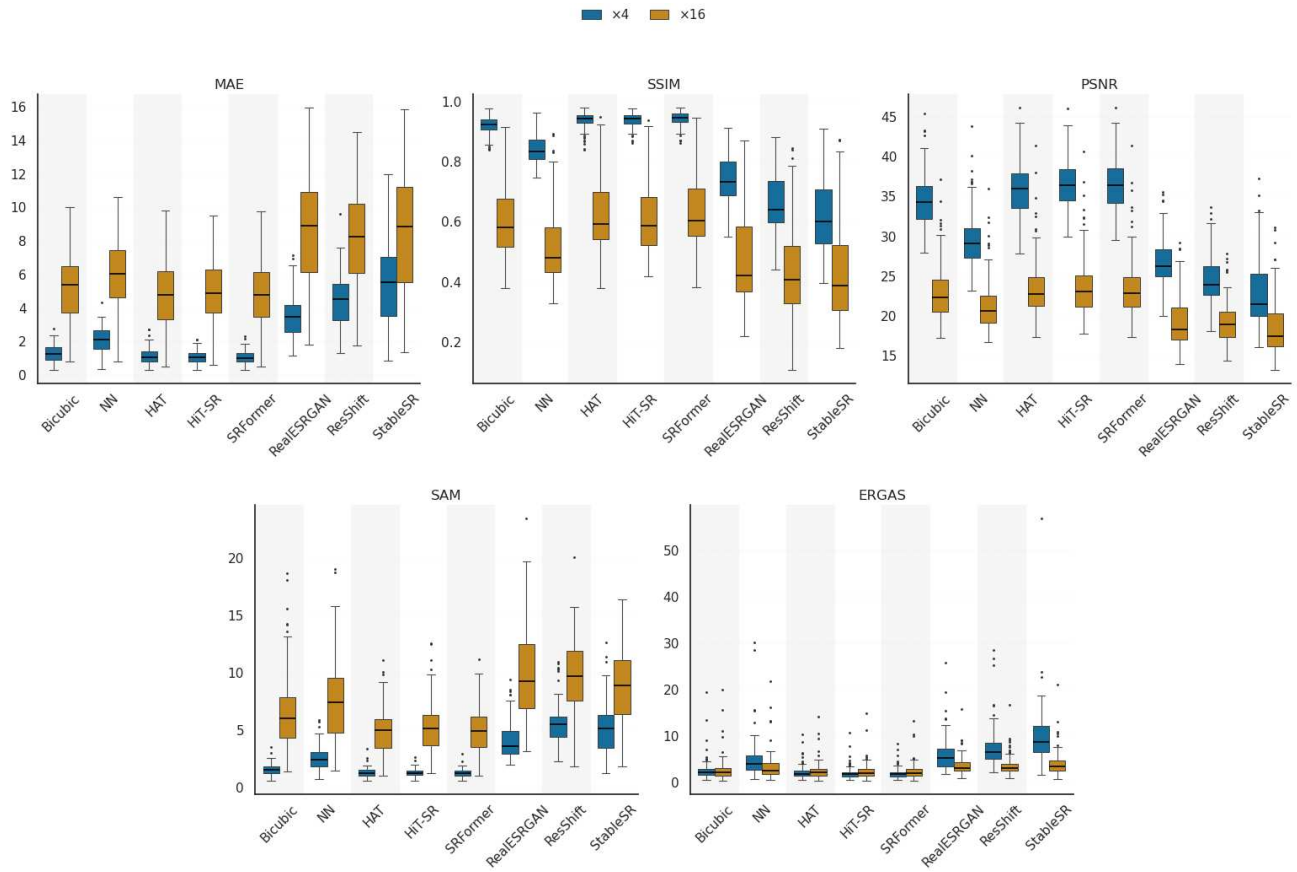


(b) Sunflower

Figure 5: Qualitative comparison of two representative 4× single-image super-resolution (SISR) reconstruction examples for (a) maize and (b) sunflower crops. For each example, the full reconstructed image is shown together with an enlarged local region (zoomed-in view) to facilitate detailed comparison of crop structures, textures, and phenotypic characteristics. Results from all SISR methods are accompanied by the corresponding reconstruction metrics: MAE↓, PSNR (dB)↑, SSIM↑, SAM (°)↓, and ERGAS↓. The zoomed-in regions highlight differences in the preservation of fine-scale agricultural structures and spectral-spatial details. While the generative approaches (last row) produce visually sharper and more perceptually appealing reconstructions, they may introduce artificial structures or textures that are not supported by the groundtruth, indicating a potential trade-off between perceptual sharpness and faithful reconstruction of agriculturally relevant information.



(a) Whole image analysis



(b) Masked region analysis

Figure 6: Distributional comparison of RGN image reconstruction quality across SR models at 4× and 16× scaling factors. Figure (a) shows the whole image results (Sec. 3.3) and (b) shows masked leaf region results (Sec. 3.4).

Table 4: Quantitative evaluation of vegetation index (VI) reconstruction derived from RGN imagery for different SR methods at 4× and 16× scaling factors using MAE ↓, PSNR (dB) ↑, and SSIM ↑. Best results in **bold**, and worst underlined. For readability, MAE values are scaled by 100. (a) Whole-image experiments. (b) Leaf-level experiments.

(a) Performance of super-resolution methods evaluated across the entire image.

SR Technique	Scale	DVI	GNDVI	MSAVI	MSR	NDVI	OSAVI	RVI	SAVI
		MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM
Bicubic	4×	1.69 / 37.66 / 0.92	2.15 / 34.57 / 0.91	1.88 / 37.34 / 0.90	4.27 / 32.68 / 0.82	2.18 / 36.42 / 0.88	2.04 / 38.80 / 0.91	8.47 / 34.07 / 0.86	1.91 / 40.80 / 0.94
	16×	3.59 / 30.94 / 0.78	6.92 / 26.19 / 0.70	3.79 / 31.07 / 0.76	8.48 / **26.45** / 0.59	4.23 / **30.51** / **0.72**	4.01 / 32.77 / 0.79	17.13 / **27.65** / 0.69	3.83 / 34.60 / 0.84
Nearest Neighbor	4×	2.03 / 35.63 / 0.89	2.79 / 32.00 / 0.87	2.20 / 35.60 / 0.87	4.92 / 31.10 / 0.78	2.49 / 34.94 / 0.85	2.36 / 37.23 / 0.89	9.83 / 32.44 / 0.83	2.23 / 39.11 / 0.92
	16×	3.86 / 30.12 / 0.75	7.03 / 25.55 / 0.67	4.03 / 30.32 / 0.73	8.92 / 25.79 / 0.56	4.42 / 29.92 / 0.69	4.23 / 32.12 / 0.77	18.10 / 26.98 / 0.66	4.06 / 33.88 / 0.82
HAT	4×	**1.55** / **38.45** / **0.93**	1.82 / 36.04 / 0.93	**1.75** / **38.01** / 0.92	**4.01** / **33.17** / **0.84**	**2.05** / **37.00** / **0.89**	**1.92** / **39.41** / **0.93**	**7.97** / 33.95 / **0.88**	**1.78** / **41.46** / **0.95**
	16×	**3.35** / **31.36** / **0.79**	**5.36** / **26.94** / **0.74**	**3.60** / **31.34** / **0.77**	**8.33** / 26.18 / **0.60**	**4.11** / 30.47 / **0.72**	**3.86** / **32.90** / **0.80**	**17.06** / 26.67 / **0.70**	**3.64** / **34.85** / **0.85**
HiT-SR	4×	1.57 / 38.39 / **0.93**	1.80 / 36.41 / 0.93	1.77 / 37.93 / 0.91	4.05 / 33.11 / **0.84**	2.07 / 36.91 / **0.89**	1.94 / 39.32 / 0.92	8.06 / **34.11** / **0.88**	1.80 / 41.37 / **0.95**
	16×	3.45 / 31.16 / 0.78	5.72 / 26.87 / 0.73	3.75 / 31.01 / 0.76	8.74 / 25.84 / 0.59	4.29 / 30.14 / 0.71	4.03 / 32.54 / 0.79	17.98 / 26.47 / 0.68	3.79 / 34.52 / 0.84
SRFormer	4×	1.56 / 38.44 / **0.93**	**1.72** / **36.49** / **0.94**	1.76 / 37.97 / **0.92**	4.04 / 33.12 / **0.84**	2.06 / 36.95 / **0.89**	1.93 / 39.36 / **0.93**	8.03 / 33.94 / **0.88**	1.79 / 41.42 / **0.95**
	16×	3.41 / 31.25 / **0.79**	5.39 / 26.88 / **0.74**	3.68 / 31.16 / 0.76	8.53 / 26.00 / 0.59	4.20 / 30.29 / **0.72**	3.95/ 32.70 / 0.79	17.49 / 26.55 / 0.69	3.72 / 34.67 / 0.84
Real ESRGAN	4×	3.35 / 32.26 / 0.82	4.87 / 28.50 / 0.79	3.83 / 31.58 / 0.79	9.09 / 26.24 / 0.63	4.53 / 30.35 / 0.74	4.21 / 32.87 / 0.81	18.40 / 26.80 / 0.72	3.89 / 35.03 / 0.86
	16×	5.58 / 27.70 / 0.70	9.16 / 23.26 / 0.64	6.19 / 27.39 / 0.68	14.41 / 22.21 / 0.50	7.14 / 26.46 / 0.64	6.69 / 28.90 / 0.72	29.33 / 22.65 / 0.64	6.26 / 30.90 / 0.77
ResShift	4×	5.19 / 28.57 / 0.70	5.18 / 26.87 / 0.74	5.93 / 27.94 / 0.65	15.09 / 20.89 / 0.43	6.92 / 26.86 / 0.59	6.47 / 29.33 / 0.68	34.52 / 17.2 / 0.54	6.01 / 31.44 / 0.75
	16×	8.86 / 24.26 / 0.57	8.88 / 22.48 / 0.61	9.86 / 23.87 / 0.54	25.62 / 16.52 / 0.32	11.04 / 23.16 / 0.50	10.47 / 25.51 / 0.59	64.72 / 12.37 / 0.44	9.89 / 27.45 / 0.65
Stable SR	4×	3.65 / 30.78 / 0.78	5.09 / 26.67 / 0.74	3.96 / 30.63 / 0.75	9.82 / 24.02 / 0.57	4.71 / 29.24 / 0.69	4.33 / 31.99 / 0.77	20.95 / 22.27 / 0.66	4.02 / 34.13 / 0.83
	16×	6.07 / 26.08 / 0.67	8.23 / 23.22 / 0.63	6.30 / 26.37 / 0.65	14.51 / 20.86 / 0.47	6.92 / 25.83 / 0.60	6.60 / 28.14 / 0.68	31.19 / 19.67 / 0.59	6.34 / 29.92 / 0.74

(b) Performance of super-resolution methods evaluated within manually segmented leaf regions.

SR Technique	Scale	DVI	GNDVI	MSAVI	MSR	NDVI	OSAVI	RVI	SAVI
		MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM	MAE / PSNR / SSIM
Bicubic	4×	1.83 / 38.58 / 0.93	2.90 / 33.09 / 0.92	1.73 / 39.04 / 0.93	4.60 / 27.57 / 0.79	1.71 / 39.05 / 0.93	1.70 / 41.08 / 0.95	11.60 / 27.30 / 0.79	1.71 / 42.67 / 0.96
	16×	6.34 / 28.59 / 0.71	14.46 / 21.07 / 0.53	5.53 / 29.72 / 0.74	13.46 / **18.85** / 0.37	5.06 / **30.34** / **0.75**	5.20 / 32.11 / **0.80**	33.66 / 18.65 / 0.37	5.43 / 33.39 / 0.83
Nearest Neighbor	4×	2.73 / 34.64 / 0.87	5.02 / 28.18 / 0.83	2.51 / 35.36 / 0.88	6.32 / 24.51 / 0.70	2.39 / 35.75 / 0.88	2.41 / 37.62 / 0.90	15.78 / 24.44 / 0.70	2.47 / 39.04 / 0.92
	16×	7.34 / 26.99 / 0.62	16.34 / 19.77 / 0.43	6.48 / 27.95 / 0.65	15.38 / 17.41 / 0.27	5.90 / 28.62 / 0.67	6.08 / 30.37 / 0.73	37.93 / 17.43 / 0.28	6.35 / 31.65 / 0.77
HAT	4×	1.59 / 39.86 / **0.95**	2.41 / 35.43 / **0.95**	1.56 / 39.99 / **0.95**	4.29 / 28.22 / **0.82**	1.59 / 39.77 / **0.94**	1.56 / 41.90 / **0.96**	10.88 / 27.91 / **0.82**	1.55 / 43.62 / **0.97**
	16×	**5.30** / **29.80** / **0.76**	9.99 / 23.50 / **0.63**	**4.78** / **30.57** / **0.76**	**13.25** / 18.43 / **0.41**	**4.80** / 30.32 / 0.74	**4.70** / **32.57** / **0.80**	**34.64** / 17.71 / **0.41**	**4.71** / **34.24** / **0.85**
HiT-SR	4×	**1.54** / **40.07** / **0.95**	2.32 / 35.73 / 0.95	**1.54** / 40.04 / **0.95**	4.30 / 28.13 / 0.81	1.59 / 39.70 / **0.94**	1.55 / 41.88 / **0.96**	10.90 / 27.81 / 0.81	**1.52** / 43.66 / **0.97**
	16×	5.49 / 29.59 / 0.74	10.71 / 23.40 / 0.62	5.13 / 30.07 / 0.74	14.48 / 17.87 / 0.38	5.20 / 29.81 / 0.72	5.07 / 32.04 / 0.79	37.92 / 17.12 / 0.38	5.04 / 33.75 / 0.83
SRFormer	4×	**1.54** / 40.06 / **0.95**	**2.16** / **36.31** / **0.95**	**1.54** / **40.10** / **0.95**	**4.25** / **28.26** / **0.82**	**1.58** / **39.82** / **0.94**	**1.54** / **41.97** / **0.96**	**10.77** / **27.95** / **0.82**	**1.52** / **43.72** / **0.97**
	16×	5.39 / 29.71 / 0.75	**9.85** / **23.67** / **0.63**	4.87 / 30.42 / **0.76**	13.31 / 18.48 / 0.40	4.83 / 30.31 / 0.74	4.75 / 32.48 / **0.80**	**34.64** / **17.85** / 0.40	4.79 / 34.11 / 0.84
Real ESRGAN	4×	4.61 / 30.93 / 0.83	6.71 / 26.69 / 0.79	4.77 / 30.70 / 0.81	13.78 / 18.45 / 0.45	5.17 / 30.06 / 0.77	4.94 / 32.38 / 0.83	34.78 / 17.99 / 0.45	4.73 / 34.31 / 0.87
	16×	11.05 / 23.87 / 0.61	16.96 / 19.41 / 0.45	11.17 / 23.90 / 0.62	29.01 / 12.59 / 0.21	11.72 / 23.69 / 0.61	11.38 / 25.80 / 0.68	69.34 / 12.49 / 0.22	11.09 / 27.52 / 0.73
ResShift	4×	7.87 / 26.42 / 0.69	8.42 / 25.24 / 0.73	8.04 / 26.23 / 0.66	26.67 / 11.96 / 0.30	8.23 / 26.03 / 0.64	8.06 / 28.16 / 0.71	82.00 / 8.32 / 0.28	7.91 / 29.91 / 0.76
	16×	12.62 / 22.45 / 0.50	15.71 / 20.29 / 0.46	12.75 / 22.36 / 0.49	47.41 / 6.75 / 0.12	12.80 / 22.37 / 0.48	12.55 / 24.48 / 0.57	183.33 / 0.82 / 0.11	12.40 / 26.15 / 0.63
Stable SR	4×	6.23 / 28.46 / 0.74	9.43 / 23.76 / 0.67	5.79 / 29.03 / 0.72	17.65 / 16.28 / 0.37	6.02 / 28.65 / 0.68	5.80 / 30.94 / 0.76	49.13 / 14.94 / 0.37	5.71 / 32.69 / 0.81
	16×	11.88 / 23.03 / 0.53	16.02 / 19.37 / 0.42	10.75 / 23.77 / 0.54	29.34 / 11.90 / 0.19	10.37 / 23.85 / 0.54	10.43 / 25.87 / 0.61	82.94 / 10.35 / 0.20	10.65 / 27.38 / 0.66

The VI reconstruction results in Table 4b follow a similar pattern. Non-parametric Kruskal-Wallis tests confirmed statistically significant performance differences across models for all eight vegetation indices at both 4× ($H(7) \in [338.31, 593.72]$, $p < 0.001$) and 16× ($H(7) \in [101.41, 420.34]$, $p < 0.001$) scale factors. Post-hoc Dunn's tests with Bonferroni correction verified that Transformer architectures (HAT, HiT-SR, and SRFormer) significantly outperformed GAN- and diffusion-based baselines across all VI metrics ($p < 0.001$), while pairwise differences within the Transformer family were not statistically significant. At 4×, soil-adjusted indices (SAVI and OSAVI) were the best reconstructed indices within leaf regions. SRFormer achieved a PSNR of 43.72 dB (resp. 41.97 dB) and an SSIM of 0.97 (resp. 0.96) for SAVI (resp. OSAVI). HiT-SR and HAT produce comparable results for both indices, with HiT-SR achieving 43.66 dB (SAVI) and 41.88 dB (OSAVI), and HAT achieving 43.62 dB (SAVI) and 41.90 dB (OSAVI). In contrast, ratio-based indices such as MSR and RVI exhibited larger reconstruction errors and lower structural similarity for all the methods. At 16×, reconstruction quality decreased for all vegetation indices. HAT achieved the highest SAVI performance, with an MAE of 4.71, a PSNR of 34.24 dB, and an SSIM of 0.85. HAT also achieved the best OSAVI results, with an MAE of 4.70, a PSNR of 32.57 dB, and an SSIM of 0.80. SRFormer and HiT-SR produced similar, but slightly lower, results for these two indices. Reconstruction was weaker for ratio-based indices, most notably RVI. ResShift had the lowest RVI reconstruction quality at 16×, with an MAE of 183.33, a PSNR of 0.82 dB, and an SSIM of 0.11.

4 Discussion

The experimental results reveal key insights about applying RGB-pretrained SISR models to multispectral crop images in terms of spatial reconstruction quality and spectral consistency.

1. Behavior of super-resolution methods in multispectral reconstruction. Non-parametric Kruskal–Wallis tests followed by Dunn's post-hoc tests with Bonferroni correction confirmed that Transformer-based methods (SRFormer, HiT-SR, and HAT) significantly outperform GAN and diffusion-based approaches across both whole-image and leaf-level evaluations. This improvement is most evident in spectral metrics (SAM and ERGAS), indicating better preservation of inter-band relationships in RGN images compared to generative models. Although SRFormer, HiT-SR, and HAT achieve comparable performance overall, SRFormer shows slightly more stable behavior across scaling factors. Pairwise comparisons revealed no statistically significant differences among SRFormer, HiT-SR, and HAT, placing all three transformer architectures within the same highest-performing statistical category. In contrast, GAN-based (RealESRGAN) and diffusion-based models (ResShift, StableSR) produce visually sharper reconstructions but exhibit higher spectral distortion. This reflects their bias toward perceptual realism rather than radiometric accuracy. Bicubic interpolation remains a strong baseline in several cases and achieves competitive PSNR at high scaling factors compared to more complex pretrained models. It thus remains a robust, simple reference method for RGN-image reconstruction.

2. Impact of the upscaling factor on the quality of reconstruction. The scaling factor has a substantial impact on all the evaluated methods. At 4×, transformer-based models (SRFormer, HiT-SR, and HAT) significantly outperform classical interpolation methods (bicubic and nearest-neighbor) across both whole-image and leaf-level evaluations ($p < 0.001$). However, at 16×, the performance gap narrows considerably, with bicubic interpolation achieving statistically comparable PSNR and ERGAS values to Transformer models in several cases. This reflects the severe loss of fine image detail, making accurate reconstruction difficult regardless of model type. The performance gap between learned SISR methods and bicubic interpolation becomes much smaller in terms of PSNR and SSIM. This indicates that when the input resolution is highly limited, there is insufficient information to reliably reconstruct fine details and, consequently, reconstruction is likely driven by simple interpolation behavior rather than learned model priors. As a result, increasing model complexity provides limited additional benefit, highlighting a fundamental limitation of SISR for multispectral images under extreme downsampling. We emphasize that the evaluated SISR models were originally designed for RGB

images, which may further influence their performance when applied to multispectral data without fine-tuning.

3. Spectral consistency and vegetation index stability. The results highlight that spectral preservation is an important aspect of multispectral SR for vegetation analysis. Transformer-based approaches consistently achieved lower SAM and ERGAS values, indicating stronger preservation of spectral relationships and radiometric consistency. This advantage translated directly to VI reconstruction, where transformer-based methods obtained better performance across most indices. However, reconstruction reliability also depends on the VI formulation. Soil-adjusted indices such as SAVI and OSAVI remained comparatively robust, while ratio-based indices such as RVI and MSR were considerably more sensitive to reconstruction errors, especially at 16×. Since ratio formulations amplify small inter-band deviations, even minor spectral inaccuracies propagated into disproportionately larger index errors. This sensitivity persisted even in leaf-level evaluation, suggesting that the observed instability originates primarily from spectral error amplification rather than heterogeneous backgrounds. These findings confirm that visually convincing SR outputs alone are insufficient for quantitative agricultural analysis, especially when spectral relationships are not preserved.

4. Generalization to multispectral imagery and methodological limitations. Although the evaluated SR methods were originally developed for RGB imagery, we directly applied them to RGN multispectral data without domain-specific fine-tuning. The key challenge is to enhance spatial detail while preserving spectral information and reliable vegetation-index analysis for crop characterization and phenotyping. The evaluated methods showed varying performance across architectures and evaluation conditions, with transformer-based architectures exhibiting the strongest overall performance. We acknowledge some limitations of our study. The evaluated methods were used directly without multispectral fine-tuning, which likely constrained the reconstruction performance, especially under extreme scaling factors. We also used RGN images acquired with a single acquisition configuration. Considering additional spectral bands, sensor types, or crop conditions could have led to different results. In addition, our imagery was acquired at ground level using a horizontal or near-horizontal viewing perspective, whereas multispectral remote sensing imagery is commonly acquired from a nadir-view perspective. Due to equipment limitations, we were unable to acquire comparable nadir-view imagery, limiting our ability to evaluate the performance of SR methods under different viewing perspectives. Differences in perspective, depth-dependent scale, leaf orientation, and occlusion may influence the spatial characteristics of the imagery and the generalization of pretrained SR models. Although leaf-level evaluation provided biologically meaningful analysis, manual segmentation is not scalable and may introduce annotation uncertainty. Since leaf morphology is irregular and masked regions are not perfectly rectangular, residual background pixels may remain during annotation, masking, or reconstruction padding, introducing minor background interference into leaf-level evaluation. These limitations may affect the reliability and generalizability of SR-based spatial and spectral measurements for crop characterization and phenotyping. Future work will compare SISR models developed for multispectral and hyperspectral imagery with the RGB-based models evaluated in this study. These findings highlight the need for future approaches that better preserve spectral relationships while enhancing spatial detail, with the goal of improving the reliability of super-resolved imagery for crop characterization and phenotyping. Since the current imagery was acquired from a horizontal or near-horizontal viewing perspective, further investigation is needed to determine how viewing geometry affects reconstruction quality, crop characterization, and phenotyping outcomes compared with the nadir perspective commonly used in remote sensing. The analysis could also be extended beyond the RGN bands to assess whether model performance is consistent across spectral bands with different spectral responses. Testing additional sensors, viewing perspectives, crop types, and growth stages would provide a broader assessment of model robustness and the practical benefits and limitations of super-resolution for crop characterization and phenotyping.

5. Implications for precision agriculture applications. The results demonstrate that visual reconstruction quality is a poor guide to evaluate SR methods in agricultural remote sensing. For downstream pheno-

typing tasks, spectral consistency plays a role more important than perceptual realism. Transformer-based approaches provide the most balanced trade-off between spatial enhancement and spectral preservation, making them the most suitable among the evaluated methods for crop monitoring applications. However, at 16×, the observed degradation highlights a practical limitation, as extreme super-resolution can introduce instability that affects quantitative interpretation. This indicates that the reliability of SR outputs for agronomic analysis strongly depends on the degree of downsampling and the preservation of inter-band relationships.

5 Conclusions

We evaluated the applicability of eight super-resolution (SR) methods originally developed for RGB imagery to RGN multispectral crop images using both whole-image and leaf-level reconstruction analysis. Across both evaluation settings and scaling factors (4× and 16×), non-parametric Kruskal–Wallis tests followed by Dunn's post-hoc tests with Bonferroni correction confirmed that transformer-based approaches significantly achieved the strongest overall performance ($p < 0.001$), with HAT, HiT-SR, and SRFormer belonging to the same top-performing statistical tier with no significant pairwise differences. Bicubic interpolation remained a competitive baseline under moderate degradation, although its ability to recover fine structural information remained limited and its performance became statistically comparable to Transformers at extreme 16× scaling in several metrics. The findings show that the reliability of vegetation index reconstruction depends on the specific index: soil-adjusted indices (SAVI and OSAVI) were reconstructed more consistently, whereas ratio-based indices (RVI and MSR) exhibited substantially greater sensitivity to spectral distortions, especially for 16× scaling. Our results demonstrate that visual reconstruction quality is not a good proxy to assess SR methods for multispectral agricultural applications. Rather, preserving spectral relationships should be the determining factor for reliable quantitative analysis. The results indicate that transformer-based SR methods currently provide the most suitable balance between spatial enhancement and spectral accuracy for crop phenotyping tasks.

Supporting Information

Supporting Information is available from the corresponding author if required.

Acknowledgements

This work received support from the Agencia Nacional de Investigación y Desarrollo (ANID) under grant FONDECYT Iniciación, 11250090; ANID AC3E 250006.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors used generative AI tools, such as Google Gemini and ChatGPT, to refine the grammar and improve readability.

References

[1] J. Jiang, P. M. Atkinson, C. Chen, Q. Cao, Y. Tian, Y. Zhu, X. Liu, and W. Cao, "Combining UAV and Sentinel-2 satellite multi-spectral images to diagnose crop growth and N status in winter wheat at the county scale," *Field Crops Research*, vol. 294, p. 108860, 2023.

[2] P. R. Mwinuka, S. K. Mourice, W. B. Mbungu, B. P. Mbilinyi, S. D. Tumbo, and P. Schmitter, "UAV-based multispectral vegetation indices for assessing the interactive effects of water and nitrogen in irrigated horticultural crops production under tropical sub-humid conditions: A case of african eggplant," *Agricultural Water Management*, vol. 266, p. 107516, 2022.

[3] S. De Petris, F. Sarvia, F. Parizia, F. Ghilardi, A. Farbo, and E. Borgogno-Mondino, "Assessing mixed-pixels effects in vineyard mapping from satellite: A proposal for an operational solution," *Computers and Electronics in Agriculture*, vol. 222, p. 109092, 2024.

[4] H. Liu, Y. Qian, X. Zhong, L. Chen, and G. Yang, "Research on super-resolution reconstruction of remote sensing images: A comprehensive review," *Optical Engineering*, vol. 60, no. 10, pp. 100 901–100 901, 2021.

[5] C. Tian, Y. Zhang, W. Zuo, C.-W. Lin, D. Zhang, and Y. Yuan, "A heterogeneous group CNN for image super-resolution," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 5, pp. 6507–6519, 2022.

[6] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1905–1914.

[7] Y. Zhou, Z. Li, C.-L. Guo, S. Bai, M.-M. Cheng, and Q. Hou, "SRFormer: Permuted self-attention for single image super-resolution," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12 780–12 791.

[8] Z. Yue, J. Wang, and C. C. Loy, "ResShift: Efficient diffusion model for image super-resolution by residual shifting," *Advances in neural information processing systems*, vol. 36, pp. 13 294–13 307, 2023.

[9] V. A. Palan, S. Thakur, and N. Sumith, "Leveraging super-resolution technology in drone imagery for advanced plant disease diagnosis and prognosis," *IEEE Access*, 2025.

[10] K. Yamamoto, T. Togami, and N. Yamaguchi, "Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture," *Sensors*, vol. 17, no. 11, p. 2557, 2017.

[11] S. Wu, L. Deng, L. Guo, and Y. Wu, "Wheat leaf area index prediction using data fusion based on high-resolution unmanned aerial vehicle imagery," *Plant Methods*, vol. 18, no. 1, p. 68, 2022.

[12] M. Vella and J. F. C. Mota, "Robust single-image super-resolution via CNNs and TV-TV minimization," *IEEE Trans. Im. Proc.*, vol. 30, pp. 7830–7841, 2021.

[13] D. Hong, C. Li, N. Yokoya, B. Zhang, X. Jia, A. Plaza, P. Gamba, J. A. Benediktsson, and J. Chanussot, "Hyperspectral imaging," *Nature Reviews Methods Primers*, vol. 6, no. 1, p. 19, 2026.

[14] J. Li, K. Zheng, L. Gao, L. Ni, M. Huang, and J. Chanussot, "Model-informed multistage unsupervised network for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.

[15] J. Li, K. Zheng, L. Gao, Z. Han, Z. Li, and J. Chanussot, "Enhanced deep image prior for unsupervised hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1–18, 2025.

[16] J. Li, K. Zheng, Z. Li, L. Gao, and X. Jia, "X-shaped interactive autoencoders with cross-modality mutual learning for unsupervised hyperspectral image super-resolution," *IEEE transactions on geoscience and remote sensing*, vol. 61, pp. 1–17, 2023.

[17] C. Li, D. Hong, B. Zhang, Z. Pan, N. Yokoya, and J. Chanussot, "Mambax: Image super-resolution with state predictive control," *arXiv preprint arXiv:2511.18028*, 2025.

[18] D. Hong, B. Zhang, X. Li, Y. Li, C. Li, J. Yao, N. Yokoya, H. Li, P. Ghamisi, X. Jia *et al.*, "Spectralgpt: Spectral remote sensing foundation model," *IEEE transactions on pattern analysis and machine intelligence*, vol. 46, no. 8, pp. 5227–5244, 2024.

[19] S. K. Phang, T. H. A. Chiang, A. Happonen, and M. M. L. Chang, "From satellite to UAV-based remote sensing: A review on precision agriculture," *Ieee Access*, vol. 11, pp. 127 057–127 076, 2023.

[20] R. Fattal, "Image upsampling via imposed edge statistics," in *ACM SIGGRAPH 2007 papers*, 2007, pp. 95–es.

[21] X. Chen, X. Wang, W. Zhang, X. Kong, Y. Qiao, J. Zhou, and C. Dong, "HAT: Hybrid attention transformer for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[22] X. Zhang, Y. Zhang, and F. Yu, "HiT-SR: Hierarchical transformer for efficient image super-resolution," in *European conference on computer vision*. Springer, 2024, pp. 483–500.

[23] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, "Exploiting diffusion prior for real-world image super-resolution," *International Journal of Computer Vision*, vol. 132, no. 12, pp. 5929–5949, 2024.

[24] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–10, 2022.

[25] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.

[26] W. Atlas, "Climate and monthly weather forecast Graneros, Chile," https://www.weather-atlas.com/en/chile/graneros-climate, 2026, accessed: 2026-05-12.

[27] Open-Meteo, "Accurate weather forecasts for any location," https://open-meteo.com/, 2026, accessed: 2026-05-12.

[28] "Process Tab — MAPIR Camera Control (MCC) — mapir.gitbook.io," https://mapir.gitbook.io/mapir-camera-control-mcc/interface-tabs/process-tab, [Accessed 27-07-2025].

[29] G. Bradski, "The opencv library." *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[30] "labelme-ytu," image annotation tool (LabelMe fork), accessed 2026-05-05. [Online]. Available: https://pypi.org/project/labelme-ytu/

[31] D. Khaledyan, A. Amirany, K. Jafari, M. H. Moaiyeri, A. Z. Khuzani, and N. Mashhadi, "Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution," in *2020 IEEE Global Humanitarian Technology Conference (GHTC)*. IEEE, 2020, pp. 1–5.

[32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[33] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 114–125.

[34] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 126–135.

[35] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 606–615.

[36] Z. Yue, J. Wang, and C. C. Loy, "Efficient diffusion model for image restoration by residual shifting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 1, pp. 116–130, 2024.

[37] S. Gu, A. Lugmayr, M. Danelljan, M. Fritsche, J. Lamour, and R. Timofte, "Div8k: Diverse 8k resolution image dataset," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, 2019, pp. 3512–3516.

[38] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 4401–4410.

[39] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3365–3387, 2020.

[40] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[41] Z. Lu and Y. Chen, "Single image super-resolution based on a modified U-net with mixed gradient loss," *signal, image and video processing*, vol. 16, no. 5, pp. 1143–1151, 2022.

[42] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote sensing of Environment*, vol. 8, no. 2, pp. 127–150, 1979.

[43] Y. He, J. Peng, F. Liu, C. Zhang, and W. Kong, "Critical review of fast detection of crop nutrient and physiological information with spectral and imaging technology," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 31, no. 3, pp. 174–189, 2015.

[44] A. A. Gitelson, Y. J. Kaufman, R. Stark, and D. Rundquist, "Novel algorithms for remote estimation of vegetation fraction," *Remote sensing of Environment*, vol. 80, no. 1, pp. 76–87, 2002.

[45] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote sensing of environment*, vol. 55, no. 2, pp. 95–107, 1996.

[46] A. Bannari, D. Morin, F. Bonn, and A. Huete, "A review of vegetation indices," *Remote sensing reviews*, vol. 13, no. 1-2, pp. 95–120, 1995.

[47] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote sensing of environment*, vol. 25, no. 3, pp. 295–309, 1988.

[48] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from eos-modis," *Remote sensing of Environment*, vol. 58, no. 3, pp. 289–298, 1996.

[49] C. Wu, Z. Niu, Q. Tang, and W. Huang, "Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation," *Agricultural and forest meteorology*, vol. 148, no. 8-9, pp. 1230–1241, 2008.

[50] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *JPL, summaries of the third annual JPL airborne geoscience workshop. volume 1: AVIRIS workshop*, 1992.

[51] H. Irmak, G. B. Akar, and S. E. Yuksel, "A map-based approach for hyperspectral imagery super-resolution," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2942–2951, 2018.

[52] T. Pohlert, "The pairwise multiple comparison of mean ranks package (pmcmr)," *R package*, vol. 27, no. 2019, p. 9, 2014.