



Smart UAV-assisted blueberry maturity monitoring with Mamba-based computer vision

Fan Zhao¹ · Yinyin He² · Jian Song¹ · Jiaqi Wang¹ · Dianhan Xi¹ · Xinlei Shao¹ · Qingyang Wu³ · Yongying Liu¹ · Yijia Chen¹ · Guochen Zhang¹ · Chenyu Zhang⁴ · Yulun Chen⁵ · Jundong Chen¹ · Katsunori Mizuno¹

Accepted: 16 May 2025 / Published online: 3 June 2025
© The Author(s) 2025

Abstract

Purpose Precise segmentation of blueberry maturity is critical for optimizing harvest schedules and maintaining product quality. Traditional methods, which rely on manual inspection, are not only labor-intensive but also cost-inefficient. This study presents a novel framework that integrates deep learning-based super-resolution reconstruction (SRR) with semantic segmentation to provide a fast and accurate solution for maturity assessment.

Methods The SRR module enhances image resolution, enabling more detailed feature extraction. Semantic segmentation models—incorporating convolutional neural networks (CNNs), Transformer-based models, and the Mamba-based state space architecture—further improve segmentation precision.

Results Experimental results indicate that the MambaIR model achieves a structural similarity index measure (SSIM) of 82.26% in SRR tasks, while the Mamba-based segmentation model attains a mean Intersection over Union (mIoU) of 83.15%.

Conclusion By uniting SRR and semantic segmentation, our framework not only advances the technical accuracy of maturity detection but also holds strong potential for real-time, cost-effective deployment in precision agriculture systems, supporting intelligent decision-making at scale.

Keywords Blueberry maturity monitoring · Mamba-based deep learning · Semantic segmentation · Super-resolution reconstruction · UAV imagery

Introduction

Vaccinium corymbosum, commonly referred to as the blueberry, is a small blue fruit widely recognized for its low-calorie yet nutrient-dense composition. Blueberries are rich in anthocyanins, which are known to support retinal health and prevent eye fatigue and myopia. Additionally, their antioxidant properties have been shown to slow aging, enhance memory,

Yinyin He, Jian Song and Jiaqi Wang contributed equally to this work.

Extended author information available on the last page of the article

and reduce the risk of certain cancers (Rowland et al., 2012; Silva et al., 2020). Due to their high nutritional value and appealing flavor, blueberries have gained prominence as a global superfood. This rising demand has led to the proliferation of blueberry plantations, boosting both global production and consumption (Banerjee et al., 2020; Bauer & Visuals, 2008; Li et al., 2021a, b). Their nutritional benefits and market potential have solidified blueberries as a key cash crop in global agriculture.

Global cultivation of blueberries is largely driven by their rich nutritional profile (Bauer & Visuals, 2008). Traditionally, blueberry harvesting relies heavily on manual labor (Prasad et al., 2018). However, this process is both labor-intensive and time-consuming. Maturity assessment of blueberries has conventionally depended on the experience of the picker and visual inspection, which are prone to subjective biases. Accurate assessment of crop maturity is essential for producers, as blueberries, unlike fruits such as apples and bananas, do not continue to ripen after harvest and exhibit a firm, tart texture when immature (Kader, 1997). Immature blueberries are not suitable for sale, while overripe berries soften and rot quickly, reducing their shelf life. Consequently, precise determination of blueberry maturity is critical for maximizing yield and profitability. These challenges have prompted the exploration of automated, scalable monitoring solutions, particularly those leveraging computer vision and UAV-assisted technologies to replace or augment manual assessment.

Advances in both hardware and software technologies, coupled with growing market demand, have led to the development of mechanized harvesting solutions, such as self-propelled carts (Prasad et al., 2018), over-the-row (OTR) machines (Takeda et al., 2017), and semi-mechanical equipped with hand-held shakers (Kim et al., 2018), are used to assist manual harvesting. While these innovations have enhanced picking efficiency, they continue to face challenges in accurately assessing blueberry maturity, particularly due to environmental factors like leaf shading and the small size of the fruit. To address these limitations, computer vision algorithms have been introduced, improving the precision and efficiency of maturity detection in agricultural products.

Moltó et al. (1992) were among the first to implement a visual localization system for estimating the number of mature citrus fruits through image analysis of their reflectance spectra during harvest. Similarly, Mendoza and Aguilera (2004) employed a computer vision system to classify the maturity of various banana varieties based on color, spotting, and texture, achieving a 98% accuracy rate, thereby demonstrating the potential of computer vision in fruit maturity prediction. Building on this, Farooque et al. (2013) utilized digital color cameras on wild blueberry harvesters to quantify berry loss by comparing pre- and post-harvest yields. Kaur et al. (2018) further refined this approach by using discrete cosine transformation and other image processing techniques to analyze the color, texture, and size of plums, achieving a final error rate of less than 2.4% in determining their maturity class.

The rapid advancement of deep learning technologies has further revolutionized complex computer vision tasks. Convolutional neural networks (CNNs), which can automatically detect important image features without prior knowledge or human intervention, have become a cornerstone in deep learning applications (Alzubaidi et al., 2021). Numerous studies have explored CNN-based fruit maturity assessments. Anatyia et al. (2020) classified the maturity of five fruit types—popcorn, mango, melon, banana, and tomato—achieving a classification accuracy of 61% using a CNN model. Zhao et al. (2021) proposed a CNN-based algorithm for classifying melon maturity in complex environments, such as greenhouses. Aherwadi et al. (2022) developed a maturity classification model for bananas using three

datasets, concluding that CNN, particularly with AlexNet, is the most suitable deep learning method for detecting banana maturity and quality. Behera et al. (2021) employed Faster R-CNN to enhance model performance and modified the mean Intersection over Union (mIoU) to account for occluded fruit. In addition, the YOLO (You Only Look Once) algorithm, a CNN-based object detection framework, has been shown to significantly improve processing speed and accuracy compared to traditional R-CNN methods. Variants such as Mogo YOLO (Koirala et al., 2019), Apple YOLO (Tian et al., 2019), and Wild Blueberry YOLO (MacEachern et al., 2023) have been developed to estimate fruit maturity and yield.

Despite the significant progress achieved by deep learning techniques in assessing blueberry maturity, several critical technical challenges remain unresolved. Before these methods can be fully applied to field-scale blueberry monitoring, it is important to understand the limitations of their enabling platforms, such as UAVs. The performance of computer vision models for blueberry maturity assessment is highly dependent on the quality of images captured in diverse environmental conditions at the planting site (Liu et al., 2020). Unmanned aerial vehicles (UAVs) offer a compact, versatile solution by providing high-resolution, multispectral data, making them increasingly popular for monitoring fruits and other agricultural products (Radoglou-Grammatikis et al., 2020; Tsouros et al., 2019). UAVs can adjust flight altitudes based on different monitoring targets, offering distinct advantages over satellite-based remote sensing, such as reduced cloud and atmospheric interference, higher spatial resolution, and lower operational costs (Osco et al., 2021). Moreover, UAVs can efficiently and flexibly cover extensive farmland areas, enhancing both speed and cost control. During orchard surveillance, UAV-generated air thrust can also help uncover fruits concealed by dense foliage (Matese et al., 2015; Primicerio et al., 2012). However, challenges persist. Vibration from the UAV during image capture (Ellenberg et al., 2016) and the inability to closely approach the target often result in motion blur and reduced image resolution (Sieberth et al., 2014), which can obscure key information and complicate blueberry detection. These limitations necessitate advanced post-processing techniques to enhance image quality and support accurate downstream analysis.

To address the image degradation caused by UAV limitations, super-resolution reconstruction (SRR) is a computer vision technique aimed at mitigating image motion blur and enhancing image resolution (Park et al., 2003). SRR approaches generally rely on three categories of algorithms: interpolation-based, reconstruction-based, and machine learning-based algorithms (Li et al., 2020). Each class presents distinct limitations. Interpolation-based SRR algorithms generate pixel values by resizing images based on spatial pixel relationships. While computationally simple and efficient, these methods often fail to preserve fine details, leading to suboptimal reconstructions. Reconstruction-based algorithms utilize prior knowledge and mathematical models to generate high-resolution images but are less effective with texture-rich content. Traditional machine learning-based SRR models, which learn mappings between low- and high-resolution images from extensive datasets, offer better visual outcomes but are computationally demanding and challenging to optimize (Xiang et al., 2022). The introduction of the SRCNN algorithm marked a breakthrough in SRR, utilizing only three convolutional layers and an end-to-end training framework (Dong et al., 2016). This development paved the way for deep learning-based SRR. Subsequent innovations include EDSR (Lim et al., 2017), RCAN (Zhang et al., 2018), SRGAN (Ledig et al., 2017), Real-ESRGAN (Wang et al., 2021a, b), and MambaIR (Guo et al., 2024), all of which have been applied across various domains such as medical imaging (Hatvani et al.,

2019; Li et al., 2021a, b), facial recognition (Jiang et al., 2023), and remote sensing (Wang et al., 2022a, b; Zhao et al., 2024a, b).

Blueberry clusters present a unique challenge for maturity assessment, as individual berries within the same cluster may vary in ripeness, while others remain immature. This heterogeneity, compounded by background interference from dense foliage and branching, complicates accurate maturity evaluation. While object detection algorithms are capable of identifying individual blueberries, they are less efficient in assessing maturity across entire planting areas. To overcome these limitations, this study proposes a semantic segmentation approach, enabling the differentiation of blueberries by maturity level at the pixel level.

Existing research on fruit maturity assessment has predominantly employed ground-based vehicles (Bargoti & Underwood, 2017; Koirala et al., 2019) or near-infrared spectroscopy (Shah et al., 2020). Studies focused on blueberry maturity, particularly for wild and plantation-grown varieties, have largely concentrated on improving detection algorithms like YOLO (Liu et al., 2023; MacEachern et al., 2023; Zhao et al., 2025). Although multi-source data collection techniques have been explored, UAV-based data acquisition from blueberry fields and comparative analyses of SRR architectures in conjunction with semantic segmentation have yet to be fully investigated.

The objective of this study is to propose an ensemble approach for assessing the overall maturity of blueberry plantations using SRR and semantic segmentation models. First, a deep learning-based SRR training dataset is constructed, which is used to train SRR networks based on various deep learning algorithms. The quality of the reconstructed blueberry images is then evaluated using peak signal-to-noise ratio (PSNR) and structural Similarity Index (SSIM) metrics. Subsequently, a semantic segmentation network is trained using the original high-resolution blueberry dataset. The trained model is then applied to different test sets, and the results are compared and analyzed for overall performance.

Methodology

The methodology of this study is summarized in Fig. 1. Initially, high-resolution (HR) images of the blueberry plantation, captured with a consumer-grade UAV at a resolution of 512×512 pixels, are downsampled to create low-resolution (LR) images with a resolution of 128×128 pixels. Using a training dataset, the nonlinear mapping between these LR and HR images is learned by selected SRR models. Based on this mapping, SR images are reconstructed from the LR images. These reconstructed SR images are then segmented at the pixel level using a pre-trained blueberry segmentation model, which classifies and labels each pixel corresponding to the blueberry fruit. The performance of various SRR algorithms is then quantitatively analyzed using the segmentation results, and the distribution and ripeness of blueberries—both ripe and unripe—across the plantation are evaluated. Details of each step are provided in subsections [Data collection](#) through [Blueberry segmentation](#).

Survey site

The study was conducted in a blueberry plantation located in Chengjiang County, Yuxi City, China, as shown in Fig. 2. The region's subtropical monsoon climate, with mild winters, cool summers, and an average annual temperature of 15–20 °C, provides optimal growing

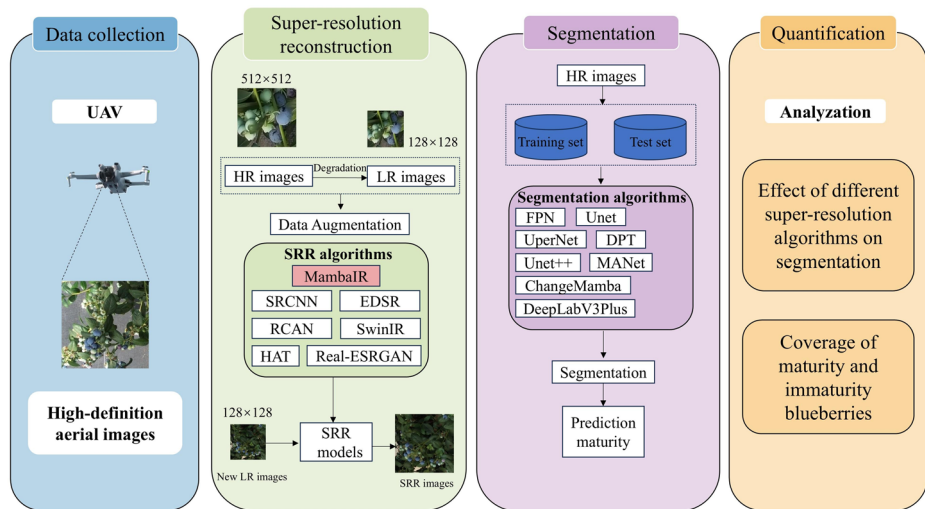


Fig. 1 Workflow of the proposed framework for blueberry maturity assessment. (Data Collection: High-resolution aerial images are collected using an unmanned aerial vehicle (UAV) over blueberry plantations. Super-resolution Reconstruction: The original high-resolution images (512×512) are downsampled to low-resolution (128×128) to train SRR models using seven different SRR algorithms. These models reconstruct low-resolution images to enhance spatial resolution. Segmentation: The reconstructed and original HR images are input into various semantic segmentation algorithms to predict pixel-wise maturity levels. Quantification: The segmentation results are analyzed to evaluate the effect of different SRR methods on segmentation performance and to quantify the coverage of mature and immature blueberries across the dataset.)

conditions for blueberries. The area receives 1,000 to 1,200 millimeters of precipitation annually, ensuring sufficient moisture for the crop. Chengjiang's average altitude of 1,800 m creates substantial diurnal temperature variation, which promotes the accumulation of sugars and anthocyanins in blueberries. The region's acidic to slightly acidic soil, rich in organic matter and minerals, further supports blueberry cultivation, as blueberries favor acidic soils.

Data collection

For data acquisition, a DJI Mini 3 drone was used, weighing 248 g and featuring a flight time of 38 min, a field of view (FOV) of 82.1° , and a maximum image resolution of 4000×3000 pixels. To optimize image quality, data collection took place under clear skies with light breezes. The drone flew at an altitude of 3 m, capturing the entire plantation in approximately 0.5 h. Although the dataset was acquired from a single plantation, it exhibits considerable diversity. Due to staggered harvesting practices, blueberries at various stages of maturity coexisted naturally, ensuring a wide range of color, shape, and density distributions. Additionally, the UAV was operated at varying distances and viewing angles across the field, introducing variability in image scale, occlusion, and illumination. Such diversity is essential for improving the robustness of machine learning models in agricultural settings and aligns with other UAV-based remote sensing applications (Muksimova et al., 2024; Zhao et al., 2024c, d).

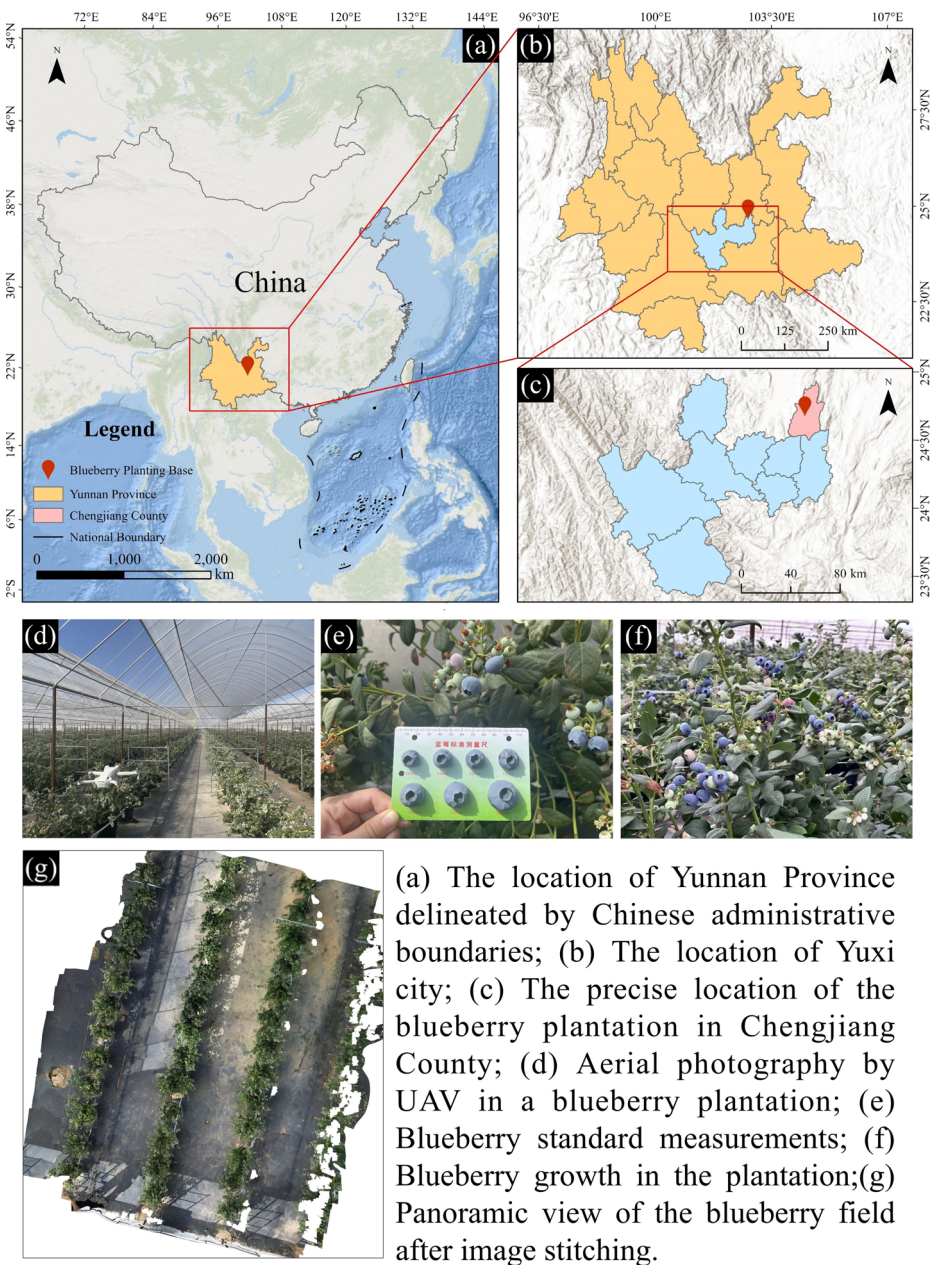


Fig. 2 Location map of blueberry plantation areas

Data preprocessing

During the data preprocessing stage, the UAV-captured images were cropped using a sliding window technique, where each window extracted a smaller section of the original image, producing a series of image segments. Each original high-resolution (HR) image had a resolution of 512×512 pixels, ensuring adequate detail for subsequent processing. From each original high-resolution image, 1,000 cropped images were generated. These HR images were then downsampled to 128×128 pixels using a degradation model, creating LR images. The image preprocessing phase also included data augmentation techniques such as rotating the images by 180 degrees, applying horizontal and vertical flips, and scaling the images by factors of 0.6, 0.7, 0.8, and 0.9. This augmented the dataset, enhancing the diversity of visual perspectives and expanding its size by a factor of 30. Parameters for window size and stride were optimized through preliminary experiments to strike a balance between computational efficiency and model performance. Specifically, a window size of 128 pixels and a stride of 32 pixels were used, as determined through preliminary experiments to balance detail capture and computational cost.

Super-resolution reconstruction

Deep learning-based SRR is utilized to enhance the quality of the LR images. To train the SRR model, a training dataset was prepared by downsampling the HR images with a degradation model. These LR images were used as input, with the HR images serving as the ground truth. Given that different SRR models employ various architectures and strategies for handling image features like edges and textures, seven deep learning-based SRR algorithms were selected for this study: Real-ESRGAN, SRCNN, EDSR, SwinIR, RCAN, HAT, and MambaIR. During training, the LR images were input into the SRR models, and a loss function was applied to quantify the difference between the predicted SR images and the actual HR images, thus optimizing the SRR models. After training, new LR images were processed by the trained SRR models to generate corresponding SR images. The quality of these SR images was evaluated using two metrics: Peak Signal-to-Noise Ratio (PSNR) and SSIM.

Architectures of SRR networks

Numerous researchers have developed various SRR models based on distinct network architectures. In this study, seven SRR models employing different network strategies were selected for comparison to achieve optimal blueberry maturity assessment, as illustrated in Fig. 3. In this diagram, “Conv” represents convolutional layers, and “Deconv” denotes deconvolutional layers.

The SRCNN model, one of the earliest SRR models based on deep learning, begins by applying bicubic interpolation to LR images. It then proceeds with three stages: feature extraction, nonlinear mapping, and reconstruction, followed by optimizing the loss between the interpolated and reference images. Although SRCNN’s simple architecture allows for fast computation, it struggles with capturing complex image features and demonstrates suboptimal performance when handling large datasets. The EDSR model improves upon SRCNN by removing the BatchNorm (BN) layers, which consume memory resources.

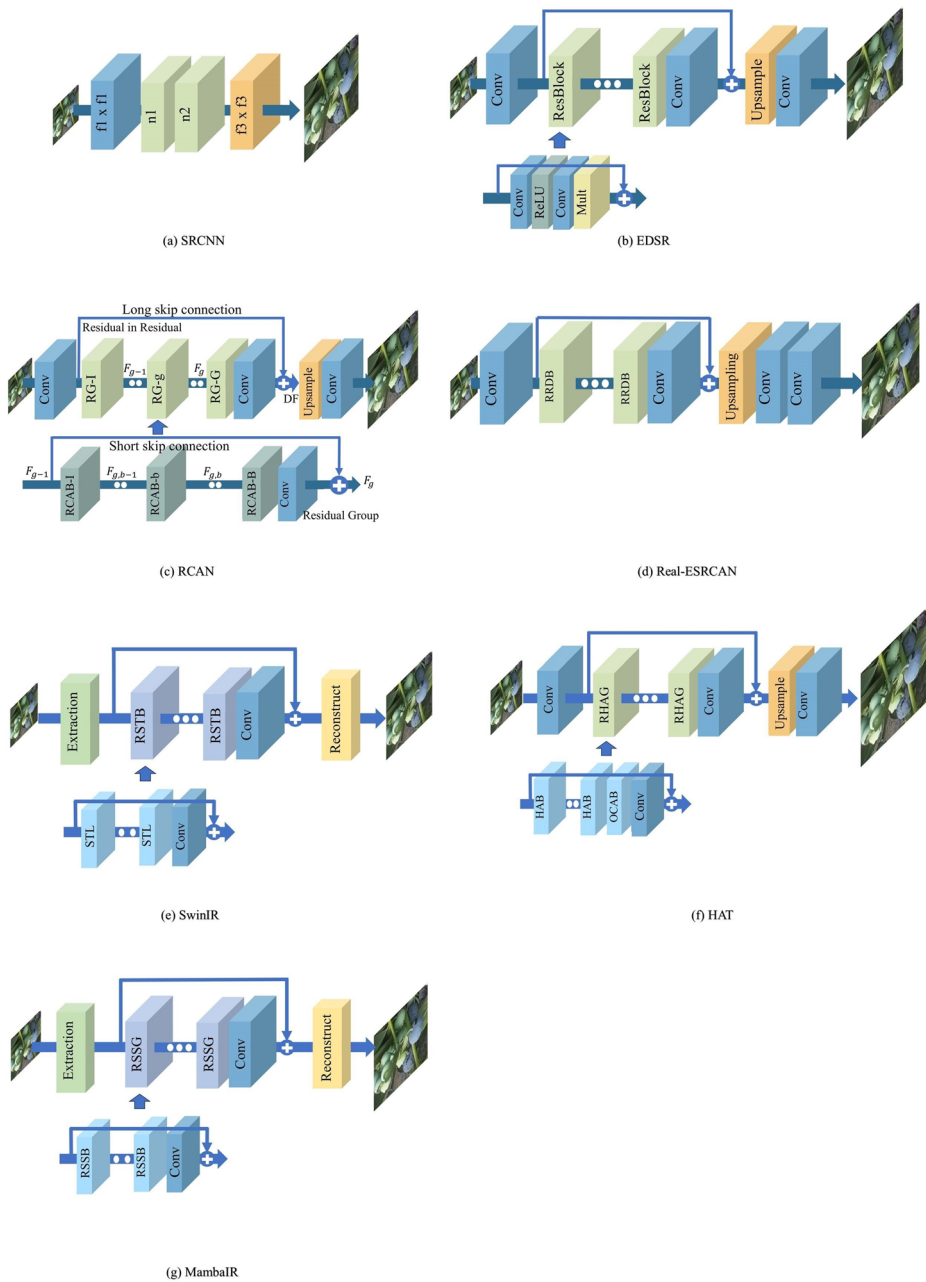


Fig. 3 Schematic depiction of SRR network architecture

This change allows more residual blocks to be stacked, thereby enhancing model performance. The RCAN model incorporates a channel attention mechanism with residual blocks, enabling it to adaptively readjust feature channels within a deep network, improving accuracy in complex image restoration tasks. Real-ESRGAN, an improvement over ESRGAN, integrates perceptual and adversarial losses, making it particularly effective at reconstructing images with complex textures and fine details, closely approximating real-world visuals. Transformer-based architectures have also been explored in SRR. Initially developed for natural language processing (Vaswani et al., 2017) Transformer architectures have since been adapted for computer vision tasks. SwinIR, introduced by, builds on the Swin Transformer (Liu et al., 2021) and utilizes a sliding window mechanism for self-attention within local image windows, efficiently capturing both local and global image features. HAT (Hybrid Attention Transformer) (Chen et al., 2023) combines multiple attention mechanisms, including channel attention and window-based self-attention, to capture global dependencies across layers for improved feature extraction and accurate image reconstruction. Finally, MambaIR, based on a selectively structured state space mode (Guo et al., 2024), leverages the global receptive field advantage to capture long-distance dependencies in images. Its Residual State Space Block (RSSB) enhances the model's ability to address issues like local pixel forgetting and channel redundancy.

Training of the networks

During training, iterative optimization is conducted to update model parameters based on the loss between each LR image input and the corresponding HR image. For CNN-based models, the Adam optimizer (Singarimbun et al., 2019) is used with the L1 loss function, as defined below:

$$L_1(O, R) = \frac{1}{pq} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} \|O_{i,j} - R_{i,j}\| \quad (1)$$

where $O_{i,j}$ and $R_{i,j}$ represent the pixel values at position (i, j) in the HR image and the SR image, respectively, and $p \times q$ is the image resolution.

In generative adversarial network (GAN)-based models, perceptual loss and adversarial loss are employed to enhance the realism of reconstructed images. Perceptual loss measures the difference between generated and target images based on high-level features extracted by a pre-trained network (e.g., VGG), ensuring that the SR images resemble real-world images beyond simple pixel-level accuracy:

$$L_{\text{perceptual}} = \sum_{i=1}^N \| \phi_i(I_{SR}) - \phi_i(I_{HR}) \|^2 \quad (2)$$

$$L_{\text{adv}} = -\log(D(G(z))) \quad (3)$$

$$L_D = -(\log(D(x)) + \log(1 - D(G(z)))) \quad (4)$$

here, ϕ_i denotes the feature map extracted by the pre-trained network at layer i , I_{SR} is the generated SR image, and I_{HR} is the target HR image. Adversarial loss is a core component in Generative Adversarial Networks (GANs) used to train the generator to make its

generated images as realistic as possible to fool the discriminator. For the generator, $G(z)$ represents the image generated by the generator, D is the discriminator, and z is the input noise or the low-resolution image. For the discriminator, x is the true image.

Evaluation metrics

The performance of the SRR models was evaluated using PSNR and SSIM. PSNR quantifies image similarity by comparing the mean squared error (MSE) between the original and reconstructed images, with a higher PSNR indicating closer resemblance. SSIM evaluates similarity by incorporating brightness, contrast, and structural information, with values closer to 1 signifying greater similarity (Wang et al., 2021a, b).

In the following Eq. (5), $R(i, j)$ represents the SRR image, while $O(i, j)$ denotes the original HR image with dimensions $p \times q$. PSNR, defined in Eq. (6), quantifies the similarity between the SRR and original HR images. Here, MAX_I represents the highest gray value within the image, conventionally set to 255. A higher PSNR value indicates greater image similarity. In Eq. (7), μ_O and σ_R represent the mean of the total pixels in the image O and the variances of image R . The covariance between O and R is denoted by σ_{OR} . Constants C_1 and C_2 are introduced to prevent division by zero in the denominator:

$$MSE = \frac{1}{pq} \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} [O(i, j) - R(i, j)]^2 \quad (5)$$

$$PSNR = 10 \times \lg_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (6)$$

$$SSIM(O, R) = \frac{(2\mu_O \mu_R + C_1)(2\sigma_{OR} + C_2)}{(\mu_O^2 + \mu_R^2 + C_1)(\sigma_O^2 + \sigma_R^2 + C_2)} \quad (7)$$

Blueberry segmentation

Network architecture

Image semantic segmentation, a fundamental task in computer vision, aims to assign a pre-defined category label to each pixel in an image (Csurka & Perronnin, 2011). Deep learning-based segmentation methods offer significant advantages by training on large pixel-labeled datasets, enabling neural networks to learn complex mappings between visual features and semantic labels. By leveraging scene information and high-level semantic features, these models can accurately interpret detailed image content, leading to a comprehensive understanding of diverse semantic categories (Zeiler & Fergus, 2014). In this study, eight semantic segmentation architectures were employed to assess blueberry ripeness across an entire plantation: FPN, U-Net, DeepLabV3Plus, Unet++, MANet, DPT, ChangeMamba, and UperNet.

FPN Feature Pyramid Networks (FPN) were introduced in 2017 to tackle multi-scale object detection by constructing pyramid features (Lin et al., 2017). Prior approaches relied solely

on final feature map outputs, where lower-level features excelled in spatial accuracy but lacked semantic richness, and upper-level features offered stronger semantics but weaker localization. FPN addresses this by connecting high-level, low-resolution features with low-level, high-resolution ones in a top-down manner. This integration enhances the detection of small objects without increasing computational load.

UperNet The design of UperNet builds upon FPN and the Pyramid Pooling Module (PPM). UperNet is explored for the task of unified perceptual parsing, aiming to help machine vision systems recognize as many visual concepts as possible from a given image. UperNet can learn the differences between various image datasets and perform joint reasoning (Xiao et al., 2018). Additionally, it is capable of predicting pixel-level texture labels using only image-level annotations.

U-Net U-Net is a deep learning-based convolutional neural network that consists of an encoder (down-sampling path) and a decoder (up-sampling path), forming a distinctive U-shaped architecture (Ronneberger et al., 2015). It consists of an encoder for down-sampling and a decoder for up-sampling, connected by skip connections, which retain spatial information for better localization. Despite its widespread use, U-Net has limitations, such as the uncertainty regarding optimal network depth and skip connections being restricted to the same scale.

Unet++ Unet++, an enhanced version, addresses these issues by efficiently integrating networks of varying depths and redesigning skip connections to aggregate multi-scale features. This redesign increases flexibility and improves inference speed (Zhou et al., 2018).

DeepLabV3Plus DeepLabV3Plus is another encoder-decoder architecture that uses atrous convolution and atrous separable convolution to expand the receptive field and extract multi-scale features (Chen et al., 2018). These features are fused, and a 1×1 convolution is applied to adjust channel numbers. However, DeepLabV3Plus faces challenges in computational efficiency when processing large images.

MA-Net MA-Net addresses these challenges by introducing the Position Attention Block (PAB), which leverages a self-attention mechanism to model the interdependencies between features in the spatial dimension, thus capturing pixel-to-pixel spatial dependencies across the entire image (Li et al., 2022). Additionally, MA-Net employs the Multi-Scale Fusion Attention Block (MFAB) to capture channel dependencies between feature maps. This

attention mechanism enables MA-Net to effectively capture rich contextual dependencies and adaptively integrate local features with global dependencies.

DPT DPT is a dense prediction network built on the Transformer architecture, which processes representations at a fixed high resolution and provides a global receptive field at each stage (Ranftl et al., 2021). The use of Transformers allows DPT to excel in dense prediction tasks by capturing long-range dependencies within the image.

ChangeMamba ChangeMamba, based on the VMamba architecture, introduces spatio-temporal state space models for image segmentation (Chen et al., 2024). It employs a cross-scanning module to process image blocks across multiple spatial directions simultaneously, effectively capturing long-range contextual information. This multi-dimensional approach enhances performance on high-resolution images while maintaining computational efficiency. ChangeMamba's structural design allows it to handle subtle variations in dynamic scenes, making it robust and adaptable for complex image tasks where high accuracy and efficiency are essential.

Evaluation metrics

To assess the performance of blueberry maturity segmentation, four prevalent metrics are utilized: Precision, Recall, F1-score, and Intersection over Union (IoU). The IoU metric quantitatively evaluates the extent of overlap between the predicted and actual results, with a higher IoU reflecting enhanced segmentation accuracy.

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (11)$$

$$IoU = \frac{TP}{TP + FN + FP} \quad (12)$$

Here, TP stands for true positives, i.e., pixels correctly segmented as targets; FP stands for false positives, i.e., non-target pixels mistaken for targets; and FN stands for false negatives, i.e., real targets mistaken for non-target pixels. In this study, both IoU and mean IoU (mIoU) were calculated to comprehensively evaluate the segmentation performance across three regions: the background, the mature part of the blueberry fruit, and the unripe part.

Result

Datasets and experiment setup

Datasets description

The SRR model in this study was trained using a specialized blueberry dataset, tailored specifically for the assessment of maturity levels. This dataset comprises 1,000 images, each with a resolution of 512×512 pixels, collected directly from a blueberry plantation to ensure the inclusion of domain-specific characteristics. These curated images cover a wide range of ripeness stages, providing diverse visual features essential for accurate super-resolution reconstruction.

The corresponding LR image set was created by applying a controlled degradation process to the original high-resolution HR images. This process simulated the types of blurring and resolution loss commonly observed in real-world UAV imagery. By mimicking realistic conditions, this approach enhanced the model's capacity to generalize across diverse image qualities and improved its robustness in both natural and agricultural environments.

$$g = (f \otimes h) \downarrow_s^{bicubic} + \eta \quad (13)$$

in this equation, g represents the LR image, while f denotes the HR image. The function h corresponds to the point spread function (PSF) modeled under uniform linear motion, describing the degradation process. The symbol $\otimes$ indicates the convolution operation applied between h and f . The downsampling operation is represented by $\downarrow$, which reduces the image resolution by a factor s (the magnification factor). The bicubic interpolation algorithm is employed to perform the downsampling operation, reducing the resolution of the image by a factor of s . Additionally, η represents Gaussian white noise, which is added to simulate random noise encountered during image acquisition.

To diversify the training dataset and increase the efficiency of the training process, LR sub-images of size $l_{sub} \times l_{sub}$ pixels are randomly extracted and paired with their corresponding HR sub-images, resized to $sl_{sub} \times sl_{sub}$ pixels, where s represents the magnification factor. These patch pairs provided training samples that capture varied spatial patterns and scale variations, thereby boosting reconstruction performance.

To simulate real-world scenarios of low-resolution and blurry imagery, the corresponding LR image set was generated by applying degradation techniques to the HR images. This degradation process allowed the model to generalize effectively across various image qualities, enhancing its robustness in both natural and agricultural environments. The approach ensures that the model performs well even under challenging conditions, such as suboptimal image captures during field operations.

Experimental setup

In this study, seven SRR models—Real-ESRGAN, SRCNN, EDSR, SwinIR, RCAN, HAT, and MambaIR—were selected to reconstruct low-resolution blueberry images. Each model brings distinct strengths in addressing challenges such as noise reduction, texture preservation, and high-resolution feature extraction. This selection allows for a comprehensive com-

parison across diverse architectural designs. The SRR process was conducted in three key phases: (1) Training: The seven SRR networks were trained on the same dataset, consisting of HR blueberry images. (2) Reconstruction: The trained models were used to reconstruct LR images from the test dataset, producing super resolution (SR) outputs. (3) Evaluation: The performance of each model was evaluated using Peak Signal-to-Noise Ratio (PSNR) and SSIM metrics to assess reconstruction quality.

For the semantic segmentation task, the procedure was structured as follows: (1) Training: The proposed semantic segmentation network was trained on the HR blueberry image dataset, using both training and validation sets. (2) Segmentation: The trained model was applied to perform pixel-wise semantic segmentation on the test set. (3) Evaluation: The segmentation results were quantitatively assessed using Intersection over Union (IoU) and mean Intersection over Union (mIoU) metrics to measure the accuracy of the blueberry maturity classification at the pixel level.

This dual approach ensures a thorough evaluation of both image reconstruction and semantic segmentation performance, enhancing the accuracy of blueberry maturity assessment.

To ensure reproducibility and fairness, all SRR models were trained under a unified configuration: 300 training epochs, a batch size of 16, an initial learning rate of 0.00005 (decayed by half every 50 epochs), and the L1 loss function optimized using the Adam algorithm. Training was performed on a local workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM) and 64 GB RAM, providing sufficient computational resources for high-resolution image reconstruction.

For the semantic segmentation task, training was conducted on a dedicated high-performance computing server equipped with a Tesla V100 GPU (32 GB), an Intel® Xeon® E5-2698 v4 CPU, and running Ubuntu 20.04. The environment utilized PyTorch 2.2.1 with CUDA 12.2 and cuDNN 8.8.0. The models were trained using 448×448 input images, a batch size of 8, and an initial learning rate of $1e-5$ for a total of 20,000 iterations. The Adam optimizer and cross-entropy loss were used across all segmentation networks. Encoder-decoder pairs such as UNet++, DPT, and ChangeMamba were initialized with pretrained weights, where available, to facilitate faster convergence and improved performance.

Inference for both SRR and segmentation tasks was conducted on the same local workstation used for SRR training (RTX 4090 GPU), ensuring a consistent environment for runtime benchmarking. A detailed summary of the training environments and configuration parameters for both tasks is provided in Table 1.

Analysis of the super-resolution reconstruction

This study employed deep learning-based SRR models, each leveraging different architectural strategies, to enhance the resolution of LR blueberry images. Table 2 presents the performance evaluation results of various SRR methods on a test set of LR images with a fourfold magnification factor. As shown in Table 1 and Fig. 4, most deep learning-based SRR methods outperform the traditional Bicubic interpolation method in terms of Peak Signal-to-Noise Ratio (PSNR) and SSIM, demonstrating superior image reconstruction capabilities. Notably, while Real-ESRGAN achieves a PSNR of 26.06 dB and an SSIM of 67.62%, slightly lower than those of the Bicubic method, this does not indicate poor performance. Real-ESRGAN incorporates a perceptual loss function, which is specifically designed to

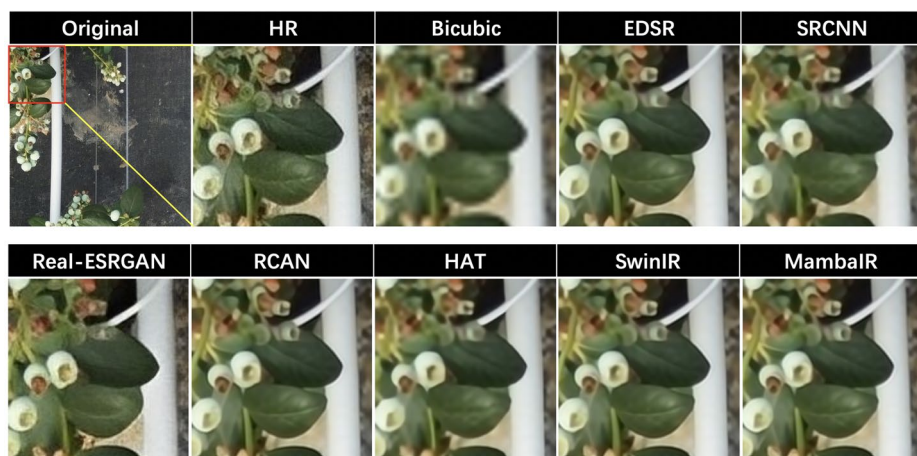
Table 1 Training configurations for SRR and semantic segmentation models

Experiment configuration	SRR Models	Segmentation Models
Operating system	Ubuntu 18.04	Ubuntu 20.04
Bit architecture	32 bit	32 bit
Random Access Memory	64GB	64GB
GPU	NVIDIA GeForce RTX 4090	Tesla V100
Memory	24GB	32GB
CPU	Intel(R) Xeon(R) Platinum i9-13900k	Intel(R) Xeon(R) E5-2698 v4
Pytorch	1.10	2.2.1
CUDA	11.1	12.2
Cudnn	8.0.4	8.8.0

Table 2 Evaluation metrics of different methods on the LR testsets with the magnification factor of 4

Metrics	Bicubic	Real-ESRGAN	SRCNN	EDSR	SwinIR	RCAN	HAT	MambaIR
PSNR(dB)	28.19	26.06	29.80	30.68	30.78	30.81	30.86	30.87
SSIM(%)	73.93	67.62	79.11	81.79	82.01	82.13	82.25	82.26

Bold values indicate the best results

**Fig. 4** Comparison of the visual effects of the reconstructed images based on the eight methods

prioritize high-frequency details and enhance texture realism (Johnson et al., 2016). This approach may result in lower PSNR and SSIM values, which focus on pixel-level accuracy, but it excels in generating visually realistic images that closely resemble natural scenes. Consequently, Real-ESRGAN is particularly advantageous in applications where perceptual quality is paramount. Among the SRR models evaluated, MambaIR exhibited the best performance, achieving a PSNR of 30.87 dB and an SSIM of 82.26%. The images reconstructed by MambaIR closely matched the original high-resolution images, underscoring the effectiveness of its selection mechanism and hardware-aware state-space architecture.

This superior performance highlights MambaIR's potential for high-fidelity image reconstruction tasks, particularly in agricultural contexts requiring detailed feature preservation for accurate maturity assessments.

Figure 4 illustrates the comparison between the original HR image and images reconstructed by various SRR methods. The Bicubic method shows a significant loss in image quality, particularly in edge clarity and texture detail, resulting in blurred lines and degraded visual features. In contrast, deep learning-based SRR methods exhibit superior performance in retaining edge sharpness and preserving texture details, yielding images that are visually closer to the original HR images. However, challenges remain. Figure 4 highlights that the SR images produced by the Bicubic method are the most blurred, reflecting the lowest perceptual quality. This observation is consistent with the Bicubic algorithm's performance in Table 2, where it ranks near the bottom in PSNR and SSIM metrics. Although Real-ESRGAN scores lowest in these quantitative measures, its SR images are visually superior, appearing sharper and with reduced noise compared to other methods. However, the details and textures generated by Real-ESRGAN still differ from those in the true HR images. Excluding Bicubic and Real-ESRGAN, other SRR methods produce images that closely match the original HR images in terms of detail and texture, indicating their effectiveness in generating high-quality SR images. Overall, qualitative visual assessments complement quantitative analyses, with MambaIR achieving an optimal balance between visual and quantitative quality.

Building on this analysis, MambaIR stands out for its significant global receptive field, allowing it to capture intricate details and preserve textures more effectively than other methods. This ability to reproduce complex features in blueberry images highlights the model's strengths, particularly for agricultural applications where texture and detail are crucial for assessing fruit maturity. Despite advancements in clarity and detail restoration, deep learning-based methods face inherent trade-offs between sharpness and realism, such as potential over-smoothing of edges. Nonetheless, these SRR models represent a substantial improvement over traditional methods like Bicubic, making them well-suited for blueberry image reconstruction.

Analysis of semantic segmentation

Single model performance

For the semantic segmentation task, the dataset captured by the UAV was divided into training, validation, and test sets at a 7:2:1 ratio. All images were meticulously labeled using the Labelme tool to ensure high-quality ground truth data. The study employed three encoder types—CNN, Transformer, and Mamba architectures—combined with eight different segmentation architectures: FPN, U-Net, DeepLabV3Plus, Unet++, MANet, DPT, Change-Mamba, and UperNet, resulting in 27 unique semantic segmentation models. Additionally, two ensemble models were constructed to further enhance performance. The performance is summarized in Table 3.

IoU (Intersection over Union) was used to evaluate segmentation accuracy across three target categories: ripe blueberries, unripe blueberries, and background. A higher IoU indicates better segmentation accuracy. The mean IoU (mIoU) results reveal that CNN-based encoders, particularly ResNeXt101_32 × 8 d and ResNet101 variants of Unet++, demon-

Table 3 Segmentation performance of various models

#	Decoder	Encoder	IoU Mature	IoU Immature	mIoU
<i>CNN-based Encoder</i>					
1	FPN	EfficientNet-b5	75.18	63.73	78.46
2	UNet	EfficientNet-b5	76.39	64.52	79.17
3	Deep-LabV3++	EfficientNet-b5	71.04	59.04	75.34
4	UNet++	EfficientNet-b5	77.49	64.91	79.69
5	FPN	ResNeXt101_32 × 8 d	77.95	66.06	80.26
6	UNet	ResNeXt101_32 × 8 d	78.64	66.16	80.53
7	Deep-LabV3+	ResNeXt101_32 × 8 d	75.96	64.38	78.95
8	UNet++	ResNeXt101_32 × 8 d	79.93	67.70	81.54
9	FPN	ResNet101	75.32	63.30	78.36
10	UNet	ResNet101	77.34	64.66	79.54
11	Deep-LabV3+	ResNet101	74.35	60.94	77.08
12	UNet++	ResNet101	78.82	66.39	80.64
13	MANet	EfficientNet-b5	73.82	63.98	78.08
14	MANet	ResNeXt101_32 × 8 d	77.71	66.20	80.24
15	MANet	ResNet101	76.21	64.48	79.12
<i>Transformer-based Encoder</i>					
16	FPN	SegFormer(mit_b5)	77.47	64.00	79.33
17	UNet	SegFormer(mit_b5)	75.49	64.50	78.88
18	MANet	SegFormer(mit_b5)	77.52	64.25	79.44
19	DPT	DINOv2(vit_1)	80.67	67.99	81.87
20	DPT	DINOv2(vit_b)	80.50	66.79	81.37
21	DPT	DINOv2(vit_s)	80.10	66.87	81.31
<i>Mamba-based Encoder</i>					
22	Change-Mamba	VMamba (base)	78.27	65.73	80.25
23	Change-Mamba	VMamba (tiny)	78.38	66.23	80.48
24	Change-Mamba	VMamba (small)	78.09	65.17	79.98
25	UperNet	VMamba (base)	77.92	64.38	79.61
26	UperNet	VMamba (tiny)	78.17	65.31	80.06
27	UperNet	VMamba (small)	77.53	64.33	79.47
<i>Ensemble Model</i>					
8 + 19 + 23 (ours)			81.82	70.29	83.13

Bold values indicate the best results

strated strong segmentation capabilities, achieving mIoU values of 81.54 and 80.64, respectively. ResNeXt101_32 × 8 d excelled in handling background complexity, outperforming other models in background classification accuracy.

In the Transformer-based experiments, DPT and its variants using the DINOv2 model family (DINOv2(vit_l), DINOv2(vit_b), DINOv2(vit_s)) achieved impressive results. DINOv2(vit_l), in particular, achieved the highest accuracy for ripe and unripe fruit classification tasks, with an mIoU of 81.87, surpassing other models in detailed fruit segmentation.

Additionally, the Mamba-based Encoder demonstrated strong performance, with all six tested models showing high robustness, maintaining mIoU values around 80. This consistent performance not only highlights the adaptability of the Mamba architecture across different semantic segmentation tasks but also suggests its significant potential for further optimization and extension.

By analyzing these models in detail, this study finds that networks based on different Encoders each have unique strengths in addressing the semantic segmentation of blueberry fruits. While CNN, Transformer, and Mamba architectures differ in their segmentation performance across various categories, overall, Transformer and Mamba architectures exhibit greater potential in handling high-complexity image details, whereas the CNN architecture excels in computational efficiency and background processing.

Based on the Precision, Recall, F1-score, mIoU, and Mature IoU metrics discussed in Sect. 2.5.3, this study compares the segmentation performance of the best-performing Ensembled fusion methods across different test sets. These include the HR test set, the test set generated using the Bicubic method, and the SR test sets produced by six other SRR models. As shown in Table 4, SRCNN delivers the highest Precision at 89.64%, while the HR test set achieves the best results in Recall, F1-score, mIoU, and Mature IoU, with values of 91.97%, 90.39%, 83.13%, and 81.82%, respectively. In contrast, the Bicubic test set exhibits the poorest performance across all evaluation metrics. Among the SRR methods, the five SR test sets show similar results, with MambaIR leading the group, delivering the best performance across the SR test sets, second only to the HR test set. The other SRR methods display comparable outcomes, further highlighting MambaIR's edge in segmentation accuracy.

Figure 5 visually compares the segmentation results for partially mature and immature blueberries, with the second column providing ground truth labels for reference. The HR test set demonstrates the most accurate segmentation, closely matching the ground truth, while the Bicubic method shows significant errors and omissions. SRR-based test sets markedly improved segmentation accuracy, with MambaIR performing particularly well. Figure 6 further highlights that all SRR-based test sets significantly outperform the Bicubic

Table 4 Segmentation results of the testsets reconstructed by different SRR models using ensemble segmentation model

Metrics (%)	HR	Bicubic	Real-ESRGAN	SRCNN	EDSR	SwinIR	RCAN	MambaIR
Precision	88.90	86.98	88.02	89.64	87.88	88.12	87.92	88.93
Recall	91.97	85.19	87.76	88.14	89.08	91.31	91.55	91.64
F1-score	90.39	87.45	87.88	88.85	89.00	89.66	89.66	89.67
mIoU	83.13	78.84	79.36	80.86	81.06	82.01	82.01	82.04
Mature IoU	81.82	78.22	77.09	80.27	80.30	80.76	80.78	80.90

Bold values indicate the best results

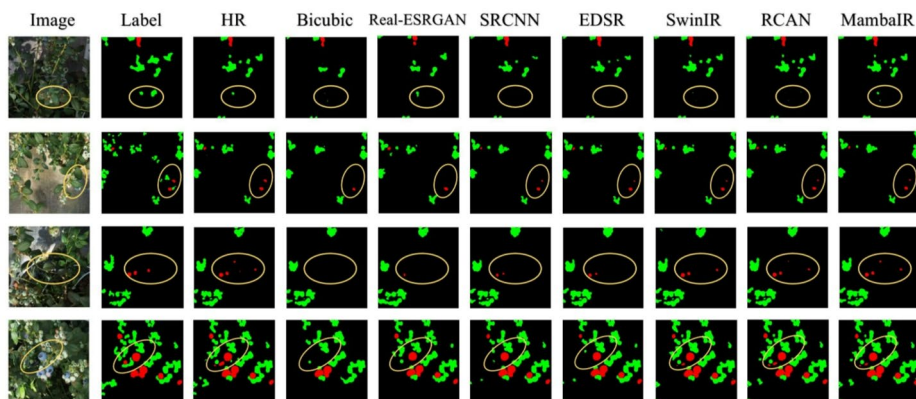


Fig. 5 Qualitative comparison of segmentation results across different SRR methods on sample test images. (Each column presents the segmentation output of a specific SRR method (e.g., SRCNN, Real-ESRGAN, MambaIR) followed by semantic segmentation using the same model. The first two columns show the original UAV image and its corresponding ground truth mask. The red regions represent mature blueberries, while green regions represent immature ones. Yellow circles highlight key regions of interest where differences in segmentation performance are most apparent—such as missed detections or over-segmentation. Models like MambaIR and SwinIR generally provide clearer and more accurate segmentation in these challenging areas.)

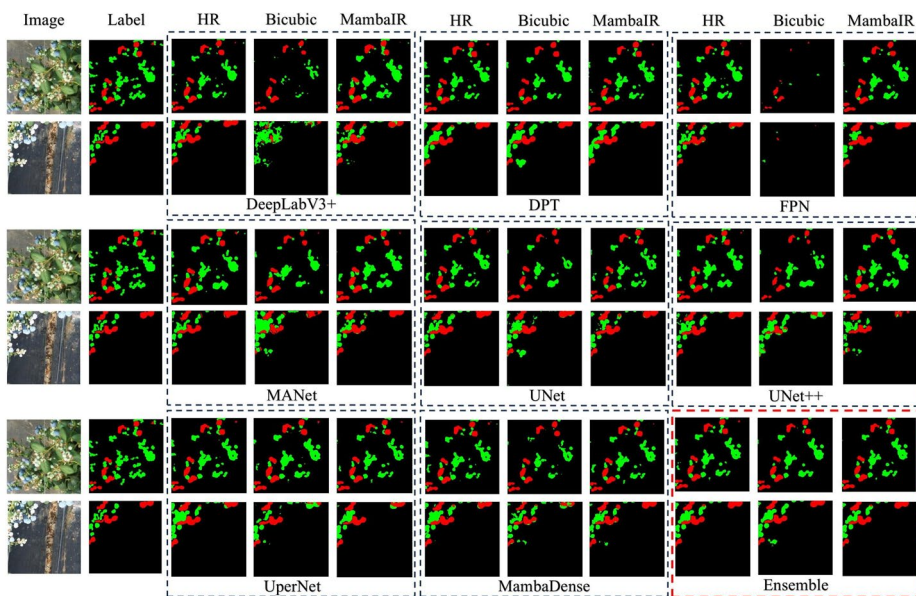


Fig. 6 Comparative results of different semantic segmentation models applied to HR, Bicubic, and MambaIR-enhanced datasets. (Each block shows the segmentation results of a specific segmentation model (e.g., DeepLabV3+, UNet++, MANet) under three different input conditions. The red masks correspond to mature blueberries, and the green masks indicate immature ones. This figure demonstrates how model performance varies with input image quality and highlights the advantage of using SRR-enhanced imagery, particularly with the MambaIR method. The results indicate that MambaIR consistently improves segmentation accuracy across all models. Ensemble results are shown at the bottom-right as a baseline for optimal performance.)

method, underscoring the efficacy of SRR models in enhancing segmentation accuracy for blueberry maturity assessment.

Ensemble learning approach

This study enhanced the accuracy and robustness of blueberry maturity segmentation by employing an ensemble learning approach. This method combines 27 segmentation models based on CNN, Transformer, and Mamba architectures. By leveraging the strengths of multiple models, ensemble learning typically achieves superior prediction performance in complex tasks, particularly in high-complexity scenarios like image segmentation (Dong et al., 2020).

To construct the final integration model, we first evaluated the segmentation performance of all 27 models and selected the top-performing ones with mIoU values exceeding 81. Specifically, three representative models were chosen: ResNeXt101_32 × 8 d in UNet++ (CNN-based), DINOv2(vit_1) in DPT (Transformer-based), and ChangeMamba (VMamba-tiny) (Mamba-based). These models were selected for their complementary strengths: ResNeXt101_32 × 8 d excels in detailed segmentation of complex backgrounds, DINOv2 captures fine texture and edge information, and ChangeMamba demonstrates superior light-weight performance and temporal modeling capabilities. By integrating these models, this study effectively merged their strengths to enhance overall segmentation performance. The ensemble was implemented by averaging the softmax outputs (pixel-wise probability maps) of the three models. This late fusion strategy was chosen to maintain the diversity of feature representations while reducing variance.

The ensemble model, which combined these top models, achieved superior performance, with an mIoU of 83.05 in the first ensemble and 83.13 in a more diverse version. These results significantly surpassed the individual models' performance, showcasing the effectiveness of ensemble learning in improving segmentation accuracy, particularly for blueberry fruits with irregular textures and edges.

Discussion

Finally, this study tested the robustness of our method on degraded input data, focusing on three aspects: (1) the effect of different scaling ratios on model performance; (2) the impact of varying degrees of Gaussian blur on the input data; and (3) the effect of different intensities of Gaussian noise on the input data.

The influence of magnification factor

The effectiveness of SRR is profoundly influenced by the magnification factor, which governs the enlargement ratio from LR to HR. Variations in magnification significantly affect the model's ability to recover intricate image details, a critical aspect in tasks such as blueberry maturity segmentation. Lin et al. (2024) demonstrated that the magnification factor not only substantially affects visual quality but also impacts performance metrics like Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Moreover, multiscale approaches have enhanced SRR models' adaptability to different magnification

factors by optimizing network architectures, thereby improving both reconstruction quality and model robustness, as highlighted by Cheng et al. (2022). In light of these findings, this study examines the effects of varying magnification factors on the performance of SRR, specifically in the context of blueberry maturity segmentation.

The experimental procedure was structured as follows: A high-resolution blueberry image dataset was first generated with a resolution of 512×512 . To simulate realistic image degradation, this HR dataset was downsampled to a LR dataset (x1-LR) with a resolution of 128×128 , using the degradation model detailed in Eq. 13. Subsequently, SRR was applied to the x1-LR dataset at magnification factors of 2, 3, 4, and 5, resulting in datasets labeled as x2-SR, x3-SR, x4-SR, and x5-SR, respectively.

During the training phase, the training and validation subsets of the HR blueberry dataset were used to train an ensemble model, from which the optimal segmentation model was selected. This model was then utilized to evaluate segmentation performance on the SRR datasets reconstructed at different magnification levels. The mean Intersection over Union (mIoU) was employed to quantify segmentation performance across scales. Figure 7 presents the segmentation results for each SRR test set, labeled as x(m)-SR to represent the results at magnification level m, alongside x1-LR for the segmentation results on the original LR dataset.

Figure 7 also illustrates the robustness of different training models under the same and varying magnification factors. The results indicate that as the magnification factor increases, the mIoU of all reconstructed images progressively improves, stabilizing after 3x magnification. These findings suggest that higher magnification ratios can enhance the segmentation accuracy of SRR models, reducing the disparity between reconstructed images and ground truth labels. Notably, the proposed ensemble method consistently outperforms other models across magnification levels, with MambaDense and DPT following closely. At 2x magnification, DPT surpasses MambaDense by 0.87% and maintains slightly higher mIoU

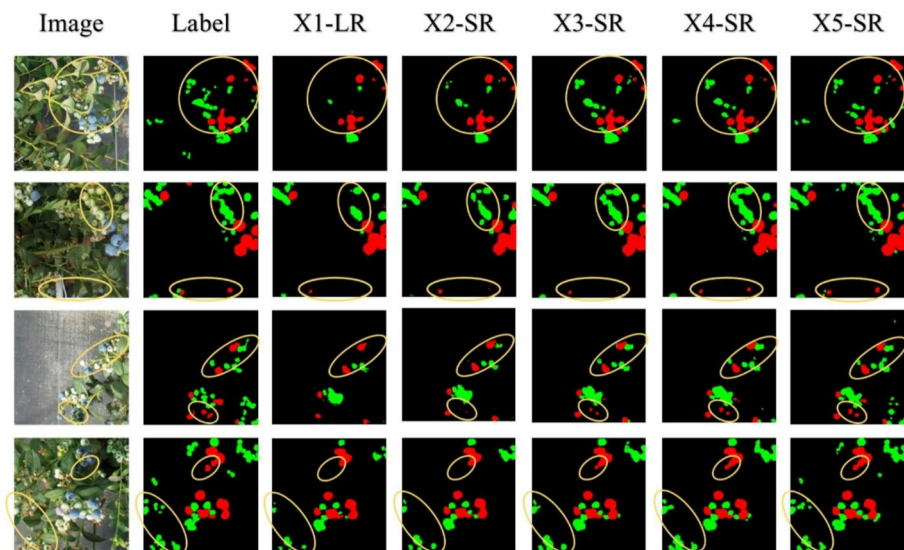


Fig. 7 Segmentation results of images from SR test sets obtained with different magnification factors

values in subsequent magnifications. DeepLab V3 + performs comparatively worse, lagging behind other algorithms by approximately 10%. However, computational cost rises significantly with increasing magnification factors. For blueberry image segmentation, a 4x magnification offers the optimal balance between segmentation accuracy and computational efficiency, although the exact computational time data require further validation. Figure 8 presents a comparative visualization of segmentation accuracy across different magnification levels, complementing the numerical results shown in Fig. 7. It provides an intuitive illustration of how increased magnification enhances segmentation accuracy, highlighting the consistent performance gains achieved by the proposed ensemble model compared to other architectures. This figure further supports the conclusion that 4× magnification offers the best trade-off between accuracy and computational efficiency for practical deployment in blueberry maturity monitoring.

Effect of the Gaussian blur

The size of the Gaussian blur kernel is a key determinant in the performance of SRR, as it influences the level of detail degradation in low-resolution images. In this study, Gaussian blur was applied during the image degradation process before SRR training and inference, simulating the defocus and motion-induced blur commonly observed in UAV imagery. Specifically, the original high-resolution images were convolved with Gaussian kernels of different sizes—11 (Blur_lv1), 15 (Blur_lv2), 21 (Blur_lv3), and 25 (Blur_lv4)—to generate corresponding blurred low-resolution inputs. These degraded images were then used as input to the SRR models to evaluate their impact on subsequent segmentation accuracy.

In terms of precision, the study observed a gradual increase as the Gaussian blur kernel size expanded from 11 to 25, with the precision rate rising from 90.07 to 90.82%. This pattern suggests that greater blurring suppresses noise, enabling the model to more effectively distinguish the boundaries of blueberry fruits, thereby enhancing segmentation accuracy in

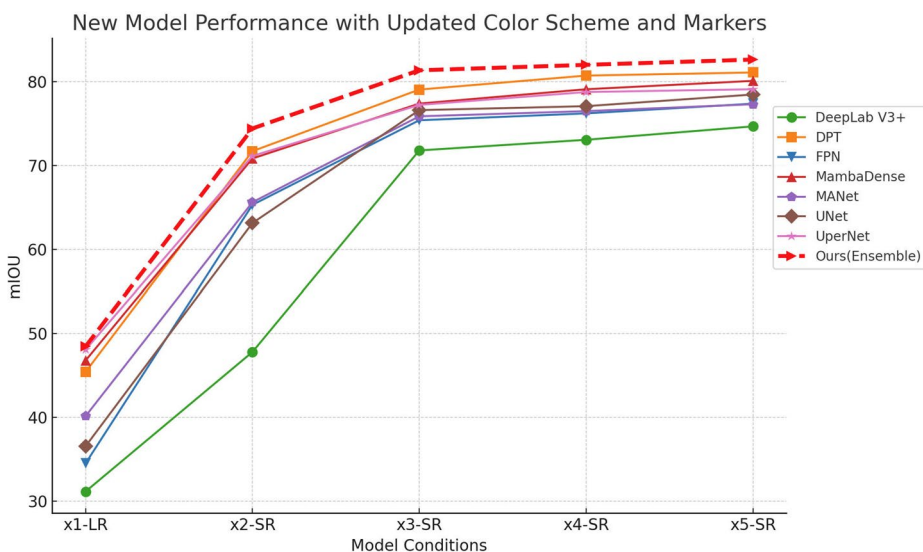


Fig. 8 Evaluating the accuracy of blueberry maturity segmentation across different magnifications

certain cases. This observation is consistent with the well-established advantages of Gaussian blur in image processing, which include noise reduction and improved local contrast by smoothing out unnecessary details(Aakerberg et al., 2022).

Despite the improvement in precision, Fig. 9 shows a decline in the mean Intersection over Union (mIoU) with increasing Gaussian blur kernel sizes. Among the models tested, the Ensemble model, MambaDense, and UperNet demonstrated relative resilience, with performance drops of 17.54%, 14.98%, and 14.94%, respectively, as blur increased. Nevertheless, the Ensemble model consistently outperformed the others. Although Gaussian blur can improve detection accuracy for certain blueberry fruits by reducing noise, it generally hampers overall segmentation performance as the kernel size increases. This is primarily due to the loss of critical edge information and fine textures, especially in blueberries with complex surfaces and irregular boundaries(Wang et al., 2019, 2022a, b). Consequently, the model's capacity to accurately segment such features diminishes, resulting in lower segmentation performance overall.

Effect of different intensities of Gaussian noise

In this experiment, varying levels of noise variance were introduced to evaluate their impact on the segmentation performance of hyperspectral SRR-reconstructed blueberry images. Gaussian noise was applied during the image degradation process prior to SRR reconstruction, simulating real-world UAV imaging conditions where sensor noise or environmental disturbances can degrade image quality. Specifically, zero-mean Gaussian white noise with different variances was added directly to the low-resolution input images used to train and evaluate the SRR models. The noise added to each pixel follows a normal distribution:

$$\eta(x, y) \sim \mathcal{N}(0, \sigma^2) \quad (14)$$

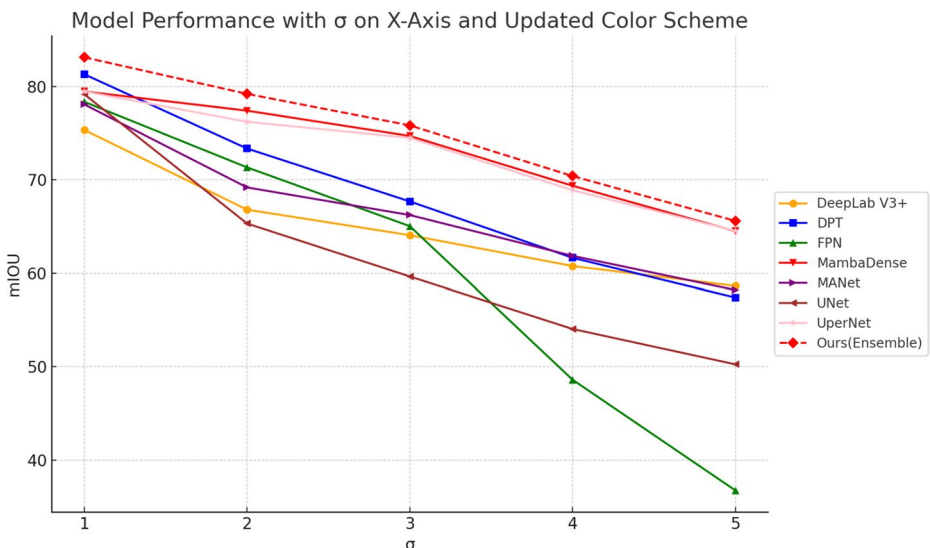


Fig. 9 Evaluation of blueberry maturity segmentation accuracy under different Gaussian blur levels

where $\eta(x, y)$ is the noise added at pixel (x, y) , and σ^2 is the variance controlling noise intensity.

Noise variance represents the intensity of noise added to each image, influencing both image quality and segmentation accuracy. As shown in Fig. 10, when $\sigma = 1$ (Noise_lv0), the noise variance is 0, meaning no noise was added. At $\sigma = 2$ (Noise_lv1), corresponding to a variance of 0.01, only minimal noise was present, causing negligible interference in the image. However, as σ increased, noise levels escalated: Noise_lv2 ($\sigma = 3$, variance = 0.05), Noise_lv3 ($\sigma = 4$, variance = 0.1), and Noise_lv4 ($\sigma = 5$, variance = 0.2). With higher noise levels, interference in the images became more pronounced, leading to noticeable degradation in segmentation quality.

The overall performance metrics, as depicted in Fig. 10 through mIoU values, show that when $\sigma = 0$, the proposed Ensemble model achieved the highest mIoU score of 83.13%. As σ increased to 2, a slight decline in performance was observed for the Ensemble, DPT, MambaDense, and UperNet models, while other algorithms exhibited a more significant drop, with mIoU reductions ranging from 24.2 to 41.5%. At $\sigma = 3$, all models experienced a sharp decline in performance. These findings indicate that higher noise levels not only degrade model accuracy but also significantly impair segmentation performance in spatial tasks. As noise variance increases, critical edges and details in the blueberry images become obscured, making accurate segmentation more challenging. While minor noise leads to only slight reductions in accuracy, excessive noise invariably causes substantial performance deterioration, undermining the model's ability to produce reliable segmentation outcomes.

Distribution map of blueberry maturity with segmentation results

Figure 11 provides a comprehensive visual representation of blueberry growth and ripeness distribution, which can be instrumental in advancing precision agriculture practices. The UAV imagery used in this figure was captured under controlled greenhouse conditions,

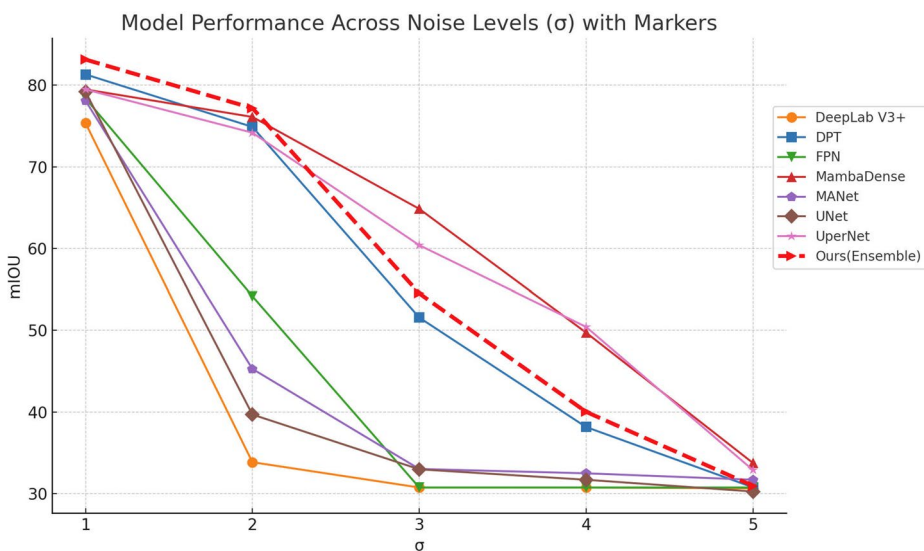


Fig. 10 Evaluation of segmentation performance under varying levels of Gaussian noise

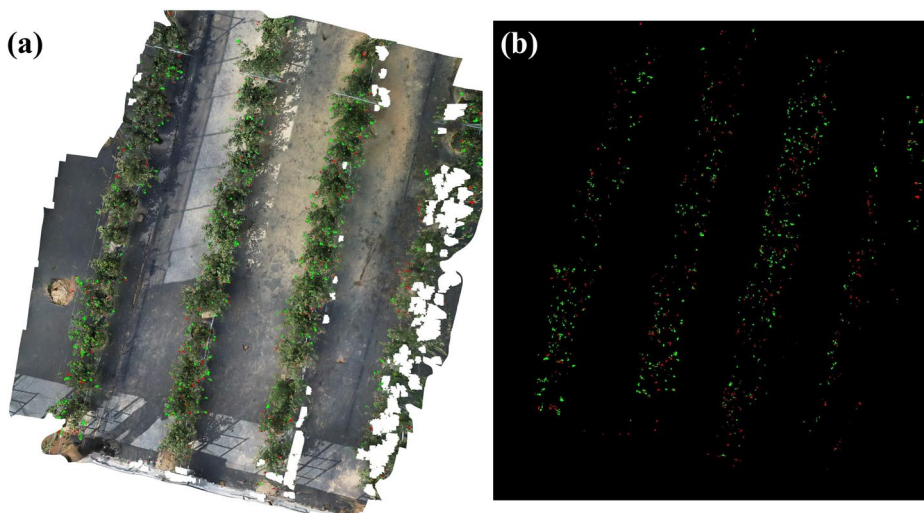


Fig. 11 Spatial Distribution and Segmentation Results of Blueberries. **(a)** Distribution map with location information captured under greenhouse conditions using UAV imagery. **(b)** Pixel-level mask visualization of blueberry maturity segmentation results, generated using the proposed ensemble segmentation model. Red indicates mature fruits and green indicates immature fruits

which ensured stable lighting and minimized environmental interference. Figure 11(a) overlays segmentation results in location data, offering an intuitive understanding of the overall field conditions. By spatially mapping blueberry maturity across the plantation, this visualization supports critical agricultural decisions, such as targeted harvesting, disease prevention, and controlled application of fertilizers and pesticides. The integration of location-specific maturity data not only facilitates more efficient resource allocation but also enhances the accuracy of yield estimates, ultimately contributing to increased crop productivity.

The mask representation (Fig. 11b) delivers additional insights by categorizing pixels based on maturity levels. This pixel-level segmentation enables precise quantification of the proportion of ripe versus unripe blueberries, with mature fruits comprising 31% of all fruits in Fig. 11b. Such detailed information allows growers to pinpoint optimal harvesting periods, reducing the risk of underripe or overripe yields. By timing harvests to the specific maturity needs of different areas, farmers can better align with market demand, minimize waste, and increase the overall quality of produce.

In regions where agriculture remains predominantly labor-intensive, including developing countries like China, the strategic timing of harvests enabled by these segmentation maps holds significant potential. By optimizing harvest schedules, growers can avoid labor shortages during peak seasons, lower associated costs, and manage workflows more efficiently. These visualizations thus support both the financial and operational aspects of precision agriculture, helping farmers make timely, data-informed decisions that increase profitability.

The mapping of maturity across large plantations further underscores the potential of UAV-based imagery for scalable, cost-effective agricultural management. By combining UAV-captured imagery with SRR and semantic segmentation, growers gain access

to high-resolution, actionable insights without the need for high-cost camera equipment. This approach broadens the accessibility of advanced monitoring techniques, especially for small- to medium-scale farmers, highlighting the feasibility and impact of computer vision technologies in supporting sustainable agriculture.

Ultimately, these mapping visualizations contribute to a more robust and scalable system for real-time monitoring, empowering farmers to adapt to dynamic field conditions and proactively manage resources to achieve optimal crop health and yield. The practical utility of these visualized outputs, as shown in Fig. 11, lies in their ability to transform raw image data into precise, easy-to-interpret information that drives informed decision-making, solidifying the role of UAV and SRR technologies as pivotal tools in modern agricultural practices.

Conclusion

This study presents a novel approach that combines deep learning-based SRR with semantic segmentation to assess blueberry maturity, addressing challenges posed by low-resolution UAV imagery and the need for efficient ripeness detection in cultivation settings. Initially, SRR techniques enhance the resolution of low-quality images, after which 27 semantic segmentation models, spanning CNN, Transformer-based architectures, and the Mamba network, are evaluated. By integrating the top-performing models from each category into an ensemble, the method achieves significant improvements in segmentation accuracy by leveraging the distinct advantages of each model type. The ensemble demonstrates superior stability and precision in determining blueberry ripeness, validating its practical applicability.

The Mamba-based network, in particular, excels in both SRR and segmentation tasks. Its architecture effectively preserves image details while minimizing blurring, resulting in high-resolution images with superior reconstruction fidelity compared to other networks. Additionally, its lightweight design and cross-scale feature extraction capabilities enable reliable detection of blueberries in complex environments, underscoring its potential to enhance both segmentation performance and model robustness.

This study also assessed the robustness of the proposed method under various data degradation scenarios, including different scaling ratios, Gaussian blur, and noise intensities. The results indicate that segmentation accuracy improves with increasing magnification up to 4x, beyond which further magnification yields diminishing returns while substantially increasing computational costs. Furthermore, while larger Gaussian blur kernels may improve recognition in certain cases, overall segmentation performance deteriorates with excessive blurring. Similarly, moderate noise levels may cause slight improvements, but higher noise intensities severely degrade segmentation accuracy. These findings highlight the importance of controlling data degradation in practical applications to maintain model effectiveness.

Future research will focus on refining the Mamba architecture by integrating SRR and semantic segmentation across diverse environmental conditions and data modalities. A particular emphasis will be placed on monitoring the detailed maturity stages of blueberry crops, allowing for more granular assessments of growth progression. Additionally, future efforts will include pixel-based analysis of sample plots for yield prediction, expanding the application from ripeness evaluation to direct estimation of crop yields. These advance-

ments are expected to significantly contribute to smart agricultural management, enhancing precision farming techniques for optimized resource use and increased productivity.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11119-025-10252-2>.

Acknowledgements The authors are grateful to the Yunnan Yijiu Agricultural Development Co., Ltd. for their support in Yuxi city, China. This research was partially supported by the Japan Science and Technology Agency SPRING Program (JST SPRING), Grant Number JPMJSP2108. This work was partially funded by JST, PRESTO Grant Number JPMJPR24G9, Japan.

Funding Open Access funding provided by The University of Tokyo.

Data availability Data will be made available on request.

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aakerberg, A., Johansen, A. S., Nasrollahi, K., & Moeslund, T. B. (2022). Semantic Segmentation Guided Real-World Super-Resolution. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 449–458. <https://doi.org/10.1109/WACVW54805.2022.00051>
- Aherwadi, N., Mittal, U., Singla, J., Jhanjhi, N. Z., Yassine, A., & Hossain, M. S. (2022). Prediction of fruit maturity, quality, and its life using deep learning algorithms. *Electronics*, *11*(24), 4100. <https://doi.org/10.3390/electronics11244100>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaria, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), 53. <https://doi.org/10.1186/s40537-021-00444-8>
- Anatya, S., Mawardi, C. V., & Hendryli, J. (2020). Fruit Maturity Classification Using Convolutional Neural Networks Method. *IOP Conference Series: Materials Science and Engineering*, *1007*(1), 012149. <https://doi.org/10.1088/1757-899X/1007/1/012149>
- Banerjee, S., Nayik, G. A., Kour, J., & Nazir, N. (2020). Blueberries. In G. A. Nayik & A. Gull (Eds.), *Antioxidants in Fruits: Properties and Health Benefits* (pp. 593–614). Springer Singapore. https://doi.org/10.1007/978-981-15-7285-2_31
- Bargoti, S., & Underwood, J. P. (2017). Image segmentation for fruit detection and yield Estimation in Apple orchards. *Journal of Field Robotics*, *34*(6), 1039–1060. <https://doi.org/10.1002/rob.21699>
- Behera, S. K., Rath, A. K., & Sethy, P. K. (2021). Fruits yield Estimation using faster R-CNN with MIOU. *Multimedia Tools and Applications*, *80*(12), 19043–19056. <https://doi.org/10.1007/s11042-021-10704-7>
- Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision—ECCV 2018* (Vol. 11211, pp. 833–851). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_49

- Chen, X., Wang, X., Zhang, W., Kong, X., Qiao, Y., Zhou, J., & Dong, C. (2023). Hat: Hybrid attention transformer for image restoration. *arXiv Preprint arXiv:2309.05239*.
- Chen, H., Song, J., Han, C., Xia, J., & Yokoya, N. (2024). ChangeMamba: Remote sensing change detection with Spatio-Temporal state space model. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1–20. <https://doi.org/10.1109/TGRS.2024.3417253>
- Cheng, Y., Chen, S., Liao, Z., & Zhou, N. (2022). Multi-Scale factor image Super-Resolution algorithm with information distillation network. *Applied Sciences*, 12(9), 4131. <https://doi.org/10.3390/app12094131>
- Csurka, G., & Perronnin, F. (2011). An efficient approach to semantic segmentation. *International Journal of Computer Vision*, 95(2), 198–212. <https://doi.org/10.1007/s11263-010-0344-8>
- Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14*, 391–407.
- Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14(2), 241–258. <https://doi.org/10.1007/s11704-019-8208-z>
- Ellenberg, A., Kontsos, A., Moon, F., & Bartoli, I. (2016). Bridge related damage quantification using unmanned aerial vehicle imagery: Damage quantification using unmanned aerial vehicle imagery. *Structural Control and Health Monitoring*, 23(9), 1168–1179. <https://doi.org/10.1002/stc.1831>
- Farooque, A. A., Chang, Y. K., Zaman, Q. U., Groulx, D., Schumann, A. W., & Esau, T. J. (2013). Performance evaluation of multiple ground based sensors mounted on a commercial wild blueberry harvester to sense plant height, fruit yield and topographic features in real-time. *Computers and Electronics in Agriculture*, 91, 135–144. <https://doi.org/10.1016/j.compag.2012.12.006>
- Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., & Xia, S. T. (2024). MambaIR: A Simple Baseline for Image Restoration with State-Space Model (arXiv:2402.15648). arXiv. <http://arxiv.org/abs/2402.15648>
- Hatvani, J., Horvath, A., Michetti, J., Basarab, A., Kouame, D., & Gyongy, M. (2019). Deep Learning-Based Super-Resolution applied to dental computed tomography. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 3(2), 120–128. <https://doi.org/10.1109/TRPMS.2018.2827239>
- Jiang, J., Wang, C., Liu, X., & Ma, J. (2023). Deep Learning-based face Super-resolution: A survey. *ACM Computing Surveys*, 55(1), 1–36. <https://doi.org/10.1145/3485132>
- Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision—ECCV 2016* (Vol. 9906, pp. 694–711). Springer International Publishing. https://doi.org/10.1007/978-3-319-46475-6_43
- Kader, A. A. (1997). Fruit maturity, ripening, and quality relationships. *International Symposium Effect of Pre- & Postharvest Factors in Fruit Storage*, 485, 203–208.
- Kaur, H., Sawhney, B. K., & Jawandha, S. K. (2018). Evaluation of Plum fruit maturity by image processing techniques. *Journal of Food Science and Technology*, 55(8), 3008–3015. <https://doi.org/10.1007/s13197-018-3220-0>
- Kim, E., Freivalds, A., Takeda, F., & Li, C. (2018). Ergonomic evaluation of current advancements in blueberry harvesting. *Agronomy*, 8(11), 266. <https://doi.org/10.3390/agronomy8110266>
- Koirala, A., Walsh, K. B., Wang, Z., & McCarthy, C. (2019). Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘mangoyolo’. *Precision Agriculture*, 20(6), 1107–1135. <https://doi.org/10.1007/s11119-019-09642-0>
- Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., & Shi, W. (2017). Photo-Realistic single image Super-Resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 105(114). <https://doi.org/10.1109/CVPR.2017.19>
- Li, K., Yang, S., Dong, R., Wang, X., & Huang, J. (2020). Survey of single image super-resolution reconstruction. *IET Image Processing*, 14(11), 2273–2290. <https://doi.org/10.1049/iet-ipr.2019.1438>
- Li, Y., Pei, J., Chen, L., & Sun, H. (2021a). *China blueberry industry report 2020*.
- Li, Y., Sixou, B., & Peyrin, F. (2021b). A review of the deep learning methods for medical images super resolution problems. *IRBM*, 42(2), 120–133. <https://doi.org/10.1016/j.irbm.2020.08.004>
- Li, R., Zheng, S., Zhang, C., Duan, C., Su, J., Wang, L., & Atkinson, P. M. (2022). Multiattention network for semantic segmentation of Fine-Resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–13. <https://doi.org/10.1109/TGRS.2021.3093977>
- Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1132–1140. <https://doi.org/10.1109/CVPRW.2017.151>
- Lin, T. Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 936–944. <https://doi.org/10.1109/CVPR.2017.106>
- Lin, X., Liu, X., Yang, H., He, X., & Chen, H. (2024). Perception- and Fidelity-aware Reduced-Reference Super-Resolution Image Quality Assessment (arXiv:2405.09472). arXiv. <http://arxiv.org/abs/2405.09472>

- Liu, Y., Yeoh, J. K. W., & Chua, D. K. H. (2020). Deep Learning–Based enhancement of motion blurred UAV concrete crack images. *Journal of Computing in Civil Engineering*, 34(5), 04020028. [https://doi.org/10.1061/\(ASCE\)JCP.1943-5487.0000907](https://doi.org/10.1061/(ASCE)JCP.1943-5487.0000907)
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012, 10022.
- Liu, Y., Zheng, H., Zhang, Y., Zhang, Q., Chen, H., Xu, X., & Wang, G. (2023). Is this blueberry ripe? A blueberry ripeness detection algorithm for use on picking robots. *Frontiers in Plant Science*, 14, 1198650. <https://doi.org/10.3389/fpls.2023.1198650>
- MacEachern, C. B., Esau, T. J., Schumann, A. W., Hennessy, P. J., & Zaman, Q. U. (2023). Detection of fruit maturity stage and yield Estimation in wild blueberry using deep learning convolutional neural networks. *Smart Agricultural Technology*, 3, 100099. <https://doi.org/10.1016/j.atech.2022.100099>
- Matese, A., Toscano, P., Di Gennaro, S., Genesio, L., Vaccari, F., Primicerio, J., Belli, C., Zaldei, A., Bianconi, R., & Gioli, B. (2015). Intercomparison of UAV, aircraft and satellite remote sensing platforms for precision viticulture. *Remote Sensing*, 7(3), 2971–2990. <https://doi.org/10.3390/rs70302971>
- Mendoza, F., & Aguilera, J. M. (2004). Application of image analysis for classification of ripening bananas. *Journal of Food Science*, 69(9). <https://doi.org/10.1111/j.1365-2621.2004.tb09932.x>
- Moltó, E., Plá, F., & Juste, F. (1992). Vision systems for the location of citrus fruit in a tree canopy. *Journal of Agricultural Engineering Research*, 52, 101–110. [https://doi.org/10.1016/0021-8634\(92\)80053-U](https://doi.org/10.1016/0021-8634(92)80053-U)
- Muksimova, S., Umirzakova, S., Mardieva, S., Abdullaev, M., & Cho, Y. I. (2024). Revolutionizing wildfire detection through UAV-Driven fire monitoring with a Transformer-Based approach. *Fire*, 7(12). <https://doi.org/10.3390/fire7120443>
- Oscó, L. P., Marcato Junior, J., Marques Ramos, A. P., De Castro Jorge, L. A., Fatholahi, S. N., De Andrade Silva, J., Matsubara, E. T., Pistori, H., Gonçalves, W. N., & Li, J. (2021). A review on deep learning in UAV remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 102, 102456. <https://doi.org/10.1016/j.jag.2021.102456>
- Park, S. C., Park, M. K., & Moon Gi Kang. (2003). Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3), 21–36. <https://doi.org/10.1109/MSP.2003.1203207>
- Prasad, K., Jacob, S., & Siddiqui, M. W. (2018). Chapter 2—Fruit Maturity, Harvesting, and Quality Standards. In M. W. Siddiqui (Ed.), *Preharvest Modulation of Postharvest Fruit and Vegetable Quality* (pp. 41–69). Academic Press. <https://doi.org/10.1016/B978-0-12-809807-3.00002-0>
- Primicerio, J., Di Gennaro, S. F., Fiorillo, E., Genesio, L., Lugato, E., Matese, A., & Vaccari, F. P. (2012). A flexible unmanned aerial vehicle for precision agriculture. *Precision Agriculture*, 13(4), 517–523. <https://doi.org/10.1007/s11119-012-9257-6>
- Radoglou-Grammatikis, P., Sarigiannidis, P., Lagkas, T., & Moscholios, I. (2020). A compilation of UAV applications for precision agriculture. *Computer Networks*, 172, 107148. <https://doi.org/10.1016/j.comnet.2020.107148>
- Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision Transformers for Dense Prediction. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 12159–12168. <https://doi.org/10.1109/ICCV48922.2021.01196>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and Computer-Assisted Intervention– MICCAI 2015* (pp. 234–241). Springer International Publishing.
- Rowland, L. J., Alkharouf, N., Darwish, O., Ogden, E. L., Polashock, J. J., Bassil, N. V., & Main, D. (2012). Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. *BMC Plant Biology*, 12(1), 46. <https://doi.org/10.1186/1471-2229-12-46>
- Shah, S. A., Zeb, S., Qureshi, A., Arslan, W. S., Ullah Malik, M., Alasmay, A., W., & Alanazi, E. (2020). Towards fruit maturity Estimation using NIR spectroscopy. *Infrared Physics & Technology*, 111, 103479. <https://doi.org/10.1016/j.infrared.2020.103479>
- Sieberth, T., Wackrow, R., & Chandler, J. H. (2014). Motion blur disturbs– the influence of motion-blurred images in photogrammetry. *The Photogrammetric Record*, 29(148), 434–453. <https://doi.org/10.1111/phor.12082>
- Silva, S., Costa, E. M., Veiga, M., Morais, R. M., Calhau, C., & Pintado, M. (2020). Health promoting properties of Blueberries: A review. *Critical Reviews in Food Science and Nutrition*, 60(2), 181–200. <https://doi.org/10.1080/10408398.2018.1518895>
- Singarimbun, R. N., Nababan, E. B., & Sitompul, O. S. (2019). Adaptive moment estimation to minimize square error in backpropagation algorithm. *2019 International Conference of Computer Science and Information Technology (ICoSNiKOM)*, 1–7.

- Takeda, F., Yang, W., Li, C., Freivalds, A., Sung, K., Xu, R., Hu, B., Williamson, J., & Sargent, S. (2017). Applying new technologies to transform blueberry harvesting. *Agronomy*, 7(2), 33. <https://doi.org/10.3390/agronomy7020033>
- Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Computers and Electronics in Agriculture*, 157, 417–426. <https://doi.org/10.1016/j.compag.2019.01.012>
- Tsuros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-Based applications for precision agriculture. *Information*, 10(11), 349. <https://doi.org/10.3390/info10110349>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
- Wang, X., Liang, X., Zheng, J., & Zhou, H. (2019). Fast detection and segmentation of partial image blur based on discrete Walsh–Hadamard transform. *Signal Processing: Image Communication*, 70, 47–56. <https://doi.org/10.1016/j.image.2018.09.007>
- Wang, X., Xie, L., Dong, C., & Shan, Y. (2021a). Real-ESRGAN: Training Real-World blind Super-Resolution with pure synthetic data. *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 1905–1914. <https://doi.org/10.1109/ICCVW54120.2021.00217>
- Wang, Z., Chen, J., & Hoi, S. C. H. (2021b). Deep learning for image Super-Resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10), 3365–3387. <https://doi.org/10.1109/TPAMI.2020.2982166>
- Wang, X., Yi, J., Guo, J., Song, Y., Lyu, J., Xu, J., Yan, W., Zhao, J., Cai, Q., & Min, H. (2022a). A review of image Super-Resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14(21), 5423. <https://doi.org/10.3390/rs14215423>
- Wang, Z., Li, K., Wang, X., & Lee, A. (2022b). An image edge detection algorithm based on Multi-Feature fusion. *Computers Materials & Continua*, 73(3), 4995–5009. <https://doi.org/10.32604/cmc.2022.029650>
- Xiang, C., Wang, W., Deng, L., Shi, P., & Kong, X. (2022). Crack detection algorithm for concrete structures based on super-resolution reconstruction and segmentation network. *Automation in Construction*, 140, 104346. <https://doi.org/10.1016/j.autcon.2022.104346>
- Xiao, T., Liu, Y., Zhou, B., Jiang, Y., & Sun, J. (2018). Unified Perceptual Parsing for Scene Understanding. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision—ECCV 2018* (Vol. 11209, pp. 432–448). Springer International Publishing. https://doi.org/10.1007/978-3-030-01228-1_26
- Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision—ECCV 2014* (Vol. 8689, pp. 818–833). Springer International Publishing. https://doi.org/10.1007/978-3-319-10590-1_53
- Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., & Fu, Y. (2018). Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision—ECCV 2018* (Vol. 11211, pp. 294–310). Springer International Publishing. https://doi.org/10.1007/978-3-030-01234-2_18
- Zhao, H., Xu, D., Lawal, O., & Zhang, S. (2021). Muskmelon Maturity Stage Classification Model Based on CNN. *Journal of Robotics*, 2021, 1–12. <https://doi.org/10.1155/2021/8828340>
- Zhao, F., Liu, Y., Wang, J., Chen, Y., Xi, D., Shao, X., Tabeta, S., & Mizuno, K. (2024a). Riverbed litter monitoring using consumer-grade aerial-aquatic Speedy scanner (AASS) and deep learning based super-resolution reconstruction and detection network. *Marine Pollution Bulletin*, 209, 117030. <https://doi.org/10.1016/j.marpolbul.2024.117030>
- Zhao, F., Xi, D., Chen, Y., Ma, B., Liu, Y., Wang, J., & Mizuno, K. (2024b). Basic Study of Deep Learning Based Efficient Hermit Crabs Detection from Drone-Captured Images. *OCEANS 2024 - Singapore*, Singapore, Singapore, 2024, pp. 1–5. <https://doi.org/10.1109/OCEANS51537.2024.10706264>
- Zhao, F., Huang, J., Liu, Y., Wang, J., Chen, Y., Shao, X., & He, Y. (2024c). A Deep Learning Approach Combining Super-Resolution and Segmentation to Identify Weed and Tobacco in UAV Imagery. In *2024 9th International Conference on Electronic Technology and Information Science (ICETIS)* (pp. 594–597). IEEE.
- Zhao, F., Song, J., Wang, J., Chen, Y., Xi, D., Shao, X., & Zhang, C. (2024d). Mamba-Based Super-Resolution and Segmentation Network for UAV-Captured Blueberry Farmland Imagery. In *2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS)* (pp. 644–649). IEEE.
- Zhao, F., Ren, Z., Wang, J., Wu, Q., Xi, D., Shao, X., & Mizuno, K. (2025). Smart UAV-assisted rose growth monitoring with improved YOLOv10 and Mamba restoration techniques. *Smart Agricultural Technology*, 10, 100730. <https://doi.org/10.1016/j.atech.2024.100730>
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, 3–11.

Bauer, N., & Visuals, N. (2008). *World Blueberry Acreage & Production*.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Fan Zhao¹  · Yinyin He² · Jian Song¹ · Jiaqi Wang¹ · Dianhan Xi¹ · Xinlei Shao¹ · Qingyang Wu³ · Yongying Liu¹ · Yijia Chen¹ · Guochen Zhang¹ · Chenyu Zhang⁴ · Yulun Chen⁵ · Jundong Chen¹ · Katsunori Mizuno¹

✉ Fan Zhao
zhaofan25ut@163.com

Jian Song
songjianjp@gmail.com

Jiaqi Wang
leodwang97@gmail.com

¹ Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277- 8563, Chiba, Japan

² Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

³ Department of Environmental Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA 90095, US

⁴ Department of Geographic Information Science, Fujian Normal University, Fuzhou 350108, China

⁵ Department of Environmental Science, Southwest Forest University, Kunming 650224, China