

# A Comprehensive Review of Diffusion Models in Smart Agriculture: Progress, Applications, and Challenges

Xing Hu<sup>a,\*</sup>, Haodong Chen<sup>a</sup>, Qianqian Duan<sup>b</sup>, Danfeng Hong<sup>c</sup>, Ruijiao Li<sup>d</sup>, Huiliang Shang<sup>e</sup>, Linghua Jiang<sup>f</sup>, Haima Yang<sup>a</sup> and Dawei Zhang<sup>a</sup>

<sup>a</sup>School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, No. 516, Jungong Road, Shanghai, 200093, China

<sup>b</sup>School of Electronics and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China

<sup>f</sup>Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, 100049, China

<sup>c</sup>College of Intelligent Robotics and Advanced Manufacturing, Fudan University, Shanghai, 200433, China

<sup>d</sup>Future Information Innovation Institute, Fudan University, Shanghai, 200433, China

<sup>e</sup>Hangzhou Institute of Technology, Xidian University, Hangzhou, 311231, China

## ARTICLE INFO

### Keywords:

diffusion model, smart agriculture, precision agriculture, pest and disease detection, data augmentation, image generation, remote sensing images, crop monitoring, deep learning.

## ABSTRACT

With the global population growing and arable land resources becoming increasingly scarce, smart agriculture and precision agriculture have emerged as key directions for the future of agricultural development. Artificial intelligence (AI) technologies, particularly deep learning models, have found widespread applications in areas such as crop monitoring and pest detection. As an emerging generative model, diffusion models have shown significant promise in tasks like agricultural image processing, data augmentation, and remote sensing. Compared to traditional generative adversarial networks (GANs), diffusion models offer superior training stability and generation quality, effectively addressing challenges such as limited agricultural data and imbalanced image samples. This paper reviews the latest advancements in the application of diffusion models in agriculture, focusing on their potential in crop pest and disease detection, remote sensing image enhancement, crop growth prediction, and agricultural resource management. Experimental results demonstrate that diffusion models significantly improve model accuracy and robustness in data augmentation, image generation, and denoising, especially in complex environments. Despite challenges related to computational efficiency and generalization capabilities, diffusion models are expected to play an increasingly important role in smart and precision agriculture as technology advances, providing substantial support for the sustainable development of global agriculture.

## 1. Introduction

With the continuous growth of the global population and the increasing scarcity of arable land resources, traditional agriculture is facing unprecedented challenges. To improve agricultural production efficiency, ensure food security, and achieve sustainable development, countries worldwide are actively advancing the modernization of agriculture. In this context, smart agriculture and precision agriculture, as key components of agricultural modernization, are gradually becoming the mainstream directions for future agricultural development.

Smart agriculture leverages modern information technologies such as artificial intelligence (AI), the Internet of Things (IoT), big data, and remote sensing to enable dynamic sensing and data-driven decision-making regarding key factors such as crop growth environments, agricultural machinery operational status, soil moisture, and climate change. By enabling precise fertilization, targeted irrigation, and intelligent pest and disease identification, smart agriculture

This note has no numbers. In this work we demonstrate  $a_b$  the formation Y\_1 of a new type of polariton on the interface between a cuprous oxide slab and a polystyrene micro-sphere placed on the slab.

\*Corresponding author

\*\*Principal corresponding author

✉ huxing@usst.edu.cn (X. Hu); 18856715535@163.com (H. Chen); dqq1019@163.com (Q. Duan); hongdf@aircas.ac.cn (D. Hong); rjli20@fudan.edu.cn (R. Li); shanghai@fudan.edu.cn (H. Shang); jianglinhua@xidian.edu.cn (L. Jiang); snowyhm@sina.com (H. Yang); dwzhang@usst.edu.cn (D. Zhang)

ORCID(s): 0000-0003-1930-0372 (X. Hu)

<sup>1</sup>This is the first author footnote, but is common to third author as well.

<sup>2</sup>Another author footnote, this is a very long footnote and it should be a really long footnote. But this footnote is not yet sufficiently long enough to make two lines of footnote text.

not only significantly improves resource utilization but also enhances the intelligence, automation, and sustainability of agricultural production. As artificial intelligence technology continues to evolve, deep learning models, especially image generation models, have become a focal point in agricultural intelligence research. Among these, diffusion models, a new and advanced type of generative model in recent years, have demonstrated capabilities far exceeding traditional methods in tasks such as image generation [1, 2, 3, 4, 5, 6, 7], image restoration [1, 3, 4, 8, 9, 10], image segmentation [11, 12, 13, 14], and image enhancement [15]. They have achieved significant results in fields such as medicine [11, 13], industrial inspection [16], and artistic creation. Introducing these models into agricultural scenarios provides new solutions to real-world challenges, such as limited agricultural data, imbalanced samples, and insufficient data diversity. Therefore, systematically reviewing the current state of research and typical applications of diffusion models in agriculture can help promote the deep integration of AI technology with modern agriculture and accelerate the practical application of agricultural intelligence. In the process of promoting agricultural intelligence, deep learning, with its powerful feature extraction and pattern recognition capabilities, has become an essential technological support for agricultural informatization and automation. Compared to traditional machine learning methods, deep learning can automatically learn complex nonlinear relationships through an end-to-end approach, significantly enhancing the analysis of images, time series, and multi-source data. It is widely applied in various agricultural subfields, such as crop identification, pest and disease detection [17, 18], yield prediction [19], and environmental monitoring [20, 21, 22].

In image processing, convolutional neural networks (CNNs) [23] are widely used for tasks such as crop classification [24], weed identification, and leaf disease diagnosis [17, 18], leveraging high-resolution images to identify minute differences in agricultural fields. In remote sensing agriculture, deep neural networks can effectively process remote sensing data such as radar [25, 26], and multispectral images [27, 28, 29], enabling the automatic identification and analysis of crop types, planting areas, growth conditions, and moisture status. In time series prediction, recurrent neural networks (RNNs) and their variants, such as long short-term memory (LSTM) networks, are employed to model agricultural time series data like soil moisture, meteorological parameters [22], and crop growth indicators, providing valuable data for precise fertilization and irrigation decisions.

Additionally, generative models such as generative adversarial networks (GANs) [15, 30] are being increasingly used for agricultural data synthesis and augmentation to address sample scarcity and data imbalance in real-world agricultural scenarios. However, GANs suffer from issues like training instability and model collapse, which limit their further application in agriculture. Against this backdrop, the stability and high-quality generation capabilities of diffusion models are increasingly demonstrating their potential value in agricultural applications.

Diffusion models are a class of deep generative models that have made breakthrough progress in generative modeling in recent years. Their core concept originates from the mathematical modeling of diffusion processes in non-equilibrium thermodynamics. Unlike generative adversarial networks (GANs) [15, 30] and variational autoencoders (VAEs) [31, 32], diffusion models generate high-quality samples by gradually adding noise to the data during the training phase (the forward process) and then gradually denoising it during the generation phase using the learned reverse process. Since their introduction, diffusion models have undergone significant development and achieved breakthroughs across various fields. The concept of diffusion models originally emerged from diffusion processes in physics, aiming to simulate the gradual evolution of data from a true distribution to a noisy distribution and then recover the data from noise through the inverse process. In 2020, Ho et al. proposed Denoising Diffusion Probabilistic Models (DDPM), marking the first large-scale application of diffusion models in deep learning and addressing the issue of training instability in generative models, particularly in the field of image generation. The introduction of DDIM further optimized generation efficiency by reducing the number of generation steps through an implicit diffusion process, thus improving generation speed. Meanwhile, Score Matching Generative Models (SGM) introduced a new generation approach for diffusion models by learning the score function of data for generation, thus avoiding the instability issues present in traditional generative models. As the technology evolved, conditional diffusion models were proposed to handle more flexible conditional generation tasks, such as text-to-image generation, expanding the application scenarios of diffusion models. In 2022, latent diffusion models (LDMs) significantly improved generation efficiency and reduced computational requirements by performing diffusion in the latent space, further promoting the widespread application of diffusion models, especially in high-resolution image generation and artistic creation. The introduction of Stable Diffusion marked a new era for diffusion models, becoming an important tool in the field of text-to-image generation due to its efficient generation capabilities and open-source nature. Although diffusion models excel in terms of generation quality and training stability, they still face challenges such as computational resource demands, dataset quality, and generalization capabilities. In the future, with improvements in computational efficiency, the development of multi-modal data fusion technologies, and model optimization in low-resource environments,

diffusion models will play a significant role in a broader range of application scenarios, driving the further development of generative artificial intelligence.

The overall structure of this paper is as follows: Section 1 introduces the challenges and technological needs facing agriculture, as well as the potential, advantages, and challenges of diffusion models in agriculture. Section 2 covers the theory, evolution, and variants of diffusion models, as well as their relationship with other generative models. Section 3 discusses the specific applications of diffusion models in image processing in the agricultural field, including agricultural image classification, data augmentation, image generation, and pest and disease diagnosis. Section 4 compares diffusion models with other methods and conducts experiments and analyses. Section 5 explores the future prospects of diffusion models in the field of agricultural images. Section 6 summarizes the content of the paper.

The contributions of this paper are as follows: (1) We systematically reviewed seven typical application areas of diffusion models in smart agriculture (as shown in Figure 10), marking the first comprehensive review in this field; (2) We compared and analyzed the performance differences between diffusion models and GANs in agricultural imaging tasks, conducting the first small-scale experimental validation of the advantages of diffusion models in data augmentation for long-tail distributions; (3) We summarized the key challenges currently limiting the large-scale application of diffusion models (such as computational efficiency, data bottlenecks, and model generalization) and outlined future research directions.

## 2. Overview of the agricultural sector and background to the application of diffusion models

### 2.1. Major Challenges and Technological Needs Facing Agriculture

The primary challenges facing agriculture include crop health and pest management, precision agriculture and resource management, as well as remote sensing image processing and data analysis. Pests and diseases are key factors affecting agricultural production. Traditional manual inspection methods are inefficient and often lead to misdiagnosis, particularly in the early stages of disease. As agricultural scales expand, intelligent technologies have become essential for improving the efficiency and accuracy of pest and disease detection. Modern computer vision and artificial intelligence technologies can monitor crop health in real-time through remote sensing devices and sensors, accurately identify pests and diseases, and reduce reliance on manual labor.

Precision agriculture uses technology to achieve efficient resource utilization, maximizing crop yields while minimizing environmental pollution. By dynamically adjusting resource allocation, optimizing planting plans, reducing pesticide and fertilizer use, and lowering carbon footprints, precision agriculture promotes sustainable agricultural practices. However, the implementation of precision agriculture depends on IoT sensors, remote sensing technology, and big data analysis to monitor soil, climate, and crop health factors in real-time, thereby optimizing agricultural decision-making.

The application of remote sensing technology in agriculture is widespread, particularly in crop health monitoring and large-scale farmland surveillance. As shown in Figure 1, mobile platforms collect weed images. While remote sensing images can provide real-time data, factors such as weather and climate changes affect image quality, necessitating improvements in clarity, resolution, and noise reduction in existing technologies. Additionally, efficiently processing and analyzing large-scale remote sensing data to extract useful information remains an urgent issue.

Therefore, there is an urgent need for the adoption of automated and intelligent technologies, such as deep learning-based crop pest and disease detection systems, precision agriculture decision support systems, and remote sensing image enhancement and analysis technologies. These technologies will help enhance production efficiency, accuracy, and sustainability, driving the transformation of agriculture from traditional models to intelligent and precision-oriented development.

### 2.2. Potential applications of diffusion models in agriculture

Diffusion models show great promise in agricultural applications, particularly in crop monitoring, pest and disease identification, and remote sensing image processing. The agricultural sector often faces challenges such as data scarcity and sample imbalance, but diffusion models can effectively address these issues due to their high-quality generation capabilities.

First, diffusion models have significant potential for pest and disease identification in crops. Traditional pest and disease detection methods rely on manual inspections, which are inefficient and prone to misdiagnosis, especially during the early stages of disease onset. Diffusion models can generate high-quality images of pests and diseases,



**Figure 1:** Weed image acquisition by a mobile platform with the OpenWeedGUI in two different fields. [33]

particularly in the early stages when symptoms are not yet evident, providing clear diagnostic information to enhance the accuracy and efficiency of agricultural monitoring systems.

Second, in crop growth prediction, diffusion models can generate simulated crop growth data, as shown in Figure 2. By combining environmental factors such as soil moisture and climate change, they assist agricultural experts in predicting the health status and yield of crops at different growth stages, providing valuable support for scientific planting decisions.

Additionally, diffusion models show significant potential in enhancing and denoising remote sensing images. Remote sensing technology provides crucial data for agricultural monitoring, but image quality is often limited by weather and environmental conditions. Diffusion models can effectively improve the quality of remote sensing images, enabling agricultural experts to obtain clearer farmland data and enhance the accuracy of tasks such as crop health monitoring, land use assessment, and climate change impact analysis.

### 2.3. Advantages and challenges of diffusion models in agriculture

The advantages of diffusion models in agriculture primarily lie in their image generation quality, training stability, and flexible conditional generation capabilities. Compared to traditional generative adversarial networks (GANs), diffusion models exhibit higher stability during training, avoiding common issues such as pattern collapse that are associated with GANs, and can generate more diverse and high-quality agricultural images.

First, diffusion models generate high-quality images through gradual denoising, with details and clarity particularly prominent in complex agricultural environments. Whether in pest and disease monitoring or crop growth prediction, diffusion models can generate image data rich in detail, closely aligned with actual agricultural needs, significantly enhancing the accuracy and reliability of image generation.

Second, diffusion models offer a clear advantage in training stability compared to GANs. GANs often encounter instability during training and are prone to getting stuck in local optima. Diffusion models ensure training stability while improving generation quality through their gradual denoising process, making them suitable for handling large-scale data generation tasks in agriculture.

Additionally, diffusion models support flexible conditional generation, enabling the creation of images tailored to specific needs based on crop types, pest and disease categories, or environmental conditions. This capability allows diffusion models to provide customized solutions in agriculture, assisting farmers and agricultural experts with more precise crop monitoring, resource allocation, and precision agriculture decision-making.

Although diffusion models demonstrate significant advantages in agricultural applications, they still face challenges such as computational efficiency and data quality. Future research should focus on improving the computational

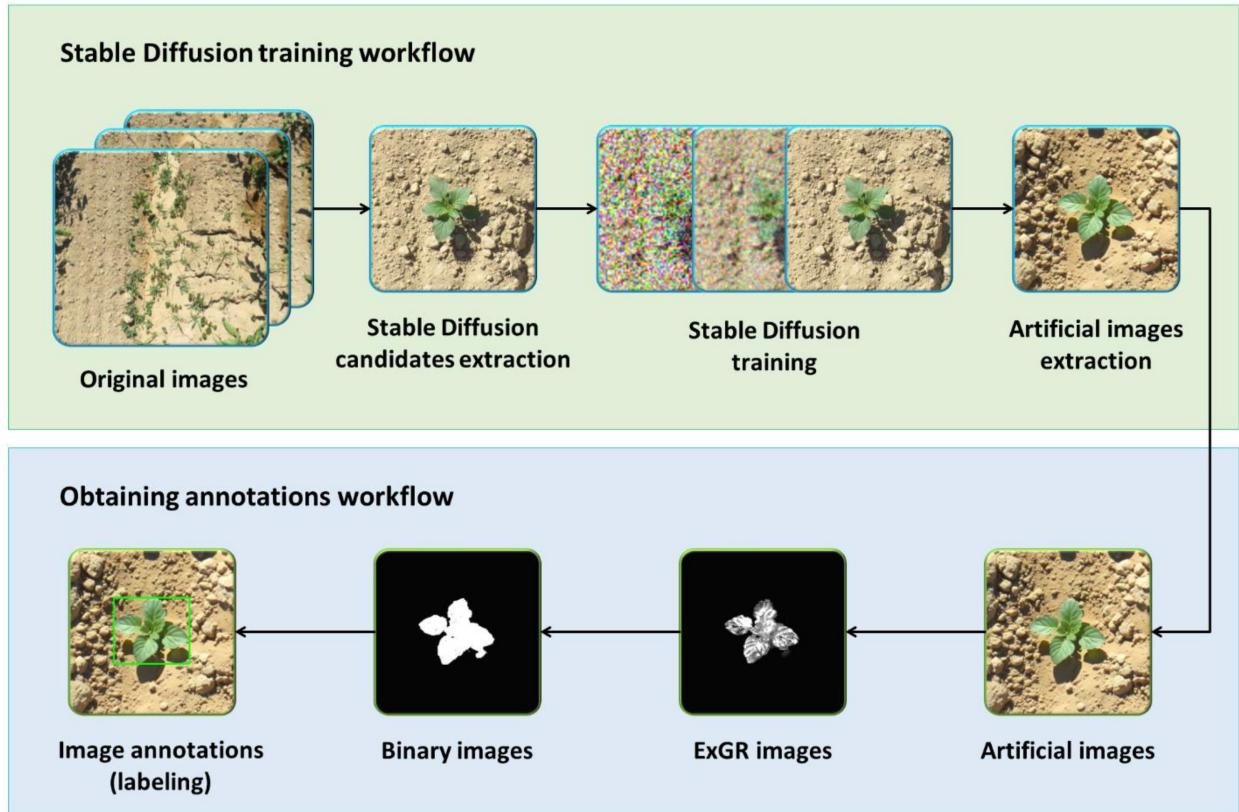


Figure 2: Proposed pipeline to train Stable Diffusion to generate artificial images.[34]

efficiency of these models, enriching agricultural datasets, and exploring multimodal data fusion techniques to further enhance the application potential of diffusion models in agriculture.

### 3. Overview of Diffusion Model Technology

Diffusion models (DM) are a class of deep generative models based on probabilistic graphical models. They simulate the random perturbation and deperturbation processes of data [35] to gradually generate high-quality samples from noise. The core idea of diffusion models is to gradually transform data into noise through a forward diffusion process, and then recover the original data from the noise through a reverse diffusion process. This method originated from diffusion processes in non-equilibrium thermodynamics and has achieved breakthrough results in multiple fields such as image generation [19, 36], speech synthesis [37], medical imaging [38], and agricultural remote sensing [39, 40] in recent years. It performs exceptionally well in high-dimensional data processing, image generation, and denoising tasks, particularly demonstrating unique advantages in hyperspectral image processing [40].

#### 3.1. Basic Principles of Diffusion Models

Diffusion models are a class of deep generative models based on Markov processes. As shown in Figure 3. Their core idea is to simulate the forward diffusion process of data contaminated by noise, then train a neural network to reverse this process and restore the original data, i.e., the reverse process, thereby achieving controllable generation from noise to data.

##### 3.1.1. Forward Diffusion Process

Given the initial sample  $x_0 \sim q(x_0)$ , the forward diffusion process is defined as gradually adding Gaussian noise over T time steps, in the form:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I) \quad (1)$$

Where  $\beta_t \in (0, 1)$  is the noise variance scheduling parameter for time step  $t$ , which is typically applied linearly, cosine, or learning-based across the entire time series.

This process constitutes a Markov chain. After  $t$  cumulative executions, the intermediate state  $x_t$  at step  $t$  can be directly sampled from  $x_0$ , satisfying:

$$q(x_t | x_0) = \mathcal{N}\left(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)I\right) \quad (2)$$

Where, The parameter  $\alpha_t$  represents the cumulative factor for retaining image information and is defined as:

$$\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s) \quad (3)$$

Where, The above formula shows that during training, it is possible to skip explicit sampling at each step of the Markov chain and instead generate  $x_t$  at any time step from  $x_0$ , greatly simplifying the training process.

To intuitively understand the evolution path of the diffusion process, the following diagram can be used: Each step of noise addition is slight, but as time  $t$  increases, the image structure and semantic information are gradually lost, eventually approaching an isotropic standard Gaussian distribution.

### 3.1.2. Reverse Generation Process

In diffusion models, the reverse generation process is a critical stage in the model's generation of new samples. The goal of this process is to gradually remove noise from a set of completely noisy data (typically a random vector sampled from a multidimensional Gaussian distribution) and restore an approximation of the original data, ultimately generating high-quality images or other forms of data.

The reverse process is the inverse of the forward diffusion process and is a parameterized Markov chain, represented as:

$$p_\theta(x_{t-1} | x_t) \quad (4)$$

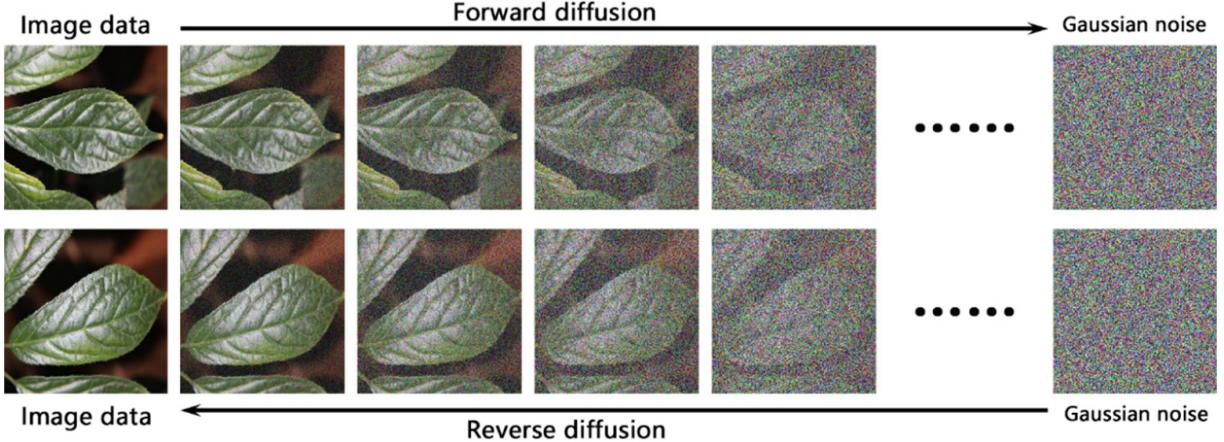
Where, The model generates a series of samples  $X_{T-1}, X_{T-2}, \dots, X_0$  by gradually sampling from noise  $x_T \sim \mathcal{N}(0, I)$ , ultimately obtaining high-quality generated samples  $X_0$ . Each sampling step can be modeled as follows:

$$p_\theta(x_{t-1} | x_t) = \mathcal{N}\left(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)\right) \quad (5)$$

Where,  $\mu_\theta$  and  $\Sigma_\theta$  are the mean and covariance of the denoising network output, respectively, which are usually parameterized by a deep neural network.

### 3.2. Denoising Diffusion Model (DDPM)

The Denoising Diffusion Model (DDPM) [42, 43] is a classic diffusion model framework proposed by Ho et al. in 2020, regarded as a pioneering work in modern diffusion models. DDPM demonstrates extremely high generation quality in tasks such as image generation and image reconstruction, and significantly outperforms early generative models like GAN [15, 30] in terms of stability and theoretical completeness. Its core idea is to simulate the “diffusion process” of an image transitioning from a clean state to pure noise using a Markov chain, and to use a neural network to learn the inverse of this process—the “reverse diffusion process” of recovering the real image from noise. The former is typically manually designed to transform any data distribution into a simple prior distribution (e.g., a standard Gaussian distribution), while the latter Markov chain reverses the former by learning transition kernels parameterized by deep neural networks.



**Figure 3:** Forward propagation and backward generation processes of the diffusion model. [41]

As mentioned earlier, DDPM transforms the original sample  $x_0$  into a Gaussian distribution  $x_t$  by progressively adding Gaussian noise. At step  $t$ , the image is transformed into:

$$x = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (6)$$

Where,  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$  controls the noise intensity.

The reverse process is the conditional probability distribution  $p_\theta(x_{t-1}|x_t)$ , parameterized by a neural network. DDPM does not directly predict the sample itself, but rather predicts the noise  $\epsilon$  added in the forward process and recovers the sample mean in the following form:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(x_t, t) \right) \quad (7)$$

During sampling, an intermediate sample is generated from a conditional Gaussian distribution at each step until  $X_0$  is restored.

DDPM adopts the U-Net architecture [44] as the main denoising network, which is a widely used convolutional neural network in medical image segmentation. It features an encoder-decoder structure and integrates cross-layer features. First, the encoder progressively extracts image features while reducing resolution; second, the decoder progressively upscales the features to restore image resolution; Finally, skip connections preserve local spatial information, aiding in better denoising and detail restoration. Additionally, DDPM introduces time embeddings into the network, typically using sine position encoding, to fuse information from the current diffusion step  $t$  into the network, enabling the model to adaptively denoise across different noise levels.

The training objective of DDPM introduces the KL divergence DKL, aiming to minimize the variational bound of the negative log-likelihood.

$$E_q [-\log p_\theta(x_0)] \leq E_q [D_{KL} (q(x_T|x_0) \| p(x_T))] + \sum_{t>1} D_{KL} (q(x_{t-1}|x_t, x_0) \| p_\theta(x_{t-1}|x_t)) - \log p_\theta(x_0|x_1) \quad (8)$$

Where,  $L_T$  and  $L_0$  represent the prior loss and reconstruction loss, respectively;  $\mathcal{L}_{1:T-1}$  represents the sum of the divergences of the posterior distributions after the forward and backward steps. By simplifying  $L_{t-1}$ , we obtain the simplified training objective  $L_{simple}$  based on the posterior distribution  $q(x_{t-1}|x_t, x_0)$ , whose expression is:

$$q(x_{t-1}|x_t, x_0) = \mathcal{N} \left( x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I \right) \quad (9)$$

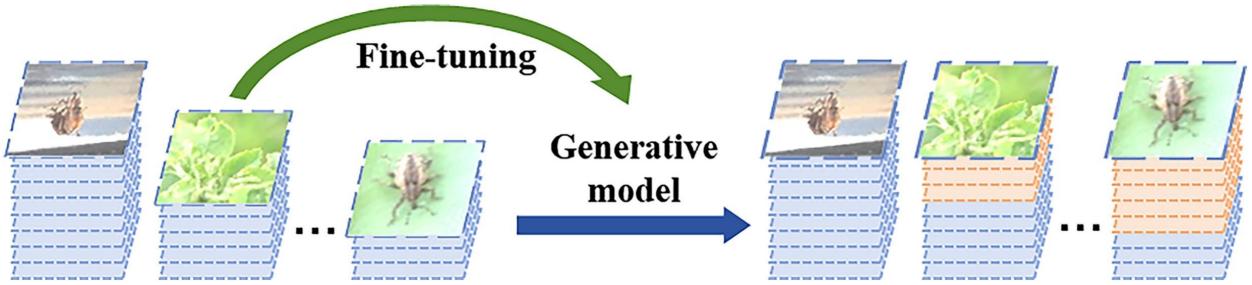


Figure 4: Overview of the data balancing process.[49]

Where  $\tilde{\beta}_t$  depends on  $\beta_t$ . While keeping the above parameterization, reparameterize  $x_t$  as  $x_t(x_0, \sigma)$ ,  $\mathcal{L}_{t-1}$  can be expressed as the expected value of the  $\ell_2$  – loss between two mean coefficients:

$$\mathcal{L}_{t-1} = \mathbb{E}_q \left[ \frac{1}{2\sigma_t^2} \|\tilde{\mu}_t(x_t, x_0) - \mu_\theta(x_t, t)\|^2 \right] + C \quad (10)$$

This is related to the denoising score matching discussed in the next section. By reparameterizing  $\mu_\theta$  W. r. t  $\epsilon_\theta$  to simplify  $\mathcal{L}_{t-1}$ , we obtain the simplified training objective  $\mathcal{L}_{\text{simple}}$ :

$$\begin{aligned} \mathcal{L}_{\text{simple}} := \\ \mathbb{E}_{x_0, \epsilon} \left[ \frac{\beta_t^2}{2\sigma_T^2 \alpha_t (1 - \bar{\alpha}_t)} \right] \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon)\|^2 \end{aligned} \quad (11)$$

Most diffusion models use the DDPM training strategy. However, the improved DDPM proposes combining  $\mathcal{L}_{\text{simple}}$  with other objectives. After training, the prediction network  $\epsilon_\theta$  is used for original sampling in the reverse process.

### 3.3. Extension and improvement of diffusion models

Since the introduction of DDPM, diffusion models have undergone continuous development and optimization in terms of theory, structure, and application. Researchers have proposed various improvement methods to address issues such as slow generation speed and weak conditional control, including enhancing sampling mechanisms (e.g., Denoising Diffusion Implicit Model-DDIM) [45], introducing more robust modeling approaches (e.g., Score-based models) [46], improving generative control capabilities (e.g., Conditional Diffusion Models) [47], and integrating more expressive network architectures (e.g., Transformers). These improvements have significantly enhanced the practical applicability and deployment efficiency of diffusion models in the agricultural sector.

#### 3.3.1. Denoising Diffusion Implicit Models (DDIM)

Traditional diffusion models (such as DDPM) [42, 43] produce high-quality results, their multi-step iterative sampling process results in low computational efficiency, severely limiting their application in agricultural scenarios with high real-time requirements. To address this issue, Denoising Diffusion Implicit Models (DDIM) [48] achieve a significant improvement in generation speed by redesigning the diffusion path, while retaining the core advantages of diffusion models. As shown in Figure 4, in the field of long-tail pest classification, DDIM outperforms traditional models by accelerating the image generation process. DDPM requires multiple denoising steps to generate images, while DDIM simplifies the generation process by skipping certain steps, thereby achieving faster and more efficient generation. This improvement is particularly useful for data augmentation in tasks such as pest classification, where balanced datasets are critical.

The Denoising Diffusion Implicit Models (DDIM) [48][36] represent a significant breakthrough in the field of diffusion models, with its innovation primarily lying in the optimized design of the sampling process. This model achieves significant sampling acceleration while fully preserving the training objectives of traditional diffusion models (DDPM) by restructuring the mathematical framework of the diffusion process. From a technical perspective, the core

innovation of DDIM lies in constructing a non-Markovian diffusion process, which has the following key features: maintaining the same marginal distribution as DDPM to ensure that the model training objective remains unchanged; redesigning the forward and backward processes to give them deterministic mapping properties; and mathematically reconstructing the backward process to allow it to skip intermediate steps and calculate directly. The construction principle of this non-Markovian diffusion process is as follows: while maintaining the consistency of data distributions across time steps with DDPM, it breaks the strict Markov chain constraint, thereby allowing the adoption of a jump-based sampling strategy; achieves deterministic generation paths; and significantly reduces the number of required sampling steps. Since the marginal distribution remains unchanged, DDIM can directly reuse the pre-trained model parameters from DDPM without the need for retraining. This feature enables DDIM to: fully compatible with existing diffusion model architectures; maintain the original model's generation quality; achieve efficiency improvements solely through sampling algorithm enhancements. Experiments demonstrate that this design can enhance sampling speed by 10–50 times while maintaining generation quality, providing critical technical support for the deployment of diffusion models in practical agricultural applications.

DDIM improves upon the standard diffusion model by introducing the concept of “implicit” in the denoising process. Traditional diffusion models require each step to generate the next image through a Markov chain, which, while producing high-quality images, is computationally expensive, especially when numerous denoising steps are required. The key innovation of DDIM lies in the implicit modeling of the reverse process. DDIM does not require explicit noise removal at each step but instead uses an improved reverse process to enable image generation in fewer steps. Specifically, DDIM introduces a parameter to control the intensity of denoising and maintains flexibility during the denoising process, ensuring that each step is no longer a simple linear mapping, thereby enhancing the quality of the generated images.

The reverse process of DDIM is as follows:

$$p_{\theta}(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_{\theta}(x_t, t), \sigma_t^2 I) \quad (12)$$

Where,  $\mu_{\theta}(x_t, t)$  is the denoised mean predicted by the neural network, and  $\sigma_t$  is the standard deviation of the control noise added. This model achieves a more efficient reverse process by selecting appropriate noise addition methods and generation parameters.

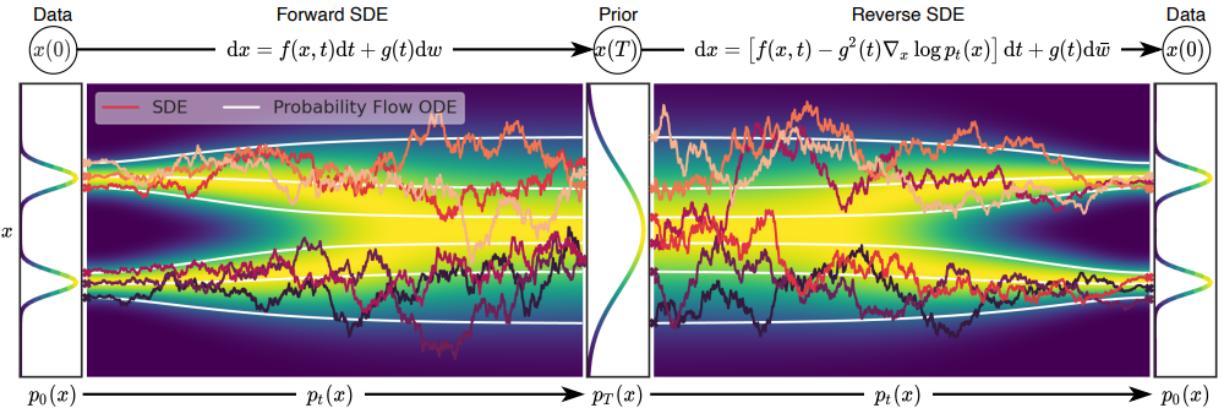
Unlike traditional diffusion models, DDIM does not require multiple Markov chain sampling to generate images. It employs an implicit modeling approach to generate high-quality images in fewer steps, with its efficiency primarily attributed to the refinement of each generation step and flexible denoising methods. By introducing an implicit reverse denoising process, DDIM addresses the high computational overhead of traditional diffusion models. It achieves high-quality image generation by efficiently denoising through a neural network in a small number of steps during the generation process, significantly improving efficiency in image generation speed.

### 3.3.2. Score-based Models (SDE)

Score-based models [46]are a class of generative models that have made significant progress in the field of generative modeling in recent years. The core idea of these models is to generate data by learning the score function (i.e., the logarithmic gradient of the probability density) of the data. Score models adopt a different approach from traditional generative models (such as generative adversarial networks (GANs) [15, 30]and variational autoencoders (VAEs) [50]). They have a strong theoretical foundation and broad application potential, particularly demonstrating outstanding performance in tasks such as image generation [51], super-resolution [52, 53], and image restoration [54, 55].

The objective of generative models is to generate new samples from the latent distribution of data. Early generative models, such as variational autoencoders (VAEs) and generative adversarial networks (GANs), typically learn to generate samples by either directly modeling the data distribution or through adversarial training of generative and discriminative models. However, these methods often face challenges during training, such as training instability and mode collapse. Unlike these methods, score-based models generate samples by learning the gradient of the data distribution (i.e., the score function). The score function is the gradient of the log-probability density of the data, providing the direction of change in the data distribution. Through the score function, the structure of the data can be inferred, and new samples can be generated.

Suppose we have a true data distribution  $p_{\text{data}}(x)$ , where  $x$  is a data sample. The score function  $s_{\theta}(x)$  is the gradient of the log-density function of this data distribution, i.e.:



**Figure 5:** Score-based generative model using SDE.[46]

$$s_\theta(x) = \nabla_x \log p_{\text{data}}(x) \quad (13)$$

This means that the score function describes how the data distribution changes in the input space, and it can tell us how to adjust the data samples to better fit the underlying distribution of the data.

To train the score model, we use the “score matching” method to approximate the true score function of the data. The goal of score matching is to minimize the difference between the model’s score function and the true score function of the data. Since directly calculating the true score function of the data (i.e., the gradient of the log probability) is infeasible, the score matching method approximates the true score function by introducing noise and simplifying assumptions. The loss function for score matching can be expressed as:

$$L_{\text{SM}} = \mathbb{E}_{x \sim p_{\text{data}}} [\|s_\theta(x) - \nabla_x \log p_{\text{data}}(x)\|^2] \quad (14)$$

Where, by minimizing this loss function, the model can learn a good scoring function.

During training, the scoring model approximates the scoring function of the data by minimizing the scoring matching loss. Specifically, the model parameterizes the scoring function using a neural network and learns the gradient information of the data. During training, the model not only needs to consider the data distribution, but also needs to approximate the true distribution of the data by adding and reconstructing appropriate noise.

The sampling process of the fractional model is typically generated using the gradient ascent method. This process starts with a noise sample and gradually adjusts the sample through a fractional function to ultimately generate data that conforms to the target distribution. Assuming we start with a standard Gaussian noise sample  $x_T \sim \mathcal{N}(0, I)$ , we generate samples through the following iterative process:

$$x_{t-1} = x_t + \Delta t \cdot s_\theta(x_t) \quad (15)$$

Where, In this context,  $\Delta t$  represents the time step size for each iteration, and  $s_\theta(x_t)$  denotes the score function at the current time step  $t$ , indicating the adjustment direction for the current sample. By repeatedly applying the score function, the sample gradually transitions from noise to real data.

Using multiple noise scales to perturb data was a key factor in the success of previous methods. We further extend this idea to an infinite number of noise scales, allowing the perturbed data distribution to evolve according to a stochastic differential equation (SDE) as the noise intensity increases. The framework is summarized in Figure 5.

Score-based models represent a significant breakthrough in the field of generative modeling. Instead of directly modeling the distribution of data, they generate samples by learning the score function of the data. Score models not only offer significant advantages in terms of generation quality, training stability, and theoretical support, but

also provide new insights into other generative models such as diffusion models. By combining score functions with diffusion models, score models demonstrate broad application potential, particularly excelling in tasks such as image generation and restoration. As research progresses, score-based models are expected to play an increasingly important role in a broader range of generative tasks.

### 3.3.3. Noise-conditioned score networks

The Score-based Diffusion Model (SGM/NCSN) [56, 57, 58] is a generative model based on the concept of score matching, first proposed by Yang Song et al. in 2019. NCSNs aim to achieve high-quality data generation by learning the score function (i.e., the gradient of the log probability density) of data distributions under different noise levels. The score function of a given data distribution  $p(x)$  is defined as the gradient of the log-probability density function:  $s(x) = \nabla_x \log p(x)$ . Here, the score function indicates how to move along the gradient direction in the data space to bring data points closer to regions with higher probability densities. In other words, it reveals the “preference” direction of the data distribution at each point. To estimate this score function, we can train a shared neural network with score matching. Specifically, the score network  $S_\theta$  is a neural network parameterized by  $\theta$ , trained to approximate the score of  $p(x)$ , i.e.,  $s_\theta(x) \approx \nabla_x \log p(x)$ , with the optimization objective being:

$$\mathbb{E}_{x \sim p(x)} \|s_\theta(x) - \nabla_x \log p(x)\|_2^2 \quad (16)$$

However, due to the computational burden of calculating  $\nabla_x \log p(x)$ , score matching cannot be extended to deep networks and high-dimensional data. To address this issue, [56] proposed denoised score matching [59] and slice score matching [60]. Additionally, Song and Ermon [56] highlighted the main challenges of score-based generative modeling applications. The key challenge is that the estimated score function is inaccurate in low-density regions, as real-world data often concentrates on low-dimensional manifolds embedded in high-dimensional spaces (manifold assumption). The authors demonstrated that these issues can be addressed by perturbing data with Gaussian noise at different scales, as it makes the data distribution more suitable for score-based generative modeling. They suggest estimating scores for all noise levels by training a single noise-conditioned scoring network (NCSN). By selecting the noise distribution  $p_{\sigma_t}(x_t|x) = \mathcal{N}(x_t; x, \sigma_t^2 \cdot I)$ , the gradient of  $\nabla_x \log p(x)$  with respect to  $x$  is derived as the gradient of  $\nabla_{x_t} \log p_{\sigma_t}(x_t|x) = -\frac{x_t - x}{\sigma_t}$  where  $x_t$  is the noisy version of  $x$ . Therefore, for a given sequence of Gaussian noise scales  $\sigma_1 < \sigma_2 < \dots < \sigma_T$ , equation 13(13) can be expressed as:

$$\frac{1}{T} \sum_{t=1}^T \lambda(\sigma_t) \mathbb{E}_{p(x)} \mathbb{E}_{x_t \sim p_{\sigma_t}(x_t|x)} \left\| s_\theta(x_t, \sigma_t) + \frac{x_t - x}{\sigma_t} \right\|_2^2 \quad (17)$$

Where,  $\lambda(\sigma_t)$  is the weighting function, and this derivation is performed using an iterative process known as “Langevin dynamics.” Langevin dynamics designs an MCMC program that samples from the distribution  $p(x)$  using only the fractional function  $\nabla_x \log p(x)$ . Specifically, to convert a random sample  $x_0 \sim \pi(x)$  into a sample from  $p(x)$ , it undergoes the following iterative steps:

$$x_i = x_{i-1} + \frac{\gamma}{2} \nabla_x \log p(x) + \sqrt{\gamma} \cdot \omega_i \quad (18)$$

Where  $\omega_i \sim \mathcal{N}(0, I)$ , and  $i \in \{1, \dots, N\}$ . As  $\gamma \rightarrow 0$  and  $N \rightarrow \infty$ , the samples  $x_i$  obtained from this process will converge to samples from the distribution  $p(x)$ . [42] proposed renaming the algorithm to annealed Langevin dynamics algorithm, as the noise scale  $\sigma_i$  gradually decreases (anneals) over time to mitigate some of the defects and failure modes of fractional matching [57].

### 3.3.4. Conditional Diffusion Models (CDMs)

The core idea behind diffusion models is to simulate a process of gradually adding noise, and then learn how to recover the original data from pure noise. This approach differs significantly from traditional generative models (such as GANs) in terms of their generative mechanisms, particularly in terms of training stability and generative quality. In recent years, conditional variants of diffusion models—conditional diffusion models (CDMs)—have emerged as an important innovation in generative models. Conditional diffusion models introduce conditional information (e.g.,

text, labels, images, etc.) into the generation process, enabling the generated samples to not only maintain high quality but also be directed and controlled according to specific conditions. This conditional generation capability has led to widespread applications of conditional diffusion models in tasks such as text-to-image generation, speech synthesis, and image restoration.

Diffusion models exhibit high flexibility, capable of generating data samples from both unconditional  $p_0$  distributions and conditional  $p_0(x|c)$  distributions. Here,  $c$  represents the given condition, such as a category label or text associated with the data  $x$  [61]. To achieve conditional generation, various sampling algorithms have been designed, including classifier-free guidance [62] and classifier-guided sampling [63].

Sampling with labeled conditions guides the generation process by introducing conditional gradients at each sampling step. Typically, this requires an additional classifier, which adopts a UNet encoder architecture to generate conditional gradients associated with specific labels. These labels can be text information, category labels, binary markers, or features extracted from the data [63, 64, 65, 66, 67, 68, 69, 70, 71, 72], a method first proposed by [63] to support current conditional sampling techniques.

**Unlabeled conditions:** Unlabeled conditional sampling uses self-information for guidance, typically applied in a self-supervised manner [73, 74], and is commonly used for denoising [75], image-to-painting conversion [76], and restoration tasks [77].

### 3.4. Relationship with other generative models

Next, we will explore the relationship between diffusion models and other generative models, starting with likelihood-based methods and concluding with generative adversarial networks (GANs).

Diffusion models share many similarities with variational autoencoders (VAEs). For example, both map data to a latent space and learn to transform latent representations back into data. Additionally, the objective functions of diffusion models and VAEs can both be derived as lower bounds on data likelihood. However, there are still some fundamental differences between the two. The latent representation of a VAE contains compressed information from the original image, while a diffusion model completely transforms the data into noise through the final step in the forward process. Unlike a VAE, the latent representation of a diffusion model has the same dimension as the original data, while a VAE reduces the dimension through compression, enabling it to perform better in a low-dimensional latent space. More importantly, the latent space mapping in VAE is trainable, meaning that VAE can optimize the latent space representation through training. In diffusion models, the latent space representation is obtained by progressively adding Gaussian noise to the image, so the latent space representation is not derived through training. Despite this, existing research has attempted to construct more effective diffusion models by applying diffusion models to the latent space of VAE [78, 79].

Autoregressive models [80, 81] generate data by representing images as sequences of pixels. The generation process is conditional, meaning that the generation of each pixel depends on previously generated pixels. This generation method is unidirectional, clearly highlighting the limitations of such models. Esser [82] et al. proposed combining diffusion models with autoregressive models to address these limitations. Their method learns the reverse polynomial diffusion process via a Markov chain and implements each transition step using an autoregressive model. In this approach, the previous step of the Markov chain provides global information to the autoregressive model, thereby addressing the issues arising from unidirectional generation.

Normalizing Flows [83, 84] are a class of generative models that transform simple Gaussian distributions into complex data distributions through a set of reversible transformations. These transformations are reversible and have easily computable Jacobian determinants, making probability density calculations more efficient. Compared to diffusion models, normalizing flows perform mapping operations in a deterministic manner by learning invertible and differentiable functions. However, the similarity between diffusion models and normalizing flows is limited to the fact that both involve mappings between data distributions and Gaussian noise. Normalized flows perform data generation by learning reversible mappings, while diffusion models generate data through gradual denoising. Although these two approaches differ in their generation methods, the DiffFlow method [85] combines the advantages of both models by integrating diffusion models and normalized flows, making both the reverse and forward processes trainable and stochastic.

Energy-based models (EBMs) [86, 87, 88, 89] focus on providing a non-standardized energy function (i.e., an estimate of the density function). Unlike likelihood-based models, EBMs represent the data distribution through a scoring function. One of the main advantages of EBMs is that they do not depend on a specific network structure and can be modeled using any regression neural network. However, training EBMs is typically challenging because they

require complex sampling strategies, such as Markov Chain Monte Carlo (MCMC) methods based on score functions, when generating samples. In fact, the training and sampling processes of diffusion models can be viewed as a special case of EBMs, particularly when training and sampling depend solely on the score function.

In terms of sample quality, until the recent emergence of diffusion models [63], generative adversarial networks (GANs) [90] were considered one of the best-performing models in the field of sub-stationary generative models. However, the adversarial training objective of GANs makes their training process challenging [91] and often leads to mode collapse issues. In contrast, the training process of diffusion models is more stable, and due to their likelihood-based nature, they can provide more diverse generative results. Although diffusion models have significant advantages in terms of training stability and diversity, their computational efficiency during inference remains low, requiring multiple network evaluations.

Diffusion models and GANs have significant differences in their latent spaces. GANs typically have low-dimensional latent spaces, while diffusion models preserve the original dimensions of images. The latent space of diffusion models is typically modeled as a random Gaussian distribution, similar to the latent space in VAE. In terms of semantic attributes, the latent space of GANs contains subspaces related to visual attributes, and manipulating changes in the latent space can control the attributes of images [15, 92]. In contrast, the latent space of diffusion models does not explicitly represent such semantic attributes. Therefore, in diffusion models, generating images with different attributes often relies on guidance techniques [62, 63] rather than direct manipulation of the latent space. Nevertheless, Song [46] et al. demonstrated that the latent space of diffusion models possesses a robust structure, suggesting that interpolation within this space can lead to smooth transitions in the image space. Therefore, although the exploration of the latent space in current diffusion models is far less advanced than in GANs, this area may become one of the key focuses of future research.

Diffusion models share some significant similarities and differences with other generative models (such as GANs, VAE, normalizing flows, and EBM) in terms of generation mechanisms, training methods, and latent spaces. Diffusion models demonstrate significant advantages in terms of training stability and generation quality, particularly in avoiding mode collapse and achieving diverse generation, surpassing GANs. However, diffusion models have relatively low computational efficiency and require multiple network evaluations during inference. In the future, combining the advantages of different generative models, improving inference efficiency, and deeply exploring the semantic properties of the latent space of diffusion models may become key research directions in this field. As shown in FigureFigures 6, I list the application of different models in the agricultural field over time and as shown in Table 1, the performance of different generative models in agriculture is compared.

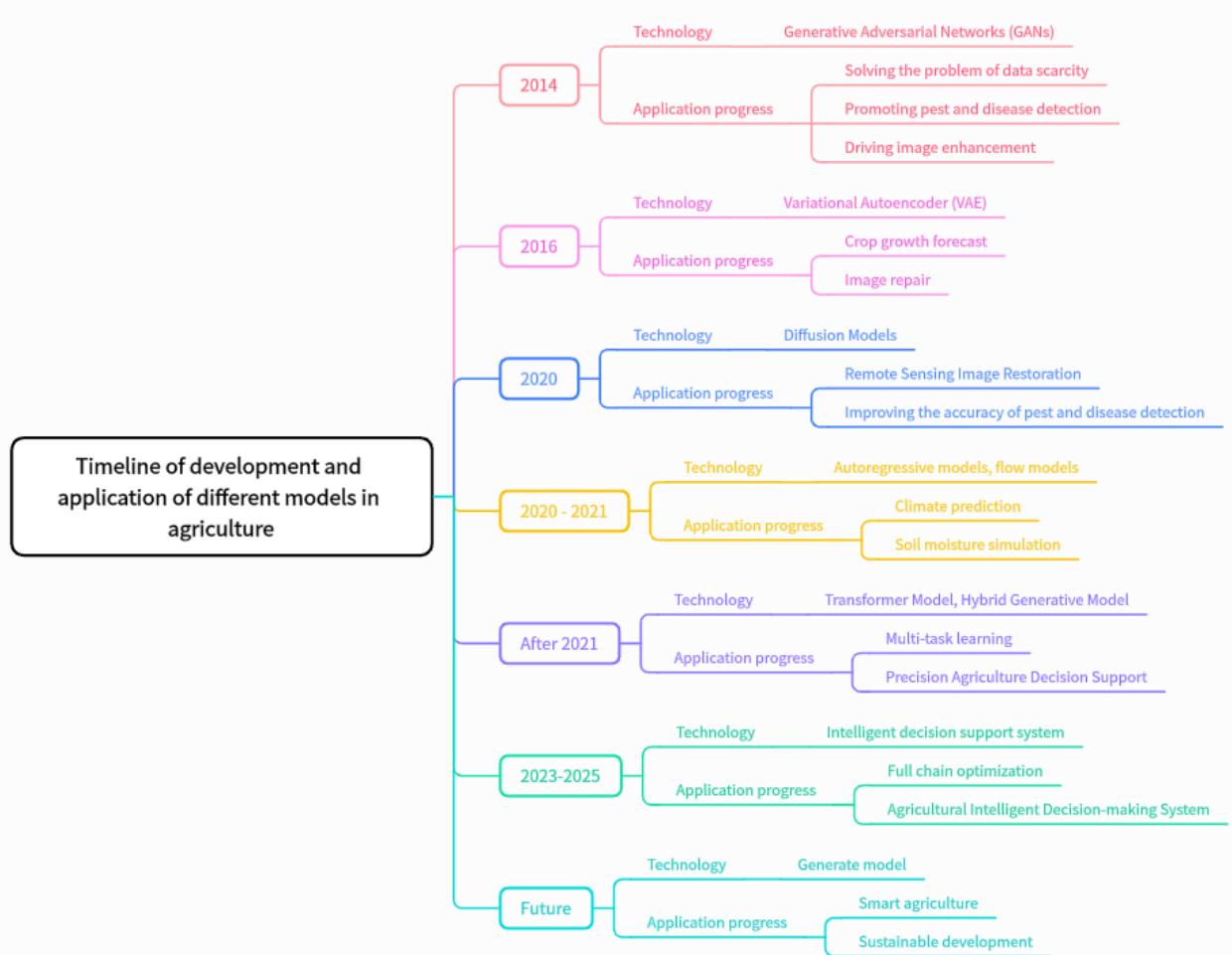
## 4. Application of diffusion models in image processing and analysis in agriculture

This section discusses the latest advancements in the application of diffusion models in agricultural image processing and divides the application areas in the agricultural field into seven categories, as shown in Figure 12. Over the past few years, generative models represented by diffusion models have played a significant role in the agricultural sector, particularly in crop health monitoring [93, 94, 95], pest and disease identification [96, 97, 98, 99, 100], and decision support [18]. By generating high-quality images, denoising, enhancing, and performing data augmentation, diffusion models can effectively improve the accuracy of crop monitoring and pest and disease detection, generate diverse agricultural images [36, 95, 101, 49], expand training datasets [17, 41], and enhance the performance of deep learning models.

### 4.1. Image Classification

Diffusion models are increasingly being used in agricultural image classification as an important technology for addressing data scarcity, class imbalance, and improving model generalization, particularly in plant disease detection [93, 41, 102] and crop growth monitoring [103]. The agricultural sector often faces challenges such as insufficient data and imbalanced samples of disease types. Diffusion models effectively address these issues by generating high-quality, diverse synthetic images, enhancing the diversity of training data, and significantly improving classifier performance. By generating precise synthetic data, diffusion models not only enhance model accuracy but also improve their adaptability under various conditions, making them a crucial tool for advancing agricultural intelligence and precision agriculture.

To improve the effectiveness of sunflower disease classification, Zhou et al. [93] proposed a diffusion model-based few-shot learning framework, combining diffusion generative models with few-shot learning to construct an



**Figure 6:** Application of different models in agriculture.

efficient end-to-end detection framework. To enhance the dataset for ginseng leaf disease classification, Wang et al. [41] proposed an image generation method based on diffusion models (DDPM) and ECA attention mechanisms. This method combines the Inception-SSNet model with skip connections, attention feature fusion, and self-calibrating convolutions (ScConv) to enhance the ability to identify small lesions and similar disease features, thereby further improving classification accuracy. To improve the accuracy of grape leaf disease classification, Zhang et al. [102] combined the diffusion model (EDM) with a data augmentation strategy. To address data scarcity and class imbalance issues, EDM was used to generate synthetic disease images, expanding the training dataset, which was then combined with convolutional neural network (CNN) models (such as ResNet50, VGG16, and AlexNet) for classification. As shown in Figure 7, images generated by EDM.

Du et al. [49] proposed an innovative ADM-DDIM framework for generating high-quality pest images. Experimental results showed that synthetic data generated using diffusion models significantly improved the performance of classification models. To address the issue of insufficient data for the intermediate state of “slightly wilted” in plant health conditions, which leads to low classification training accuracy, Lee [103] utilized DDPM. By leveraging images of normal and wilted states, they generated “slightly wilted” state using interpolation in the latent space, thereby improving classification accuracy when identifying the “slightly wilted” state.

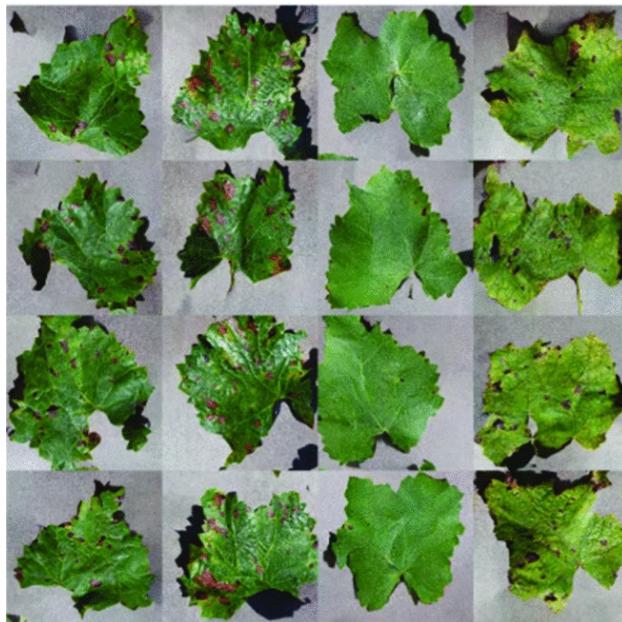
#### 4.2. Image Generation

Diffusion models have also achieved significant results in image generation. In the agricultural field, diffusion models are widely used for the generation and restoration of crop images [36, 101], such as generating missing parts

**Table 1**

Comparison of performance of different generative models in agriculture

Characteristic	Diffusion Models[42, 43]	Adversarial Networks (GAN)[15, 30]	Autoencoders (VAE)[50]	Normalizing Flows[83, 84]
Generation Mechanism	Step-by-step denoising to generate data	Generator and discriminator engage in adversarial training	Encoder and decoder generate data	Data distribution is mapped through inverse reasoning and transformation
Training Stability	Stable, avoids mode collapse	Prone to mode collapse	Stable, but generates lower-quality data	High stability, but computationally expensive
Generation Quality	High quality, diverse	Quality is unstable, may lose diversity	Generates poor quality, especially for complex distributions	Generates lower quality, simpler
Computational Efficiency	Lower, requires multiple network evaluations	Efficient, but may fall into local optima	Efficient, suitable for large-scale data	High computational cost, but strong expressive power
Mode Collapse Issue	Can avoid mode collapse	Prone to mode collapse	No significant issues	No significant issues
Applicability	High-quality generation tasks	Suitable for image, video, and similar tasks	Suitable for generation, representation learning, and dimensionality reduction	Suitable for high-dimensional data generation and modeling

**Figure 7:** Image generated by EDM.[102]

of crop images or simulating crop images at different growth stages [95, 104]. These generated images not only have high-quality visual effects but also assist agricultural experts in crop condition analysis, pest and disease prediction, and growth environment simulation. In this way, diffusion models provide robust technical support for precision agriculture, optimizing agricultural decision-making processes.

	Stable Diffusion v1.4	Stable Diffusion v1.5	Stable Diffusion v2	Stable Diffusion XL 1.0	SDXL Turbo
Cherry					
Banana					
Orange					
Apple					
Pineapple					

**Figure 8:** Sample outputs from generative models for various fruit types.[36]

In the field of precision agriculture, Zhao et al. [36] used the Stable Diffusion series of models (such as v1.4, v1.5, v2, and SDXL) to generate a fruit image dataset. As shown in Figure 8. By generating synthetic images of different fruits (such as cherries, bananas, oranges, apples, and pineapples), researchers enhanced the diversity of the dataset and the accuracy of the model. Although there are differences in accuracy and processing speed among the models, SDXL Turbo demonstrated the highest accuracy across all fruit categories, showcasing the immense potential of generative AI in fruit detection and segmentation for agricultural robots.

Plant diseases pose a serious threat to agricultural production, and traditional detection methods typically rely on manual inspection and expert knowledge, which are time-consuming and prone to errors. Generative models, particularly Stable Diffusion, offer an effective solution to address data scarcity and class imbalance issues. Huang et al. [95] generated high-quality synthetic data and combined it with ControlNet and LoRA fine-tuning techniques, enabling diffusion models to generate diverse plant disease images, thereby enhancing dataset diversity and improving model generalization capabilities. Heschl et al. [101] combined the denoising diffusion probabilistic model (DDPM) with generative adversarial networks (GANs) to design a dual diffusion model architecture capable of synthesizing images with high-quality phenotypic features and consistent labels, thereby improving the accuracy and efficiency of crop region segmentation.

To improve weed recognition accuracy and address data scarcity issues, Chen et al. [104], Deng et al. [105], and Moreno et al. [34] explored the application of diffusion model-based data augmentation methods in agricultural weed recognition. Chen et al. generated synthetic weed images and combined them with the Stable Diffusion model and convolutional neural networks (CNNs) to significantly improve the performance of weed detection systems, particularly in complex scenarios. By using the generated synthetic data to augment the training data, they optimized the model's adaptability. Deng et al. combined synthetic images with actual images, significantly improving weed classification accuracy, especially in cases of overlapping weed species. The generated data enhanced the model's generalization ability. Moreno et al. found that synthetic images generated by the Stable Diffusion model effectively improved weed detection accuracy, especially in scenarios with overlapping weed species and complex backgrounds. Data augmentation enhanced the model's adaptability and robustness in real agricultural environments. These studies demonstrate that diffusion models can provide more efficient and precise technical support for agricultural pest and disease monitoring, particularly in scenarios with scarce data and significant environmental variability.

In agricultural pest datasets, Du et al. [49] encountered a long-tail problem due to the imbalance in the frequency of occurrence of different pest types, which is a major challenge in agricultural research. Diffusion models were used to generate high-quality pest images, balancing the dataset distribution through synthetic image generation to address the long-tail issue. By fine-tuning pre-trained models and generating realistic pest images, diffusion models can enhance the diversity of the dataset, balance the representation of different categories, and thereby improve the accuracy of pest detection. Experimental results on the IP102 pest dataset show that diffusion models outperform traditional methods



**Figure 9:** Visual comparison of SR generalization studies on the Agricultural Visual (AgV) dataset. The left side shows the FOV of AgV data samples, and the right side shows the SR results, where EVADM produces more realistic views and finer details.[107]

in terms of image quality and diversity, effectively alleviating the long-tail problem and improving classification task performance.

#### 4.3. Data Enhancement

Diffusion models have made significant progress in generative tasks in recent years, particularly in agricultural applications. In terms of data augmentation, diffusion models play an important role in crop monitoring [103, 106] and plant disease detection [41, 102] by gradually generating diverse, high-quality data. By processing existing data to generate new data that is similar to the original but with greater variability, diffusion models effectively enhance the diversity of training datasets, enabling agricultural models to better identify different crop conditions and pests/diseases. In this way, diffusion models provide richer data support for agricultural applications under resource-constrained conditions, thereby improving the model's generalization capabilities.

Image super-resolution (SR) technology can significantly improve the resolution and quality of aerial imagery. In agricultural applications, particularly in the analysis of remote sensing imagery, super-resolution technology helps enhance image detail and visualization effects. Stable Diffusion has demonstrated exceptional performance in the super-resolution task of agricultural aerial imagery. Lu et al. [107] constructed the CropSR dataset and applied the VASA and EVADM models. As shown in Figure 9. Their findings revealed that diffusion models significantly improved the quality of generated images in image super-resolution tasks, enhancing the Fréchet-Inception-Distance (FID) and SRFI metrics, thereby demonstrating the immense potential of diffusion models in enhancing image quality and detail in agricultural image processing.

In plant disease classification tasks, data scarcity and class imbalance are common issues. Especially for certain plant disease categories, such as ginseng leaf disease, the quantity and quality of disease images limit the accuracy of classification models. By employing a data augmentation method based on diffusion models, Wang et al. [41] were able to generate high-quality synthetic images, enhance the diversity of the dataset, and improve model accuracy. Zhang et al. proposed the Elucidation Diffusion Model (EDM) [102], which uses diffusion models to generate sampling paths at different noise levels and fine-tunes models such as ResNet50, VGG16, and AlexNet for disease classification. This method improves the model's classification accuracy while enhancing dataset quality, particularly when dealing with diseases featuring subtle characteristics. Hirahara et al. proposed a diffusion model-based data augmentation method, D4[108], for vine branch detection in vineyards. This method uses text-guided diffusion models to generate



**Figure 10:** Shows the results of generating “slightly withered” data.[103]

new annotated images, overcoming the issue of insufficient training data. The generated synthetic images are not only tailored to the target domain but also retain the annotation information required for object detection. Experiments show that the D4 method significantly improves the average accuracy in object detection tasks and addresses the challenges of annotation difficulty and domain diversity in the agricultural field, demonstrating the powerful capabilities of diffusion models in data augmentation. For the early detection of plant diseases, traditional image enhancement methods cannot provide sufficient diversity and detail. Lee’s DDPM-based enhancement method [103] can simulate various variants of different diseases, As shown in Figure 10. For example, using diffusion models to generate images of plants in a “slightly wilted” state helps improve the accuracy of classification models, enhances the capture of subtle changes in plant health status, and strengthens the robustness of plant health monitoring models.

To address the issue of real-time detection of broken grains and impurities during grain harvesting, Zhang et al. [106] proposed a dual attention diffusion model (DADM) that combines a denoising diffusion probabilistic model (DDPM) and a spatial channel attention block (SCA-Block). DADM outperforms DDPM and generative adversarial networks (GANs) in terms of image quality and diversity. The dataset enhanced by DADM not only significantly improves the accuracy of seven segmentation models but also performs well in other agricultural tasks such as pest and disease detection.

#### 4.4. Remote sensing and hyperspectral image reconstruction

With the advancement of remote sensing technology, the reconstruction and enhancement of remote sensing images (especially high-resolution and hyperspectral images) have become important research directions in various fields such as environmental monitoring, agricultural management, and disaster response. Traditional image reconstruction and enhancement methods, such as interpolation [109, 110] and conventional convolutional neural networks (CNNs) [111], have certain limitations when handling large-scale data and complex images, particularly in cases of significant data loss or low image resolution. In recent years, generative methods based on Diffusion Models, especially their application in remote sensing and hyperspectral image reconstruction, have demonstrated significant advantages.

Satellite images often contain missing data due to cloud cover, sensor failures, or incomplete acquisition, especially in high-resolution and high-frequency missions. Traditional interpolation methods have limited effectiveness in handling large-scale missing data, while diffusion models can effectively reconstruct missing data through gradual denoising. Yu et al. proposed SatelliteMaker [112], a diffusion model-based method for reconstructing missing data in remote sensing images, which can fill spatial gaps and spatio-temporal data gaps. Combined with a digital elevation model (DEM) and customized prompts, it performs exceptionally well in agricultural management tasks. SatelliteMaker effectively addresses missing data issues caused by cloud cover and sensor failures, particularly in

high-resolution images and high-frequency tasks, overcoming the limitations of traditional methods. Research indicates that cloud removal methods based on latent diffusion models can significantly improve image quality under complex conditions, especially when combined with the AdamW optimizer and Huber loss function [113], effectively removing clouds while preserving ground details.

Hyperspectral images, due to their multi-band characteristics, are widely used in agricultural remote sensing, particularly for crop health monitoring and soil quality analysis. However, sensor resolution limits the quality of many hyperspectral images. Diffusion models excel in hyperspectral image super-resolution tasks, converting low-resolution images into high-resolution ones through a gradual denoising and refinement process. Tang et al. [40] used the CropSR dataset and the EVADM model (VASA Enhanced Diffusion Model) to significantly improve image detail and resolution. This provides a new method for detail restoration and high-resolution crop monitoring in agricultural remote sensing images, enhancing analysis accuracy.

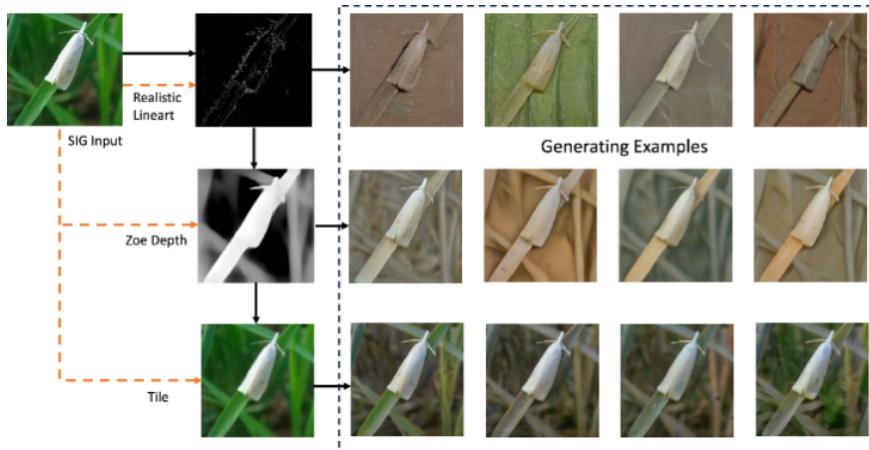
Land use and land cover (LULC) semantic segmentation is a critical task in remote sensing. Although traditional convolutional neural networks (CNNs) and vision transformers are widely applied in this task, they fail to fully utilize spatial and color texture information in high-resolution remote sensing images, leading to significant loss of details during feature extraction and limiting segmentation performance. Fan et al. proposed DDPM-SegFormer[114], which combines the denoising diffusion probability model (DDPM) and Vision Transformer (ViT) for high-precision LULC segmentation. This framework overcomes the information bottleneck of traditional models when processing remote sensing images by integrating the characteristics of diffusion models and Transformers, particularly excelling in detail capture and multi-scale semantic feature extraction. Experimental results on the LoveDA and Tarim Basin LULC datasets demonstrate that DDPM-SegFormer outperforms existing methods such as SegFormer and SegNext in terms of accuracy and recall.

Diffusion models significantly enhance the detail recovery and analysis accuracy of agricultural remote sensing images by enhancing the super-resolution of aerial images of farmland, providing support for agricultural applications such as crop monitoring and land use assessment. Lu et al. proposed a super-resolution method combining diffusion models [107], variance attention mechanisms, and self-supervised training, significantly improving the quality of aerial images. Han et al. proposed an efficient hybrid conditional diffusion model (EHC-DMSR) [115] that combines Transformer and CNN to extract features from low-resolution images, and applies Fourier high-frequency spatial constraints and feature distillation to improve super-resolution performance and reduce inference time. Experiments show that this method outperforms existing algorithms on multiple datasets. Graikos et al.'s self-supervised learning embedded diffusion model [116] generates diverse agricultural image variants to improve accuracy in crop classification and land cover monitoring tasks, addressing the challenge of high-quality image generation in agricultural remote sensing.

#### 4.5. Pest and Disease Diagnosis and Auxiliary Decision Making

To explore the application of diffusion models in pest and disease diagnosis and decision support in the agricultural sector, this study focuses on analyzing their advantages in feature extraction, disease area identification, background suppression, and decision support. Yin et al. [18] applied diffusion models to agricultural disease detection, overcoming the limitations of traditional methods. By introducing an endogenous diffusion subnetwork, they gradually optimized feature distribution, enhancing the accuracy of disease region identification in complex backgrounds, particularly for fine-grained diseases such as rust and root rot. Additionally, the designed endogenous diffusion loss function enabled multi-task optimization, dynamically adjusting the weights of diffusion steps to improve model robustness. Huang et al. proposed a stable diffusion model data augmentation method [95] that combines ControlNet and Embedding to precisely control the generation process and uses low-rank adaptation (LoRA) to fine-tune the pre-trained model, generating high-quality disease data and improving the accuracy of few-shot disease classification. Experiments show that this method outperforms traditional methods in terms of FID scores and Top-1 accuracy, enhancing the model's generalization ability. Wang et al. proposed a method combining semantic diffusion and knowledge distillation to generate agricultural pest images [97]. By optimizing pest detection through the Semantic Integration Guided (SIG) network and Multi-Level Alignment Distillation (MLAD) framework, the method improved the quality and diversity of generated images, enhanced the diversity of training data, and significantly improved detection performance in complex backgrounds. As shown in Figure 11, The experimental results show that this method performs excellently in real-time pest detection in actual agricultural environments.

Li et al. explored the application of denoising diffusion probabilistic models (DDPM) and transfer learning in citrus disease diagnosis [98], proposing two methods to address the small sample problem and improve classification accuracy. Method one uses DDPM to generate synthetic images for data augmentation, followed by pre-training and



**Figure 11:** Shows an example of image generation using SIG Block. We demonstrate the effect of gradually integrating three different semantic embeddings (Realistic Lineart, Zoe Depth, and Tile) during the generation process.[97]

fine-tuning the Swin Transformer model on the synthetic dataset, achieving an accuracy rate of 96.3%; method two combines the pre-trained Swin Transformer with synthetic data generated by DDPM, achieving an accuracy rate of 99.8%. This study effectively combines DDPM and transfer learning to address the issue of data scarcity. Guo et al. proposed a tomato disease detection method combining the MaxMin diffusion mechanism and lightweight technology [99], which dynamically adjusts attention weights and integrates time series networks to significantly improve disease segmentation accuracy and robustness in complex agricultural environments. The model performs exceptionally well in bacterial spot disease detection and holds broad potential for intelligent agriculture applications. Hu et al. [117] proposed an innovative method combining diffusion models and transformer architectures, termed the diffusion-transformer architecture, to enhance jujube tree disease detection accuracy. This method improves efficiency and accuracy through parallel attention mechanisms and parallel loss functions, achieving 95% accuracy in disease detection, outperforming traditional models. Zhang et al. [100] proposed a multimodal deep learning model combining image and sensor data for jujube tree disease detection in desert environments. The model precisely extracts disease features by integrating transformer and diffusion modules, overcoming detection challenges in complex environmental conditions. Experimental results show that the model outperforms existing methods in terms of accuracy, precision, and recall rate, particularly under low-light and high-dust conditions.

#### 4.6. Multimodal Fusion and Agricultural Knowledge Generation

Diffusion models have demonstrated significant potential in agricultural applications, particularly in multimodal fusion and agricultural knowledge generation. They effectively address the challenge of integrating multimodal information [25, 114, 100, 118] (such as images, sensor data, and environmental variables) in agricultural datasets, enhancing the performance of agricultural models by generating high-quality synthetic data.

Xiang et al. proposed the diffusion model-based domain adaptation framework DODA [119], aimed at addressing the adaptation issues of agricultural object detection models in dynamic environments. By combining external domain embeddings and generated synthetic images, DODA can rapidly adapt to new domains and support multimodal fusion in agricultural environments. Kong et al. [120] proposed a disease detection framework combining diffusion models, graph attention networks (ViG), and lightweight optimization strategies, leveraging multimodal data fusion to improve detection accuracy and robustness in complex environments. This method performed exceptionally well in tomato disease detection and is suitable for operation on mobile devices. Fan et al. [114] proposed the DDPM-SegFormer framework, which combines the denoising diffusion model (DDPM) and the visual transformer (Transformer) for high-precision land use and land cover semantic segmentation, optimizing remote sensing image segmentation and demonstrating broad application potential in the agricultural field. The JuDiffomer model proposed by Zhang et al. [100] combines image data and sensor data, utilizing diffusion mechanisms to improve the accuracy of jujube tree disease detection in desert environments, promoting agricultural knowledge generation, and optimizing disease prediction and control strategies. Chen et al. [118] proposed the IADL method, which combines diffusion models and language generation, using natural language descriptions to guide image generation, enhancing the dataset for

plant disease detection, improving detection accuracy, and advancing agricultural knowledge generation. Jiang et al. proposed a high-spectral image and lidar image fusion method based on DDPM [121], which improves the accuracy of disease detection and crop monitoring tasks through multimodal data fusion, particularly enhancing model adaptability in complex environments. Zhang et al. [122] proposed the BO-CNN-BiLSTM deep learning model, which combines remote sensing data and climate data. By fusing multimodal data, it improved the accuracy of crop yield prediction, promoting smart agriculture and sustainable development. Zhou et al. proposed the JuDiffomer model, which enhances the accuracy of jujube tree disease detection in desert environments through multimodal data fusion and agricultural knowledge generation [123], providing a new technical pathway for precision agriculture.

#### 4.7. Small Sample Data Image Generation

Diffusion models have demonstrated significant advantages in image generation with small sample data, particularly in the agricultural sector, where they effectively address data scarcity issues and enhance disease detection accuracy. Many studies have combined diffusion models with few-shot learning to successfully tackle data insufficiency challenges in pest detection.

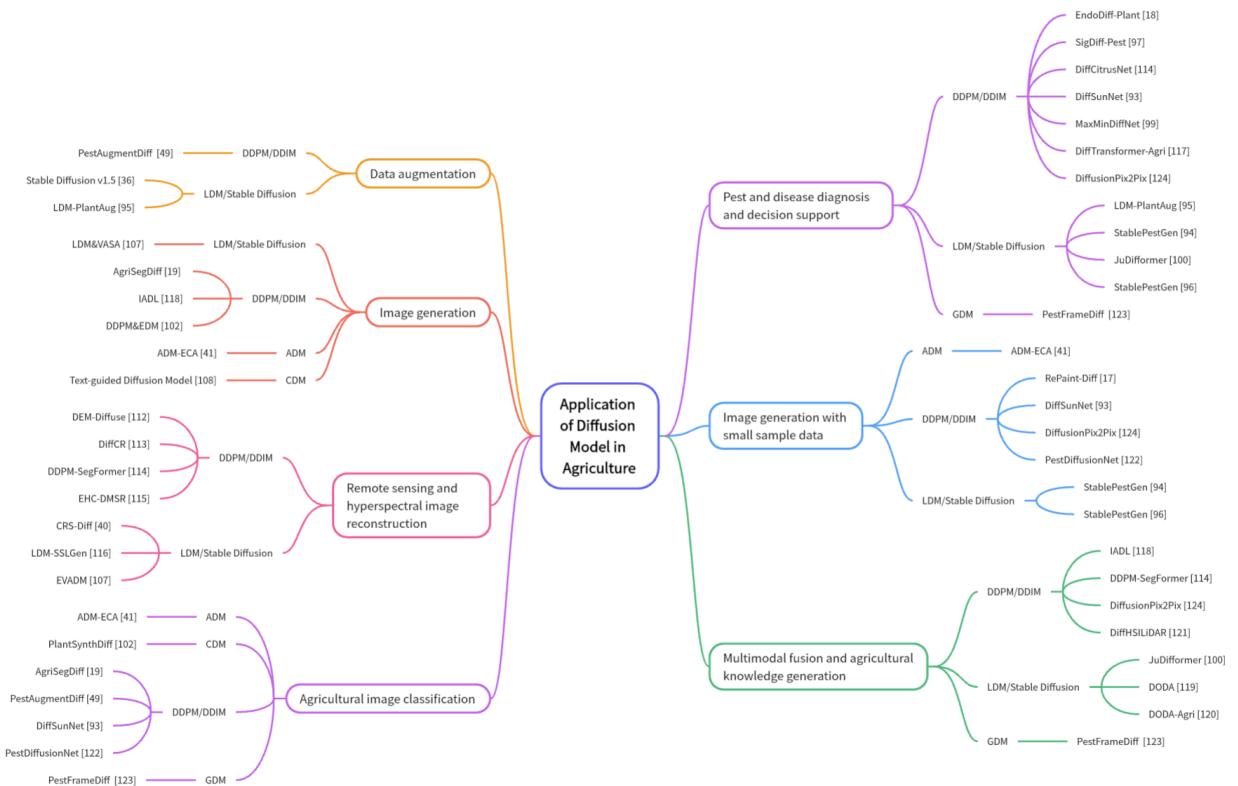
Zhou et al. proposed a diffusion model-based few-shot learning framework [93] for soybean disease detection and classification. This framework enhances detection performance by generating high-quality augmented samples and improves feature extraction of key disease-affected areas through the integration of attention mechanisms. Additionally, the study designed a new diffusion loss function to optimize the quality of generated data and model stability. This method has not only been validated in the laboratory but has also achieved good results in actual agricultural applications, particularly in soybean disease detection tasks in the Bayannur region of China, demonstrating its broad potential for practical application. Meanwhile, Wang et al. used a controllable generative model based on stable diffusion technology to generate images of emerging pests in the field [94], further addressing the issue of few-shot pest detection. By combining text prompts and coordinate information, this method can generate high-quality pest images in natural scenes, effectively alleviating the data scarcity issue. Additionally, Astuti et al. proposed a synthetic dataset generated using a convolutional visual transformer (CvT) and a stable diffusion model [96] to enhance the accuracy of potato leaf disease detection. By generating diverse synthetic datasets and training the CvT model, the study significantly improved disease detection accuracy, demonstrating the immense potential of diffusion models in addressing data scarcity issues and enhancing detection capabilities in agricultural disease detection. Furthermore, studies by Wang et al. [41] and Egusquiza et al. [124] also demonstrated the application of diffusion models in plant disease image enhancement and generation. For example, the RePaint model outperforms traditional GAN methods in generating plant disease images, particularly in terms of image quality and disease category accuracy; DiffusionPix2Pix combines disease severity labels and leaf segmentation masks to generate high-quality disease images, outperforming traditional GAN models across multiple metrics, and provides an efficient solution for enhancing agricultural disease datasets and improving the accuracy of disease detection models.

### 5. Experimental Results and Analysis

To further validate the performance advantages of diffusion models in agriculture, this paper designed and conducted comparative experiments. Using examples such as data augmentation and image generation experiments, remote sensing and hyperspectral image reconstruction experiments, pest and disease diagnosis and decision support experiments, and multimodal fusion experiments, this paper evaluated the performance of diffusion models and other existing technologies through visualization results and quantitative indicators. We designed this experiment to intuitively validate the effectiveness of diffusion models in agricultural pest image generation, providing a reference for future research.

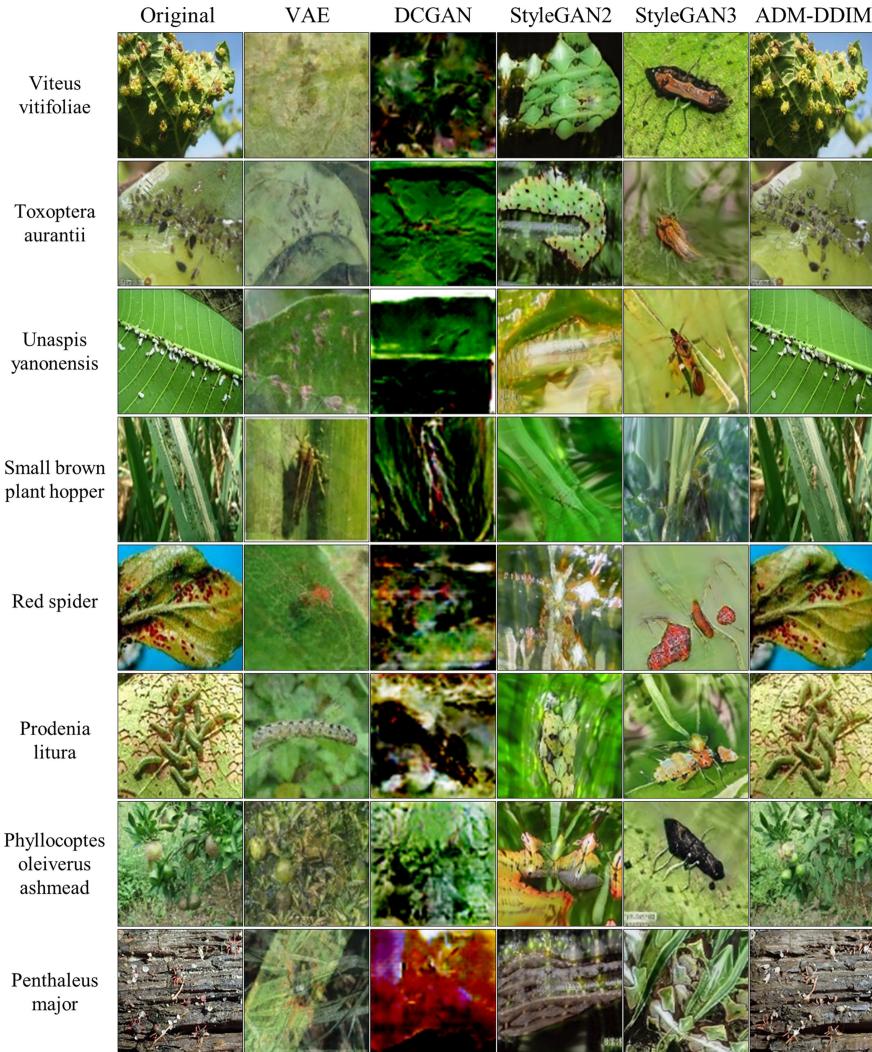
#### 5.1. Performance Evaluation and Various Indicators

SSIM (Structural Similarity Index): Another standard for evaluating image quality, with values closer to 1 indicating that the image is closer to the original; PSNR (Peak Signal-to-Noise Ratio): A metric used to measure image quality, with higher values indicating better image reconstruction results; FID (Frettschter Distance): Measures the distribution difference between generated images and real images; lower values indicate higher image quality; RFID (Relative Frettschter Distance): A relative FID score used to measure the quality difference between synthesized images and upscaled images; SRFI (Super-Resolution Relative Fidelity Index): A comprehensive evaluation metric combining structural similarity and perceptual difference; higher values indicate better super-resolution results; FLOPs (Floating-Point Operations Per Second): A metric measuring the computational complexity of the model; lower values indicate



**Figure 12:** Shows seven areas of application for diffusion models in agriculture: (1) data augmentation, (2) image generation, (3) remote sensing and hyperspectral image reconstruction, (4) pest and disease diagnosis and decision support, (5) image generation from small sample data, (6) multimodal fusion and agricultural knowledge generation, and (7) agricultural image classification. No abbreviation, first and last two initials

more efficient computational performance; Params (Number of Parameters): The number of trainable parameters in the model; fewer parameters typically indicate a more lightweight model; Precision: Measures the proportion of samples predicted as positive that are actually positive; higher values indicate more accurate positive class predictions; Recall: Measures the proportion of all actual positive samples correctly identified by the model. A higher recall rate indicates that the model is more sensitive in detecting positive samples; Accuracy: Measures the proportion of correct predictions among all predictions. A higher accuracy rate typically indicates that the model performs well overall; mAP@75 (mean average precision at 75% IOU): In object detection tasks, this evaluates the model's average precision at a 75% Intersection over Union (IoU) threshold. A higher mAP@75 indicates better performance in high-precision detection tasks; F1-Score (F1 score): The harmonic mean of precision and recall. The F1 score combines the advantages and disadvantages of precision and recall, providing a comprehensive performance metric for the model when handling imbalanced datasets; FPS (frames per second): Measures the model's processing speed, indicating the number of frames it can process per second. A higher FPS indicates stronger real-time processing capabilities; Params (number of parameters): Indicates the number of trainable parameters in the model. Fewer parameters typically mean a lighter model with lower computational requirements; FLOPs (Floating-Point Operations Per Second): A metric for measuring the computational complexity of a model. Fewer FLOPs indicate a more efficient model; Memory (Memory Usage): The amount of memory required for the model to run. Lower memory usage means the model is lighter for deployment and operation on devices.



**Figure 13:** Real samples and generated samples in the IP102 pest dataset. Images are taken from the training set. Generated samples are produced by ADM-DDIM, VAE, and GAN. Each row represents a type of pest.[49]

## 5.2. Experimental Results

### 5.2.1. Data Augmentation and Image Generation Experiments

To assess the exceptional fidelity of different generative models in generating high-quality pest images, we conducted experiments on the IP102 large-scale pest dataset using five methods, as shown in Figure 13. The experiments demonstrate that our method, ADM-DDIM, outperforms traditional generative networks such as VAE[124], DCGAN[125], StyleGAN2[126], and StyleGAN3[127] under metrics such as FID, IS, and NIQE. As shown in the figure, ADM-DDIM outperforms other models in terms of image detail, texture reproduction, and handling diverse and imbalanced datasets, making it a better choice for generating agricultural pest images.

Table 2 shows that ADM-DDIM performs exceptionally well across all metrics, particularly with an FID score of 23.66, an IS score of 7.26, and an NIQE score of 3.05, demonstrating its outstanding capability in generating high-quality agricultural pest images. In contrast, the traditional generative model VAE performs poorly in capturing image details, struggling to reproduce the complex textures of pest images, resulting in an FID score of only 264.57, an IS score of only 2.39, and an NIQE score of only 7.52. DCGAN, StyleGAN2, and StyleGAN3 can generate high-resolution, detail-rich images but perform poorly when handling datasets with diverse and imbalanced categories. For example, StyleGAN3 has a high FID score of 42.54, a low IS score of 4.70, and an intermediate NIQE score of 4.08. These

**Table 2**

Performance comparison between diffusion models and other generative models.[49]

Models	Original	VAE		GANs		Diffusion Model
		VAE	DCGAN	StyleGAN2	StyleGAN3	ADM-DDIM
FID↓	/	264.57	217.34	86.90	42.52	23.66
IS↑	1.40	2.39	2.58	2.70	4.70	7.26
NIQE↓	4.22	7.52	7.71	6.28	4.08	3.05

**Table 3**

Validation accuracy (val-acc %) of different models. The maximum value indicates the maximum accuracy improvement.[49]

	Models	ResNet	GoogLeNet	ShuffleNet	DenseNet	EfficientNet	MobileNet
Original		69.10	63.60	65.30	68.90	70.30	62.50
	VAE	68.70	64.50	64.10	67.60	70.10	63.70
	DCGAN	69.80	64.60	64.90	69.10	71.00	64.40
Generated	StyleGAN2	69.60	65.20	65.40	69.00	71.40	63.40
	StyleGAN3	69.20	65.70	65.60	69.40	70.60	63.80
	ADM-DDIM	73.80	69.30	70.70	74.00	74.80	69.10
Maximum↑		<b>4.70↑</b>	<b>5.70↑</b>	<b>5.40↑</b>	<b>5.10↑</b>	<b>4.50↑</b>	<b>6.60↑</b>

findings highlight the advantages of generating diverse and high-quality pest images and may effectively address the challenges of high intra-class variability and inter-class imbalance.

Table 3 shows the classification results of images generated by the ADM-DDIM generative model under six classification models. For example, the validation accuracy of ResNet trained on the StyleGAN2 enhanced dataset reached 69.60%, which is only 0.50% higher than the baseline model. However, with data augmentation, the validation accuracy of ResNet surged to 73.80%, which is 4.2% higher than StyleGAN2 and 4.7% higher than the baseline model. Similarly, EfficientNetV2 saw a 1.1% increase in accuracy after using StyleGAN2, but it jumped to 74.80%, which is 4.5% higher than the baseline model. These trends also extend to other models: GoogLeNet achieved an accuracy of 69.30%, which is 4.1% higher than StyleGAN2 and 3.6% higher than StyleGAN3. Notably, MobileNet's accuracy on the baseline dataset jumped from 62.50% to 69.10%, an increase of 6.6%. These results highlight the advantages of ADM-DDIM in addressing long-tail challenges and improving the classification accuracy of deep learning models. In contrast, the classification performance of datasets enhanced using DCGAN or VAE decreased. The fundamental reason lies in the presence of unnatural features or artifacts in the images generated by these models, which interfere with the training of the classifier and reduce accuracy.

### 5.2.2. Remote sensing and hyperspectral image reconstruction experiments

Table 4 shows the model performance comparison on CropSR-Test. To validate the effectiveness of the proposed variance-average spatial attention mechanism (VASA)-enhanced diffusion model (EVADM), the paper compares it with the regression model EDSR[128], the GAN model Real-ESRGAN [129], and LDM[61]. The results indicate that incorporating the VASA attention mechanism significantly improves model performance with minimal increase in computational complexity. For the  $\times 2$  SR ratio, VASR-fc1-x2 performs best, primarily due to its higher structural similarity. For the  $\times 4$  SR ratio, EVADM-x4 outperforms LDM-x4 in both FID and SRFI, with FID reduced by 5.7% and SRFI improved by 7%. Additionally, the EVADM model reduces computational complexity by over 10% while reducing parameters by nearly 50%.

Figure 14 at the top shows a visual comparison of key models in CropSR-Test. At the  $\times 2$  SR ratio, samples generated by VASR-fc1-x2 are rich in detail, vivid in color, and clear in edges; while EVADM-x2 reconstructs a more realistic crop canopy, it performs poorly in handling rare flowers. For the more challenging  $\times 4$  SR task, GAN-generated images exhibit severe distortion, while diffusion models, though unable to perfectly reconstruct ground truth, can reproduce

**Table 4**

Comparison of model performance on CropSR-Test.[107]

Model	Type	r	PSNR↑	SSIM↑	FID↓	RFID*↑	SRFI↑	FLOPs/G↓	Parms/M↓
EDSR_x2	Reg	2x	26.18	0.90	6.07	0.70	14.17	90.18	1.37
<b>VASR_fc1_x2</b>	Reg	2	<b>26.63</b>	<b>0.92</b>	<b>4.74</b>	<b>0.77</b>	<b>16.74</b>	90.87	1.45
RealESRGAN_x2	GAN	2	21.75	0.78	27.13	-0.34	1.11	294.24	37.78
EVADM_x2	Diff	2	22.69	0.83	10.01	0.50	8.29	22.97	29.14
EDSR_x4	Reg	4	<b>19.18</b>	<b>0.49</b>	51.78	0.53	2.44	130.29	1.52
RealESRGAN_x4	GAN	4	17.27	0.44	40.38	0.63	2.67	1176.61	37.77
VARDGAN_x4	GAN	4	17.43	0.45	31.73	0.71	2.95	1188.19	40.71
LDM_x4	Diff	4	16.37	0.35	31.36	0.71	2.79	40.22	113.62
EVADM_ca_x4	Diff	4	16.21	0.34	25.98	0.76	2.93	35.96	64.03
<b>EVADM_x4</b>	Diff	4	16.49	0.36	<b>25.66</b>	<b>0.77</b>	<b>2.97</b>	35.91	63.20
EDSR_x8	Reg	8	<b>16.79</b>	<b>0.20</b>	131.24	0.41	1.34	290.60	1.67
RealESRGAN_x8	GAN	8	14.84	0.13	280.12	-0.27	0.89	1453.58	37.77
<b>EVADM_x8</b>	Diff	8	14.30	0.14	<b>64.14</b>	<b>0.71</b>	<b>1.57</b>	35.92	63.21

**Table 5**

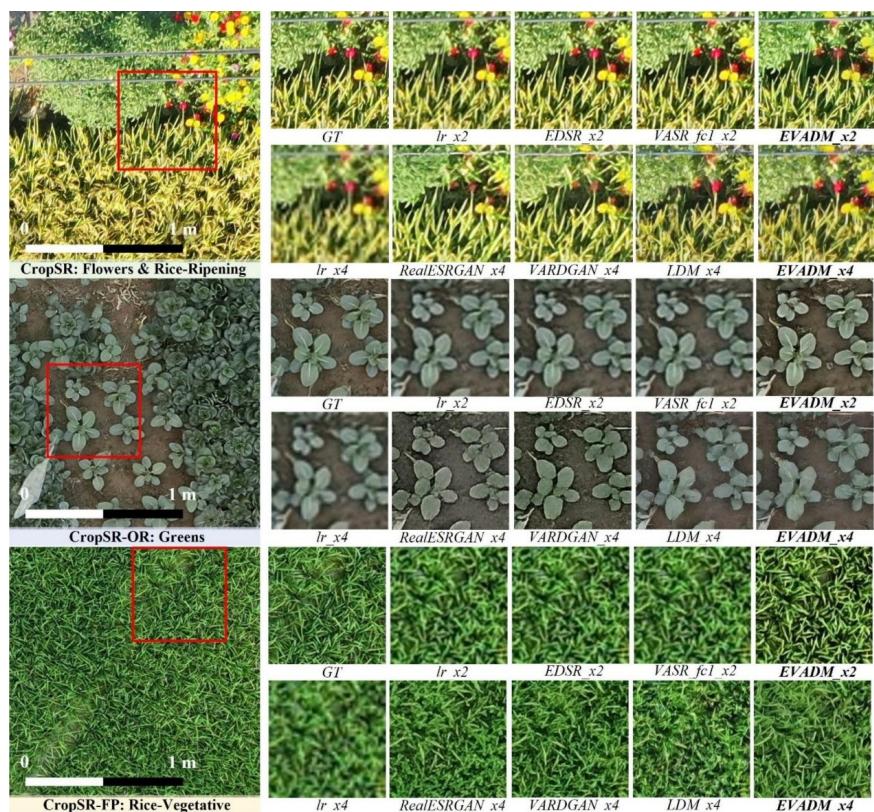
Comparison of CropSR-OR and CropSR-FP model performance.[107]

Model	r	CropSR-OR				CropSR-FP			
		SSIM↑	FID↓	RFID*↑	SRFI↑	SSIM↑	FID↓	RFID*↑	SRFI↑
EDSR_x2	2	0.55	66.60	0.13	2.54	0.50	78.71	0.14	2.46
VASR_fc1_x2	2	0.55	65.96	0.14	2.60	0.50	78.14	0.15	2.49
<b>EVADM_x2</b>	2	0.42	<b>52.30</b>	0.32	<b>3.36</b>	0.35	<b>62.50</b>	0.32	<b>3.11</b>
RealESRGAN_x4	4	0.29	124.75	0.00	1.17	0.35	141.89	0.11	1.39
VARDGAN_x4	4	0.30	116.09	0.07	1.28	0.34	126.24	0.21	1.55
LDM_x4	4	0.28	61.75	0.50	2.10	0.35	92.47	0.42	1.98
<b>EVADM_x4</b>	4	0.28	<b>52.47</b>	<b>0.58</b>	2.28	0.33	<b>85.78</b>	0.46	<b>2.06</b>
EVADM_x8	8	0.15	101.72	0.57	1.45	0.21	151.09	0.39	1.33

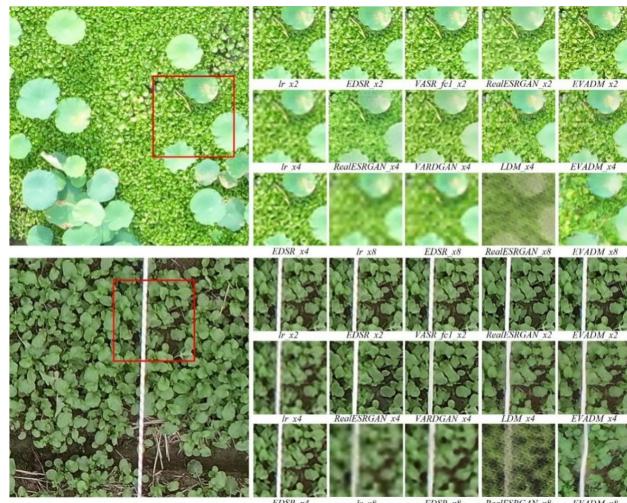
similar patterns. When directly conducting  $\times 8$  ratio enhancement SR experiments, due to severe information loss, the model can only infer similar features. However, compared to regression and GAN methods, the images generated by EVADM-x8 (Figure 15) have clear edges and discernible details, with FID reduced by 67.1, validating the feasibility of diffusion models in high-ratio image SR.

The test results on the CropSR-OR/FP dataset are shown in Table 5. Although the regression-based EDSR-x2 and VASR-fc1-x2 achieved high SSIM values, their SR results were rather blurry, approaching interpolation effects, as shown in the middle (CropSR-OR) and bottom (CropSR-FP) of Figure 14. For the  $\times 2$  and  $\times 4$  scales, the proposed EVADM model performed best on both FID and SRFI. Compared to the baseline models, EVADM-x2 reduced FID by an average of 14.6, and EVADM-x4 by 8.0; SRFI improved by 27.1% and 6.3%, respectively. GAN-based models scored close to zero on RFID, indicating they cannot generate satisfactory SR results.

Observing the VARDGAN-x4 image of the second sample, the leaf detail features are ignored, while EVADM-x4 vividly depicts the leaf veins in Figure 15. For the  $\times 4$  SR of the third row of dense crop canopy, the GAN model cannot reconstruct narrow leaves. By comparing the last two images generated by the  $\times 4$  diffusion model, it can be seen that the VASA attention mechanism significantly improves the SR model's ability to reconstruct fine details. Compared to the performance of regression models on the simulated CropSR-Test dataset, the decline in numerical and visual quality on real SR datasets indicates their limited generalization ability, while diffusion models exhibit greater robustness to domain shifts.



**Figure 14:** Visual comparison of SR results on the CropSR-Test, CropSR-OR, and CropSR-FP datasets. The left side provides an overview of different farmland scenes, while the model names are displayed below the small blocks on the right side.[107]



**Figure 15:** Visual comparison of SR results on the CropSR-Test dataset. The left side provides an overview of different farmland scenes, while the model names are displayed below the small blocks on the right side.[107]

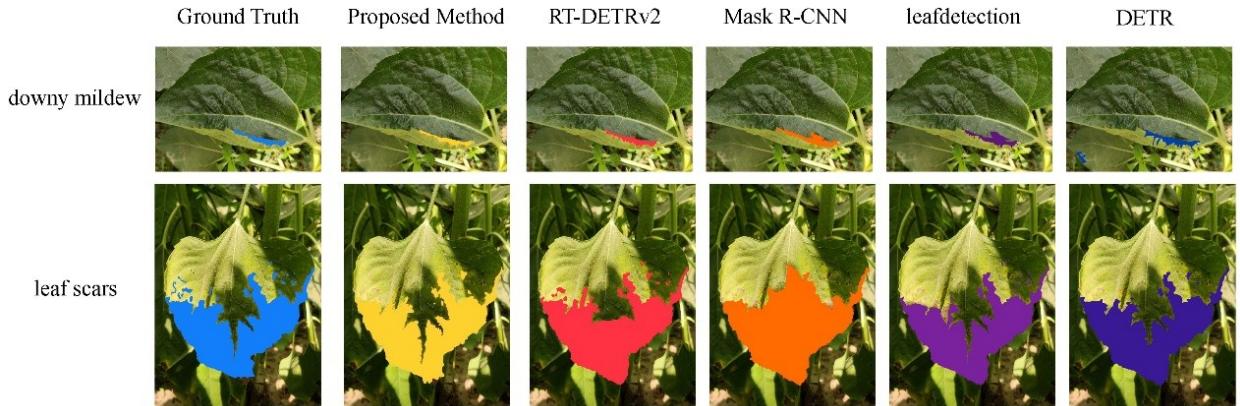
### **5.2.3. Pest and disease diagnosis and auxiliary decision-making experiments**

This experiment evaluated the performance of different models in sunflower disease detection using metrics such as precision, recall, accuracy, F1 score, and mean average precision (mAP@75). The results showed that the models

**Table 6**

Experimental results of disease detection model.[93]

Model	Precision	Recall	Accuracy	mAP@75	F1-Score
RT-DETRv2	0.83	0.80	0.82	0.81	0.81
Mask R-CNN	0.87	0.84	0.86	0.85	0.85
DETR	0.89	0.86	0.87	0.87	0.87
LeafDetection	0.92	0.89	0.91	0.90	0.90
FSDDN	0.94	0.92	0.93	0.92	0.93

**Figure 16:** Visualization analysis of experimental results.[93]

achieved significant improvements through architectural optimization and task adaptation, particularly in complex background and limited sample scenarios. Table 6 shows that RT-DETRv2 performed the worst, as it focuses on global features and lacks the ability to extract fine details; Mask R-CNN achieved good results through instance segmentation, but still has room for improvement in small spot segmentation; the DETR model performed well, but faced challenges in background noise and small spot detection; The LeafDetection[130] model, specifically designed for sunflower disease detection, demonstrates a clear advantage in fine-grained feature extraction, outperforming other models in all metrics with an accuracy of 0.94, recall of 0.92, precision of 0.93, F1 score of 0.93, and mAP@75 is 0.92, proving the effective combination of deep learning technology and customized disease feature modeling. According to the article, the method proposed by the authors can be named "Few-Shot Diffusion Detection Network (FSDDN)". This name highlights the application of few-shot learning and diffusion models for high-quality disease detection in small sample situations.

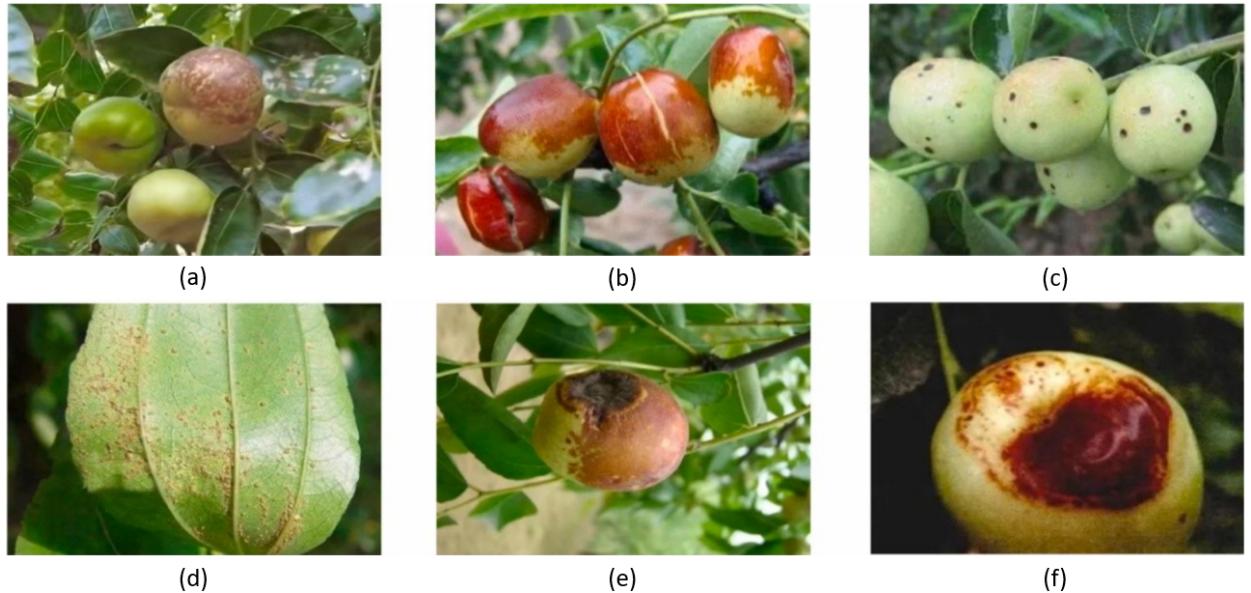
In addition to experiments on the original dataset, we also tested an independent dataset to validate the generalization performance and robustness of the proposed method. The dataset contains four categories of images: downy mildew, fresh leaves, gray mold, and leaf scars, with 120, 134, 72, and 140 images in each category, respectively. The diversity of the dataset is reflected in the significant differences in disease types and leaf conditions, particularly for diseases like downy mildew and gray mold, which have complex features and unclear boundaries, making them susceptible to background interference. The annotation process aligns with the original dataset, with expert-verified precise bounding boxes ensuring effective capture of disease features during training. To mitigate class imbalance, data balancing techniques and augmentation operations (such as rotation, cropping, and color jittering) were employed during training, enhancing model robustness and expanding the dataset. As shown in Table 7, these operations significantly improved model performance.

Figure 16 shows the visualization results of different models, including RT-DETRv2, Mask R-CNN, DETR, LeafDetection, and the proposed model. RT-DETRv2 can only capture global features and performs poorly in handling details; Mask R-CNN performs well in instance segmentation but is insufficient for small lesion segmentation; DETR has an advantage in utilizing global information but faces challenges in detecting small lesion regions; LeafDetection performs well in disease detection and excels in detail extraction. In contrast, the proposed model performs best in

**Table 7**

Experimental results of disease detection models on other datasets.[93]

Model	Precision	Recall	Accuracy	mAP@75	F1-Score
RT-DETRv2	0.84	0.80	0.82	0.81	0.82
Mask R-CNN	0.86	0.82	0.84	0.83	0.84
DETR	0.87	0.85	0.86	0.86	0.86
LeafDetection	0.90	0.89	0.90	0.90	0.89
FSDDN	0.92	0.90	0.91	0.91	0.91

**Figure 17:** Samples of different jujube tree diseases: (a) jujube wilt disease, (b) fruit cracking disease, (c) black tumor disease, (d) rust disease, (e) ring spot disease, (f) anthracnose disease.[100]

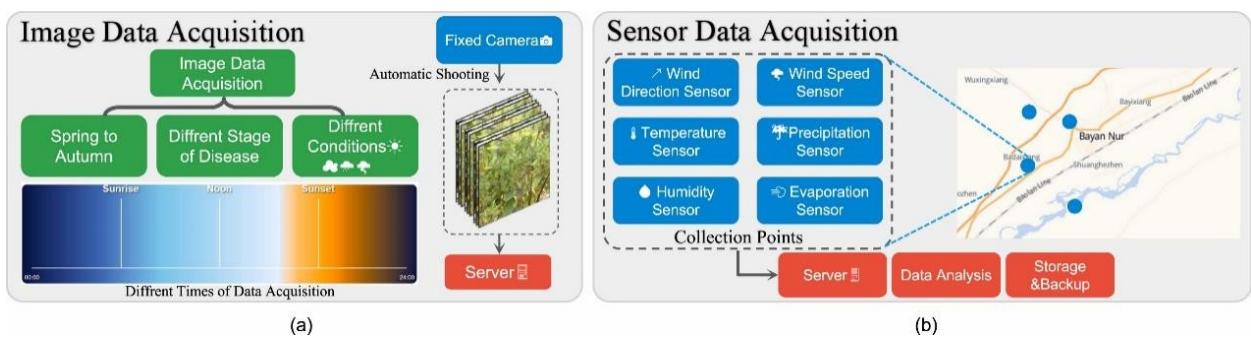
detecting all disease categories, particularly demonstrating stronger robustness in the segmentation and identification of lesions caused by downy mildew and leaf scars.

#### 5.2.4. Multimodal fusion experiment

This experiment proposes a deep learning model based on multimodal data fusion for detecting jujube tree diseases in desert environments. Considering the complexity of light and environmental conditions in desert regions, existing methods have limitations in feature extraction and accuracy. By integrating image and sensor data, this study designed a feature extraction mechanism combining Transformer and Diffusion modules to achieve precise capture of disease features.

Data collection was primarily conducted in the Bayannur Desert region of northern China, where environmental factors such as light, temperature, and humidity significantly influence jujube tree diseases. To capture disease information under various environmental conditions, data collection was conducted from spring to autumn, covering different growth stages of jujube trees and disease development phases. Images were collected daily at dawn, noon, and dusk, as shown in Figures 17 and Figures 18-(a), to ensure the model could detect diseases under different lighting conditions, thereby enhancing the model's generalization ability and avoiding overfitting. Sensors were also used to collect data around the clock, recording environmental information such as temperature, humidity, and wind speed. These data helped the model understand the relationship between diseases and environmental factors, enhancing the model's robustness.

Table 8 compares YOLOv9 [131], Retinanet [132], Efficientdet [133], YOLOv10 [134], and DETR [135] in disease detection tasks, evaluating their convergence speed throughout the training process and final detection results. Table 9



**Figure 18:** Data collection overview. (a) Image data collection: Captured at different times, seasons, disease stages, and conditions, and stored on a central server. (b) Sensor data collection: Environmental data (e.g., wind, temperature, humidity) collected from multiple sensors, transmitted for analysis and storage.[100]

Table 8

Comparison with different mainstream models[100]

Model	Precision	Recall	Accuracy	mAP	F1 Score	FPS	Params (M)	FLOPs (G)	Memory (GB)
YOLOv9	0.83	0.79	0.81	0.82	0.81	35	57.3	189	10
Retinanet	0.85	0.81	0.83	0.84	0.83	25	52	198	8
Efficientdet	0.87	0.83	0.85	0.86	0.85	30	77	410	12
YOLOv10	0.89	0.85	0.86	0.87	0.87	40	29.5	160	9
DETR	0.91	0.87	0.88	0.89	0.89	22	60	253	11
JuDiformer	0.93	0.89	0.90	0.91	0.91	28	61.32	207	10

**Table 9**

Comparison with other state-of-the-art results.[100]

Model	Precision	Recall	Accuracy	mAP	F1 Score
Thai et al [136]	0.76	0.71	0.73	0.72	0.73
Zeng et al [128]	0.78	0.73	0.75	0.74	0.75
Pang et al [137]	0.79	0.74	0.76	0.75	0.77
Li et al [138]	0.81	0.76	0.78	0.77	0.78
Sebastian et al [139]	0.82	0.78	0.79	0.79	0.80
Wang et al [140]	0.84	0.79	0.81	0.81	0.81
Zhou et al [141]	0.85	0.81	0.82	0.82	0.83
Sharma et al [142]	0.87	0.82	0.84	0.84	0.85
Wang et al [143]	0.88	0.84	0.85	0.86	0.86
Karthik et al [144]	0.90	0.86	0.87	0.88	0.88
Zhang and Lu et al [145]	0.91	0.87	0.88	0.89	0.89
JuDifformer	0.93	0.89	0.90	0.91	0.91

compares the performance of the proposed disease detection model with other state-of-the-art methods in disease detection tasks.

Table 10 shows that when using only image data, the model's metrics are as follows: Precision 0.81, Recall 0.77, Accuracy 0.79, mAP 0.78, and F1-score 0.79; when using sensor data alone, these metrics improved to 0.85, 0.82, 0.83, 0.83, and 0.83, respectively, indicating that environmental information in sensor data is closely related to disease progression. When using image and sensor data fusion, all metrics significantly improved: Precision reached 0.93, Recall was 0.89, Accuracy was 0.90, mAP was 0.91, and F1-score was 0.91, demonstrating the effectiveness of multimodal data fusion in enhancing disease detection accuracy and robustness.

### **5.3. Key findings and conclusions from the experiment**

In this section, the key findings from four experiments are summarized, demonstrating the powerful application potential of diffusion models in the agricultural field, particularly in pest and disease detection and data augmentation.

**Table 10**

Comparison with different multimodal datasets.[100]

Data	Precision	Recall	Accuracy	mAP	F1 Score
Image	0.81	0.77	0.79	0.78	0.79
Sensor	0.85	0.82	0.83	0.83	0.83
Both	0.93	0.89	0.90	0.91	0.91

In the data augmentation and image generation experiments, data augmentation using diffusion models (ADM-DDIM) significantly improved the quality of the generated images. Compared to traditional generative models such as VAE, DCGAN, StyleGAN2, and StyleGAN3, ADM-DDIM demonstrated superior performance across metrics including FID, IS, and NIQE. The generated images not only exhibit greater precision in detail and texture reproduction but also effectively address data long-tail issues, particularly achieving significant progress in the identification of rare categories. This finding indicates that diffusion models have great application potential in alleviating agricultural data scarcity and class imbalance issues.

In remote sensing and hyperspectral image reconstruction experiments, the diffusion model enhanced by the variance-average spatial attention mechanism (VASA) (EVADM) demonstrated excellent performance in image super-resolution tasks. Compared to regression models (such as EDSR) and GAN models (such as RealESRGAN and LDM), EVADM achieved significant improvements in FID and SRFI metrics, demonstrating its efficiency and stability in agricultural image super-resolution tasks. Additionally, EVADM performed well in reducing computational complexity and model parameters, enhancing its feasibility for practical agricultural applications.

Experiments show that diffusion model-based generative enhancement methods significantly improve the accuracy and robustness of agricultural pest and disease detection, particularly in tasks involving long-tail categories and limited samples. By generating high-quality synthetic images, diffusion models enhance the model's detection capabilities, particularly in sunflower disease detection and date palm disease detection in desert environments. The multimodal fusion of images and sensor data further improves the model's adaptability and accuracy in complex environments. These results indicate that diffusion models hold great potential in data augmentation, image reconstruction, and multimodal fusion, providing efficient and precise technical support for smart agriculture and agricultural pest and disease detection, thereby driving technological progress in the agricultural sector.

## 6. Challenges and Countermeasures in Diffusion Models for Agricultural Applications

Diffusion models hold great promise for advancing agricultural practices, particularly in areas like pest and disease detection, crop monitoring, and precision farming. However, several key challenges need to be addressed before these models can reach their full potential in the agricultural domain.

### 6.1. Challenges

- **Computational Efficiency:** One of the most significant hurdles is the high computational cost associated with generating high-resolution images, especially when working with large-scale datasets. Due to the iterative nature of diffusion models, which requires multiple network evaluations during the generation process, inference speeds are notably slow. This issue is particularly critical in real-time agricultural applications, where timely results are essential. To overcome this, future research should prioritize the development of faster sampling algorithms and explore distributed computing and parallel processing to enhance the real-time performance of diffusion models.
- **Data Scarcity and Imbalance:** Agricultural datasets often suffer from data scarcity, especially when dealing with rare pests or diseases. During the early stages of an outbreak, the lack of sufficient data hinders the ability of diffusion models to generate accurate training samples. Furthermore, datasets are frequently imbalanced, with overrepresentation of certain diseases or crops, leading to biased model training. Addressing this issue will require advanced data augmentation strategies and the generation of high-quality synthetic data, especially for rare agricultural conditions, to improve model robustness across diverse scenarios.
- **Generalization Across Diverse Agricultural Environments:** Another significant challenge lies in the variability of environmental conditions, such as fluctuating climates and soil types. Diffusion models often struggle to generalize effectively when applied to new or underrepresented agricultural settings. This can result in decreased

model performance when transitioning from controlled environments, such as research labs, to real-world agricultural conditions. To enhance generalization, techniques such as domain adaptation, transfer learning, and multi-modal data fusion should be explored, enabling models to adapt to different agricultural conditions.

- **Dataset Quality:** The effectiveness of diffusion models heavily relies on the quality and diversity of the datasets used during training. Low-quality data, such as noisy images or inconsistent labeling, significantly impacts model performance. Ensuring high-quality, well-labeled datasets and incorporating data from a broad range of agricultural environments will be crucial for improving the model's learning process and its ability to generate accurate results.

## 6.2. Future Research Directions

Looking ahead, several avenues for research and innovation could significantly enhance the application of diffusion models in agriculture.

- **Improving Computational Efficiency:** As previously mentioned, the computational cost of diffusion models remains a critical bottleneck. Future work should focus on model compression techniques, faster sampling algorithms, and leveraging high-performance computing to reduce inference time without compromising the quality of generated results.
- **Synthetic Data Generation and Augmentation:** To address the issues of data scarcity and imbalance, future research should explore the generation of synthetic agricultural data through advanced data augmentation methods. This could help to create labeled datasets for rare agricultural conditions and improve model performance across a wide variety of crops, pests, and diseases.
- **Cross-Domain Generalization:** Improving the generalization capabilities of diffusion models across various agricultural environments is essential. Techniques like domain adaptation and transfer learning can be employed to ensure models are adaptable to new settings. Additionally, integrating multi-modal data sources will provide a more comprehensive understanding of agricultural environments, improving the robustness and accuracy of models.
- **Integration with Other Generative Models:** Combining diffusion models with other generative models such as GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders) could lead to more flexible and efficient generative systems. Research should explore ways to integrate these models, potentially improving control over image generation processes. Furthermore, exploring semantic space could allow for more targeted generation of agricultural data, such as images of specific crop diseases or growth stages.

By addressing these challenges and pursuing these future research directions, diffusion models have the potential to significantly advance agricultural practices, especially in areas like pest and disease detection, crop monitoring, and precision agriculture.

## 7. Summary

Diffusion models have demonstrated their powerful potential in agricultural applications, particularly in crop health monitoring, pest and disease detection, and remote sensing image processing. By generating high-quality synthetic data, diffusion models can effectively address issues such as data scarcity and sample imbalance in agriculture. In fine-grained disease identification and crop growth prediction, diffusion models provide agricultural experts with clearer images of crop health status, thereby aiding in more precise decision-making. Compared to traditional methods, diffusion models generate images through gradual denoising, significantly improving image detail and quality. Additionally, diffusion models have distinct advantages in data augmentation and image generation, particularly in generating high-quality synthetic images to supplement scarce samples, expand training datasets, and thereby enhance classifier performance. In tasks such as pest and disease detection and crop growth prediction, the images generated by diffusion models are not only of high quality but can also handle complex situations in different environments, enhancing the model's adaptability and accuracy. The conditional generation capability of diffusion models allows them to generate specific images based on different crop types, pest and disease types, or environmental conditions, thereby providing customized solutions to meet the diverse needs of precision agriculture. Additionally, diffusion

models demonstrate strong capabilities in multimodal data fusion, integrating various types of information such as images, sensor data, and climate data to provide comprehensive support for agricultural decision-making. This data fusion capability gives diffusion models significant application potential in agricultural pest and disease monitoring, environmental change analysis, and precision agriculture decision-making.

## Acknowledgment

This project is jointly supported by the Shanghai Agricultural Technology Innovation Project (2024-02-08-00-12-F00032), the Science and Technology Program of Qinghai Province (2024-NK-141S-4), the High-Quality Development Special Project of the Ministry of Industry and Information Technology (TC240A9ED-56), the Academician Workstation Program of Yunnan Province (202405AF140013), and the Shanghai Municipal Commission of Science and Technology (24TS1415900).

## References

- [1] R. Valle, J. M. Buenaposada, L. Baumela, Reducing head pose estimation data set bias with synthetic data, *IEEE Access* (2025).
- [2] P. Zhang, D. Wang, C. Wu, J. Yang, L. Kang, Z. Bai, Y. Li, Q. Shen, Hyperdiff: Masked diffusion model with high-efficient transformer for hyperspectral image cross-scene classification, in: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5.
- [3] P. Zhang, D. Wang, C. Wu, J. Yang, L. Kang, Z. Bai, Y. Li, Q. Shen, Hyperdiff: Masked diffusion model with high-efficient transformer for hyperspectral image cross-scene classification, in: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5.
- [4] Y. Huang, K. Kastner, K. Audhkhasi, B. Ramabhadran, A. Rosenberg, Audio diffusion with large language models, in: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5.
- [5] Y. Li, B. Leng, Adaptive contribution modulation for multi-modal manipulation media detection and grounding, in: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5.
- [6] D. Liu, Z. Wang, L. Guo, A plug-and-play diffusion-styled conversion model for domain discrepancies in medical image segmentation, in: *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2025, pp. 1–5.
- [7] H. Li, H. Jiang, A. Luo, P. Tan, H. Fan, B. Zeng, S. Liu, Dmhomo: Learning homography with diffusion models, *ACM Transactions on Graphics* 43 (3) (2024) 1–16.
- [8] L. Dong, Diffusion probabilistic models for resource allocation in multi-carrier noma systems, in: *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2025, pp. 1–6.
- [9] C. Reinders, R. Berdan, B. Besbinar, J. Otsuka, D. Iso, Raw-diffusion: Rgb-guided diffusion models for high-fidelity raw image generation, in: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2025, pp. 8431–8443.
- [10] M. Poleski, J. Tabor, P. Spurek, Geoguide: Geometric guidance of diffusion models, in: *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2025, pp. 297–305.
- [11] Y. Shi, A. Abulizi, H. Wang, K. Feng, N. Abudukelimu, Y. Su, H. Abudukelimu, Diffusion models for medical image computing: A survey, *Tsinghua Science and Technology* 30 (1) (2024) 357–383.
- [12] T. Amit, T. Shaharbany, E. Nachmani, L. Wolf, Segdiff: Image segmentation with diffusion probabilistic models, *arXiv preprint arXiv:2112.00390* (2021).
- [13] A. Rahman, J. M. J. Valanarasu, I. Hacihaliloglu, V. M. Patel, Ambiguous medical image segmentation using diffusion models, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 11536–11546.
- [14] J. Wolleb, R. Sandkühler, F. Bieder, P. Valmaggia, P. C. Cattin, Diffusion models for implicit image segmentation ensembles, in: *International conference on medical imaging with deep learning*, PMLR, 2022, pp. 1336–1348.
- [15] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, *arXiv preprint arXiv:1511.06434* (2015).
- [16] Y. Wang, Z. Zhou, X. Tan, Y. Pan, J. Yuan, Z. Qiu, C. Liu, Unveiling the potential of progressive training diffusion model for defect image generation and recognition in industrial processes, *Neurocomputing* 592 (2024) 127837.
- [17] A. Muhammad, Z. Salman, K. Lee, D. Han, Harnessing the power of diffusion models for plant disease image augmentation, *Frontiers in Plant Science* 14 (2023) 1280496.
- [18] J. Yin, W. Li, J. Shen, C. Zhou, S. Li, J. Suo, J. Yang, R. Jia, C. Lv, A diffusion-based detection model for accurate soybean disease identification in smart agricultural environments, *Plants* 14 (5) (2025) 675.
- [19] A. Heschl, M. Murillo, K. Najafian, F. Maleki, Synthset: generative diffusion model for semantic segmentation in precision agriculture, in: *European Conference on Computer Vision*, Springer, 2024, pp. 267–283.
- [20] H. Shi, L. Tian, J. Tao, Y. Li, L. Wang, X. Liu, A spatial-temporal deep probabilistic diffusion model for reliable hail nowcasting with radar echo extrapolation, in: *Companion Proceedings of the ACM on Web Conference 2025*, 2025, pp. 2367–2376.
- [21] F. Ling, Z. Lu, J.-J. Luo, L. Bai, S. K. Behera, D. Jin, B. Pan, H. Jiang, T. Yamagata, Diffusion model-based probabilistic downscaling for 180-year east asian climate reconstruction, *npj Climate and Atmospheric Science* 7 (1) (2024) 131.
- [22] X. Ling, C. Li, F. Qin, P. Yang, Y. Huang, Srndiff: Short-term rainfall nowcasting with condition diffusion model, *arXiv preprint arXiv:2402.13737* (2024).

- [23] M. El Sakka, M. Ivanovici, L. Chaari, J. Mothe, A review of cnn applications in smart agriculture using multimodal data, *Sensors* 25 (2) (2025) 472.
- [24] A. Kamilaris, F. X. Prenafeta-Boldú, A review of the use of convolutional neural networks in agriculture, *The Journal of Agricultural Science* 156 (3) (2018) 312–322.
- [25] H. Huang, W. He, H. Zhang, Y. Xia, L. Zhang, Stfdiff: Remote sensing image spatiotemporal fusion with diffusion models, *Information Fusion* 111 (2024) 102505.
- [26] X. Ding, J. Qu, W. Dong, T. Zhang, N. Li, Y. Yang, Graph representation learning-guided diffusion model for hyperspectral change detection, *IEEE Geoscience and Remote Sensing Letters* 21 (2024) 1–5.
- [27] X. Ding, J. Qu, W. Dong, T. Zhang, N. Li, Y. Yang, Graph representation learning-guided diffusion model for hyperspectral change detection, *IEEE Geoscience and Remote Sensing Letters* 21 (2024) 1–5.
- [28] H. Zeng, J. Cao, K. Zhang, Y. Chen, H. Luong, W. Philips, Unmixing diffusion for self-supervised hyperspectral image denoising, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 27820–27830.
- [29] L. Pang, X. Rui, L. Cui, H. Wang, D. Meng, X. Cao, Hir-diff: Unsupervised hyperspectral image restoration via improved diffusion models, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3005–3014.
- [30] P. Awate, A. D. Nagne, Case studies on generative adversarial networks in precision farming: Gan for precision agriculture, in: *Generative Adversarial Networks for Remote Sensing*, IGI Global Scientific Publishing, 2025, pp. 291–320.
- [31] M. Abdulsalam, U. Zahidi, B. Hurst, S. Pearson, G. Cielniak, J. Brown, Unsupervised tomato split anomaly detection using hyperspectral imaging and variational autoencoders, in: *European Conference on Computer Vision*, Springer, 2024, pp. 101–114.
- [32] F. O. Isinkaye, M. O. Olusanya, A. A. Akinyelu, A multi-class hybrid variational autoencoder and vision transformer model for enhanced plant disease identification, *Intelligent Systems with Applications* 26 (2025) 200490.
- [33] J. Xu, Y. Lu, B. Deng, Openweedgui: an open-source graphical tool for weed imaging and yolo-based weed detection, *Electronics* 13 (9) (2024) 1699.
- [34] H. Moreno, A. Gómez, S. Altares-López, A. Ribeiro, D. Andújar, Analysis of stable diffusion-derived fake weeds performance for training convolutional neural networks, *Computers and Electronics in Agriculture* 214 (2023) 108324.
- [35] F.-A. Croitoru, V. Hondru, R. T. Ionescu, M. Shah, Diffusion models in vision: A survey, *IEEE transactions on pattern analysis and machine intelligence* 45 (9) (2023) 10850–10869.
- [36] K. Zhao, M. Nguyen, W. Yan, Evaluating accuracy and efficiency of fruit image generation using generative ai diffusion models for agricultural robotics, in: *2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, IEEE, 2024, pp. 1–6.
- [37] Z. Kong, W. Ping, J. Huang, K. Zhao, B. Catanzaro, Diffwave: A versatile diffusion model for audio synthesis, *arXiv preprint arXiv:2009.09761* (2020).
- [38] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, D. Merhof, Diffusion models in medical imaging: A comprehensive survey, *Medical image analysis* 88 (2023) 102846.
- [39] S. Teslenok, P. Dmitriev, K. Teslenok, N. Letkina, I. Fomin, Spatial analysis of innovations diffusion in the agricultural sector, in: *IV International Scientific and Practical Conference'Anthropogenic Transformation of Geospace: Nature, Economy, Society'(ATG 2019)*, Atlantis Press, 2020, pp. 281–285.
- [40] D. Tang, X. Cao, X. Hou, Z. Jiang, J. Liu, D. Meng, Crs-diff: Controllable remote sensing image generation with diffusion model, *IEEE Transactions on Geoscience and Remote Sensing* (2024).
- [41] R. Wang, X. Zhang, Q. Yang, L. Lei, J. Liang, L. Yang, Enhancing panax notoginseng leaf disease classification with inception-ssnet and image generation via improved diffusion model, *Agronomy* 14 (9) (2024) 1982.
- [42] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, *Advances in neural information processing systems* 33 (2020) 6840–6851.
- [43] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, S. Ganguli, Deep unsupervised learning using nonequilibrium thermodynamics, in: *International conference on machine learning*, pmlr, 2015, pp. 2256–2265.
- [44] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241.
- [45] J. Song, C. Meng, S. Ermon, Denoising diffusion implicit models, *arXiv preprint arXiv:2010.02502* (2020).
- [46] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, B. Poole, Score-based generative modeling through stochastic differential equations, *arXiv preprint arXiv:2011.13456* (2020).
- [47] L. Zhang, A. Rao, M. Agrawala, Adding conditional control to text-to-image diffusion models, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 3836–3847.
- [48] J. Song, C. Meng, S. Ermon, Denoising diffusion implicit models, *arXiv preprint arXiv:2010.02502* (2020).
- [49] M. Du, F. Wang, Y. Wang, K. Li, W. Hou, L. Liu, Y. He, Y. Wang, Improving long-tailed pest classification using diffusion model-based data augmentation, *Computers and Electronics in Agriculture* 234 (2025) 110244.
- [50] D. P. Kingma, M. Welling, et al., An introduction to variational autoencoders, *Foundations and Trends® in Machine Learning* 12 (4) (2019) 307–392.
- [51] A. Vahdat, J. Kautz, Nvae: A deep hierarchical variational autoencoder, *Advances in neural information processing systems* 33 (2020) 19667–19679.
- [52] Z.-S. Liu, W.-C. Siu, L.-W. Wang, Variational autoencoder for reference based image super-resolution, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 516–525.
- [53] Z.-S. Liu, W.-C. Siu, Y.-L. Chan, Photo-realistic image super-resolution via variational autoencoders, *IEEE Transactions on Circuits and Systems for Video Technology* 31 (4) (2020) 1351–1365.
- [54] K. Yi, Y. Guo, Y. Fan, J. Hamann, Y. G. Wang, Cosmo vae: Variational autoencoder for cmb image inpainting, in: *2020 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2020, pp. 1–7.

- [55] C.-T. Tu, Y.-F. Chen, Facial image inpainting with variational autoencoder, in: 2019 2nd international conference of intelligent robotic and control engineering (IRCE), IEEE Computer Society, 2019, pp. 119–122.
- [56] Y. Song, S. Ermon, Generative modeling by estimating gradients of the data distribution, *Advances in neural information processing systems* 32 (2019).
- [57] Y. Song, S. Ermon, Improved techniques for training score-based generative models, *Advances in neural information processing systems* 33 (2020) 12438–12448.
- [58] C. Luo, Understanding diffusion models: A unified perspective, *arXiv preprint arXiv:2208.11970* (2022).
- [59] P. Vincent, A connection between score matching and denoising autoencoders, *Neural computation* 23 (7) (2011) 1661–1674.
- [60] Y. Song, S. Garg, J. Shi, S. Ermon, Sliced score matching: A scalable approach to density and score estimation, in: *Uncertainty in artificial intelligence*, PMLR, 2020, pp. 574–584.
- [61] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, High-resolution image synthesis with latent diffusion models, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10684–10695.
- [62] J. Ho, T. Salimans, Classifier-free diffusion guidance, *arXiv preprint arXiv:2207.12598* (2022).
- [63] P. Dhariwal, A. Nichol, Diffusion models beat gans on image synthesis, *Advances in neural information processing systems* 34 (2021) 8780–8794.
- [64] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, D. J. Fleet, Video diffusion models, *Advances in neural information processing systems* 35 (2022) 8633–8646.
- [65] C. Lu, Y. Zhou, F. Bao, J. Chen, C. Li, J. Zhu, Dpm-solver++: Fast solver for guided sampling of diffusion probabilistic models, *Machine Intelligence Research* (2025) 1–22.
- [66] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, M. Chen, Glide: Towards photorealistic image generation and editing with text-guided diffusion models, *arXiv preprint arXiv:2112.10741* (2021).
- [67] C. Meng, R. Rombach, R. Gao, D. Kingma, S. Ermon, J. Ho, T. Salimans, On distillation of guided diffusion models, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 14297–14306.
- [68] M. Hu, Y. Wang, T.-J. Cham, J. Yang, P. N. Suganthan, Global context with discrete diffusion in vector quantised modelling for image generation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11502–11511.
- [69] J. Wolleb, F. Bieder, R. Sandkühler, P. C. Cattin, Diffusion models for medical anomaly detection, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2022, pp. 35–45.
- [70] K. Packhäuser, L. Folle, F. Thamm, A. Maier, Generation of anonymous chest radiographs using latent diffusion models for training thoracic abnormality classification systems, in: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2023, pp. 1–5.
- [71] S. Chen, P. Sun, Y. Song, P. Luo, Diffusiondet: Diffusion model for object detection, in: *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 19830–19843.
- [72] D. Baranchuk, I. Rubachev, A. Voynov, V. Khrulkov, A. Babenko, Label-efficient semantic segmentation with diffusion models, *arXiv preprint arXiv:2112.03126* (2021).
- [73] V. T. Hu, D. W. Zhang, Y. M. Asano, G. J. Burghouts, C. G. Snoek, Self-guided diffusion models, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18413–18422.
- [74] C.-H. Chao, W.-F. Sun, B.-W. Cheng, C.-Y. Lee, Quasi-conservative score-based generative models (2022).
- [75] H. Chung, B. Sim, J. C. Ye, Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12413–12422.
- [76] J. Choi, S. Kim, Y. Jeong, Y. Gwon, S. Yoon, Ilvr: Conditioning method for denoising diffusion probabilistic models, *arXiv preprint arXiv:2108.02938* (2021).
- [77] Y. Tashiro, J. Song, Y. Song, S. Ermon, Csd: Conditional score-based diffusion models for probabilistic time series imputation, *Advances in neural information processing systems* 34 (2021) 24804–24816.
- [78] A. Vahdat, K. Kreis, J. Kautz, Score-based generative modeling in latent space, *Advances in neural information processing systems* 34 (2021) 11287–11302.
- [79] K. Pandey, A. Mukherjee, P. Rai, A. Kumar, Vaes meet diffusion models: Efficient and high-fidelity generation, in: *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, Vol. 14, 2021.
- [80] V. den Oord, Conditional image generation with pixelcnn decoders, *Adv. Neural Inf. Process. Syst.* 29 (2016).
- [81] T. Salimans, A. Karpathy, X. Chen, D. P. Kingma, Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications, *arXiv preprint arXiv:1701.05517* (2017).
- [82] P. Esser, R. Rombach, A. Blattmann, B. Ommer, Imagebart: Bidirectional context with multinomial diffusion for autoregressive image synthesis, *Advances in neural information processing systems* 34 (2021) 3518–3532.
- [83] L. Dinh, D. Krueger, Y. Bengio, Nice: Non-linear independent components estimation, *arXiv preprint arXiv:1410.8516* (2014).
- [84] L. Dinh, J. Sohl-Dickstein, S. Bengio, Density estimation using real nvp, *arXiv preprint arXiv:1605.08803* (2016).
- [85] Q. Zhang, Y. Chen, Diffusion normalizing flow, *Advances in neural information processing systems* 34 (2021) 16280–16291.
- [86] Y. LeCun, S. Chopra, R. Hadsell, M. Ranzato, F. Huang, et al., A tutorial on energy-based learning, *Predicting structured data* 1 (0) (2006).
- [87] J. Ngiam, Z. Chen, P. W. Koh, A. Y. Ng, Learning deep energy models, in: *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 1105–1112.
- [88] K. Swersky, M. Ranzato, D. Buchman, N. D. Freitas, B. M. Marlin, On autoencoders and score matching for energy based models, in: *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 1201–1208.
- [89] F. Bao, K. Xu, C. Li, L. Hong, J. Zhu, B. Zhang, Variational (gradient) estimate of the score function in energy-based latent variable models, in: *International Conference on Machine Learning*, PMLR, 2021, pp. 651–661.
- [90] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, *Advances in neural information processing systems* 27 (2014).

- [91] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen, Improved techniques for training gans, *Advances in neural information processing systems* 29 (2016).
- [92] Y. Shen, J. Gu, X. Tang, B. Zhou, Interpreting the latent space of gans for semantic face editing, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 9243–9252.
- [93] H. Zhou, W. Li, P. Li, Y. Xu, L. Zhang, X. Zhou, Z. Zhao, E. Li, C. Lv, A novel few-shot learning framework based on diffusion models for high-accuracy sunflower disease detection and classification, *Plants* 14 (3) (2025) 339.
- [94] X. Wang, R. Wang, S. Wu, X. Ma, J. Du, J. Gao, A stable diffusion based image generation method for few-shot emerging pest detection in the wild, in: *2024 IEEE International Conference on Industrial Technology (ICIT)*, IEEE, 2024, pp. 1–6.
- [95] J. Huang, Z. Wang, M. Xu, L. Ma, W. Wu, J. Cao, Enhancing few-shot plant disease classification with diffusion model, in: *2024 China Automation Congress (CAC)*, IEEE, 2024, pp. 2036–2040.
- [96] T. Astuti, A. N. Umar, R. Wahyudi, Z. Rifai, Enhancing potato leaf disease detection: Implementation of convolutional vision transformers with synthetic datasets from stable diffusion, *JOIV: International Journal on Informatics Visualization* 8 (4) (2024) 2054–2065.
- [97] Z. Wang, H. Ma, C. Wu, L. Ma, Agricultural pest image generation for real-time detection with semantic diffusion and knowledge distillation, *Signal, Image and Video Processing* 19 (7) (2025) 523.
- [98] Y. Li, J. Guo, H. Qiu, F. Chen, J. Zhang, Denoising diffusion probabilistic models and transfer learning for citrus disease diagnosis, *Frontiers in plant science* 14 (2023) 1267810.
- [99] H. Guo, J. Liu, Y. Li, Y. Xu, K. Xu, A. Fan, J. Hao, Y. Hou, C. Lv, Efficient tomato disease detection using maxmin-diffusion mechanism and lightweight techniques, *Plants* 14 (3) (2025) 354.
- [100] L. Zhang, Y. Zhao, C. Zhou, J. Zhang, Y. Yan, T. Chen, C. Lv, Judifromer: Multimodal fusion model with transformer and diffusion for jujube disease detection, *Computers and Electronics in Agriculture* 232 (2025) 110008.
- [101] A. Heschl, M. Murillo, K. Najafian, F. Maleki, Synthset: generative diffusion model for semantic segmentation in precision agriculture, in: *European Conference on Computer Vision*, Springer, 2024, pp. 267–283.
- [102] H. Zhang, Q. Zhang, W. Yan, J. Gu, A generative data augmentation enhanced plant disease classification combined diffusion model, in: *2024 4th International Conference on Electronic Information Engineering and Computer Communication (EIECC)*, IEEE, 2024, pp. 874–878.
- [103] Y. Lee, Enhancing plant health classification via diffusion model-based data augmentation, *Multimedia Systems* 31 (2) (2025) 143.
- [104] D. Chen, X. Qi, Y. Zheng, Y. Lu, Y. Huang, Z. Li, Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition, *Computers and Electronics in Agriculture* 216 (2024) 108517.
- [105] B. Deng, Y. Lu, Weed image augmentation by controlnet-added stable diffusion for multi-class weed detection, *Computers and Electronics in Agriculture* 232 (2025) 110123.
- [106] S. Zhang, L. Liu, G. Li, Y. Du, X. Wu, Z. Song, X. Li, Diffusion model-based image generative method for quality monitoring of direct grain harvesting, *Computers and Electronics in Agriculture* 233 (2025) 110130.
- [107] X. Lu, J. Zhang, R. Yang, Q. Yang, M. Chen, H. Xu, P. Wan, J. Guo, F. Liu, Effective variance attention-enhanced diffusion model for crop field aerial image super resolution, *ISPRS Journal of Photogrammetry and Remote Sensing* 218 (2024) 50–68.
- [108] K. Hirahara, C. Nakane, H. Ebisawa, T. Kuroda, Y. Iwaki, T. Utsumi, Y. Nomura, M. Koike, H. Mineno, D4: Text-guided diffusion model-based domain adaptive data augmentation for vineyard shoot detection, *Computers and Electronics in Agriculture* 230 (2025) 109849.
- [109] H. Gunes, H. E. Cekli, U. Rist, Data enhancement, smoothing, reconstruction and optimization by kriging interpolation, in: *2008 Winter Simulation Conference*, IEEE, 2008, pp. 379–386.
- [110] P. Bhatt, A. Shah, S. Patel, S. Patel, Image enhancement using various interpolation methods, *International Journal of Computer Science and Information Technology & Security* 2 (4) (2012).
- [111] H. A. Salman, A. Kalakech, Image enhancement using convolution neural networks, *Babylonian Journal of Machine Learning* 2024 (2024) 30–47.
- [112] Z. Yu, M. Y. I. Idris, P. Wang, A diffusion-based framework for terrain-aware remote sensing image reconstruction, *arXiv preprint arXiv:2504.12112* (2025).
- [113] L. H. F. P. Silva, J. F. Mari, M. C. Escarpinat, A. R. Backes, Evaluating combinations of optimizers and loss functions for cloud removal using diffusion models, *Proceedings Copyright* 648 656.
- [114] J. Fan, Z. Shi, Z. Ren, Y. Zhou, M. Ji, Ddpm-segformer: Highly refined feature land use and land cover segmentation with a fused denoising diffusion probabilistic model and transformer, *International Journal of Applied Earth Observation and Geoinformation* 133 (2024) 104093.
- [115] L. Han, Y. Zhao, H. Lv, Y. Zhang, H. Liu, G. Bi, Q. Han, Enhancing remote sensing image super-resolution with efficient hybrid conditional diffusion model, *Remote Sensing* 15 (13) (2023) 3452.
- [116] A. Graikos, S. Yellapragada, M.-Q. Le, S. Kapse, P. Prasanna, J. Saltz, D. Samaras, Learned representation-guided diffusion models for large-image generation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8532–8542.
- [117] X. Hu, Z. Zhang, L. Zheng, T. Chen, C. Peng, Y. Wang, R. Li, X. Lv, S. Yan, Enhancing jujube forest growth estimation and disease detection using a novel diffusion-transformer architecture, *Plants* 13 (17) (2024) 2348.
- [118] W. Chen, P. Zhou, L. He, H. Wang, B. Cao, Performance analysis of image augmentation driven by language based on diffusion models, in: *International Conference on Applied Intelligence*, Springer, 2024, pp. 73–82.
- [119] S. Xiang, P. M. Blok, J. Burridge, H. Wang, W. Guo, Doda: Adapting object detectors to dynamic agricultural environments in real-time with diffusion, *arXiv preprint arXiv:2403.18334* (2024).
- [120] J. Kong, C. Hua, X. Jin, N. Guo, L. Peng, An effective object detector via diffused graphic large selective kernel with one-to-few labelling strategy for small-scaled crop diseases detection, *Crop Protection* 182 (2024) 106705.
- [121] Y. Jiang, S. Liu, H. Wang, Diffusion-based remote sensing image fusion for classification, *Applied Intelligence* 55 (4) (2025) 247.
- [122] L. Zhang, C. Li, X. Wu, H. Xiang, Y. Jiao, H. Chai, Bo-cnn-bilstm deep learning model integrating multisource remote sensing data for improving winter wheat yield estimation, *Frontiers in Plant Science* 15 (2024) 1500499.

- [123] C. Zhou, X. Ge, Y. Chang, M. Wang, Z. Shi, M. Ji, T. Wu, C. Lv, A multimodal parallel transformer framework for apple disease detection and severity classification with lightweight optimization, *Agronomy* 15 (5) (2025) 1246.
- [124] I. Egusquiza, L. Benito-Del-Valle, A. Picón, A. Bereciartua-Pérez, L. Gómez-Zamanillo, A. Elola, E. Aramendi, R. Espeso, T. Eggers, C. Klukas, et al., When synthetic plants get sick: Disease graded image datasets by novel regression-conditional diffusion models, *Computers and Electronics in Agriculture* 229 (2025) 109690.
- [125] H. Qi, Z. Huang, B. Jin, Q. Tang, L. Jia, G. Zhao, D. Cao, Z. Sun, C. Zhang, Sam-gan: An improved dcgan for rice seed viability determination using near-infrared hyperspectral imaging, *Computers and Electronics in Agriculture* 216 (2024) 108473.
- [126] T. Karras, S. Laine, 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.
- [127] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila, Alias-free generative adversarial networks, *Advances in neural information processing systems* 34 (2021) 852–863.
- [128] Q. Zeng, L. Niu, S. Wang, W. Ni, Sevit: a large-scale and fine-grained plant disease classification model based on transformer and attention convolution, *Multimedia Systems* 29 (3) (2023) 1001–1010.
- [129] X. Wang, L. Xie, C. Dong, 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.
- [130] Y. Zhang, S. Wa, L. Zhang, C. Lv, Automatic plant disease detection based on trnvolution detection network with gan modules using leaf images, *Frontiers in Plant Science* 13 (2022) 875693.
- [131] C.-Y. Wang, I.-H. Yeh, H.-Y. Mark Liao, Yolov9: Learning what you want to learn using programmable gradient information, in: *European conference on computer vision*, Springer, 2024, pp. 1–21.
- [132] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [133] M. Tan, R. Pang, Q. V. Le, Efficientdet: Scalable and efficient object detection, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10781–10790.
- [134] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, et al., Yolov10: Real-time end-to-end object detection, *Advances in Neural Information Processing Systems* 37 (2024) 107984–108011.
- [135] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: *European conference on computer vision*, Springer, 2020, pp. 213–229.
- [136] H.-T. Thai, K.-H. Le, N. L.-T. Nguyen, Formerleaf: An efficient vision transformer for cassava leaf disease detection, *Computers and Electronics in Agriculture* 204 (2023) 107518.
- [137] D. Pang, H. Wang, J. Ma, D. Liang, Dctn: a dense parallel network combining cnn and transformer for identifying plant disease in field, *Soft Computing* 27 (21) (2023) 15549–15561.
- [138] X. Li, X. Li, S. Zhang, G. Zhang, M. Zhang, H. Shang, Slvit: Shuffle-convolution-based lightweight vision transformer for effective diagnosis of sugarcane leaf diseases, *Journal of King Saud University-Computer and Information Sciences* 35 (6) (2023) 101401.
- [139] A. Sebastian, A. A. Fathima, R. Pragna, S. MadhanKumar, G. Y. Kannan, V. Murali, Vital: An advanced framework for automated plant disease identification in leaf images using vision transformers and linear projection for feature reduction, in: *International Conference on Computing and Machine Learning*, Springer, 2024, pp. 31–45.
- [140] H. Wang, T.-H. Nguyen, T. N. Nguyen, M. Dang, Pd-tr: End-to-end plant diseases detection using a transformer, *Computers and Electronics in Agriculture* 224 (2024) 109123.
- [141] M. Zhou, R. Xing, D. Han, Z. Qi, G. Li, Pdt: Uav target detection dataset for pests and diseases tree, in: *European Conference on Computer Vision*, Springer, 2024, pp. 56–72.
- [142] V. Sharma, A. K. Tripathi, H. Mittal, L. Nkenyereye, Soyatrans: A novel transformer model for fine-grained visual classification of soybean leaf disease diagnosis, *Expert Systems with Applications* 260 (2025) 125385.
- [143] F. Wang, R. Wang, Z. Huang, S. Dong, X. Wang, Q. Zhou, S. Zheng, L. Liu, Eact-det: An efficient adjusting criss-cross windows transformer embedding pyramid networks for similar disease detection, *Multimedia Tools and Applications* 83 (18) (2024) 55683–55707.
- [144] R. Karthik, R. Menaka, S. Ompirakash, P. B. Murugan, M. Meenakashi, S. Lingaswamy, D. Won, Grapeleafnet: A dual-track feature fusion network with inception-resnet and shuffle-transformer for accurate grape leaf disease identification, *IEEE Access* 12 (2024) 19612–19624.
- [145] Y. Zhang, C. Lv, Tinysegformer: A lightweight visual segmentation model for real-time agricultural pest detection, *Computers and Electronics in Agriculture* 218 (2024) 108740.