Noname manuscript No.
(will be inserted by the editor)

Generating Multi-Illumination Leaf Images through Single-Image Material Estimation for Robust Disease Recognition

Anonymous author

Received: date / Accepted: date

Abstract Visual conditions are critical for vision-based smart agricultural production tasks. However, traditional field acquisition methods struggle to capture complex illumination variations and are often time-consuming, labor-intensive, and expensive. In this work, we propose an innovative method for generating datasets of crop leaf images under diverse visual conditions, which synthesizes high-fidelity leaf data through physics-based simulation of realistic agricultural illumination variations. Our methodology comprises three stages. First, we use advanced techniques to estimate depth from the input image and reconstruct accurate leaf geometry. Second, we employ a physics-based inverse rendering framework to estimate the material and illumination characteristics of leaf surfaces. Finally, high dynamic range (HDR) environment maps are utilized to achieve photorealistic relighting across diverse agricultural conditions. Extensive experiments conducted on the AppleLeaf9 dataset validate the effectiveness of the proposed image generation method for leaf disease classification across various deep learning architectures. The synthesized images achieved classification accuracies of 97.10% with MobileNetV2, 95.65% with DenseNet121, 96.62% with ResNet50, and 98.07% with VGG16. These results underscore the strong potential of the proposed method in agricultural vision applications, particularly in addressing domain-specific challenges that are often insufficiently handled by existing image generation approaches.

Keywords Image synthesis · Inverse rendering · Relighting · Crop disease classification

F. Author
first address
Tel.: +123-45-678910
Fax: +123-45-678910
E-mail: fauthor@example.com

S. Author
second address

1 Introduction

Advances in computer vision have significantly enhanced smart agricultural systems. While dataset construction remains fundamental to these developments, current agricultural image collections face critical limitations. Most existing datasets are captured under controlled experimental conditions, lacking sufficient illumination diversity to represent dynamic field environments affected by seasonal variations, weather fluctuations, and diurnal changes. These environmental dynamics lead to significant variations in leaf color, surface texture, and the visual manifestation of disease symptoms. Large-scale image collection and annotation require substantial human and financial investment, making it difficult to comprehensively address practical agricultural scenarios [1]. Therefore, expanding crop leaf image datasets incorporating a variety of visual conditions is essential for advancing vision-based research in smart agricultural production.

Recent advancements in generative adversarial networks (GANs) [2] and diffusion models have demonstrated remarkable capabilities in photorealistic image synthesis, emerging as powerful tools for agricultural vision research [3,4]. To improve feature richness and image fidelity, Liu [5] proposed LeafGAN, incorporating a densely connected architecture to capture finer details. Conditional GANs [6] have been employed to synthesize tomato leaf images, leading to substantial improvements in classification accuracy. Despite these successes, GANs still face several challenges, including training instability, mode collapse, and reliance on large-scale datasets [7,8]. Li [9] introduced Denoising Diffusion Probabilistic Models [10] for data augmentation, demonstrating their effectiveness in citrus disease diagnosis under limited data conditions. Furthermore, diffusion models were applied to weed identification tasks, significantly improving the quality of generated weed images [11]. Nevertheless, diffusion models often produce images that lack physical realism and offer limited controllability over specific visual attributes.

Specifically in this work, we introduce a simple yet effective image generation method for synthesizing crop leaf images, which simulates various typical visual conditions in agricultural environments to enhance the diversity and realism of leaf disease datasets. Starting from a single image, our approach reconstructs the 3D leaf geometry and estimates material and lighting properties using advanced physics-based inverse rendering techniques. To faithfully simulate the diverse and naturally occurring illumination scenarios found in outdoor agricultural scenes, we incorporate high dynamic range (HDR) environment maps during the relighting process. Our analysis involves training classification models on various combinations of real and synthetic datasets, accompanied by quantitative performance comparisons across multiple deep learning architectures. The results demonstrate that our synthetic images not only enhance classification accuracy but also contribute to more diverse and representative training data. Through this work, we validate the effectiveness of our synthetic generation method in capturing complex illumination scenarios and faithfully replicating real-world conditions, thereby supporting the development of more accurate and robust vision systems for agricultural applications.

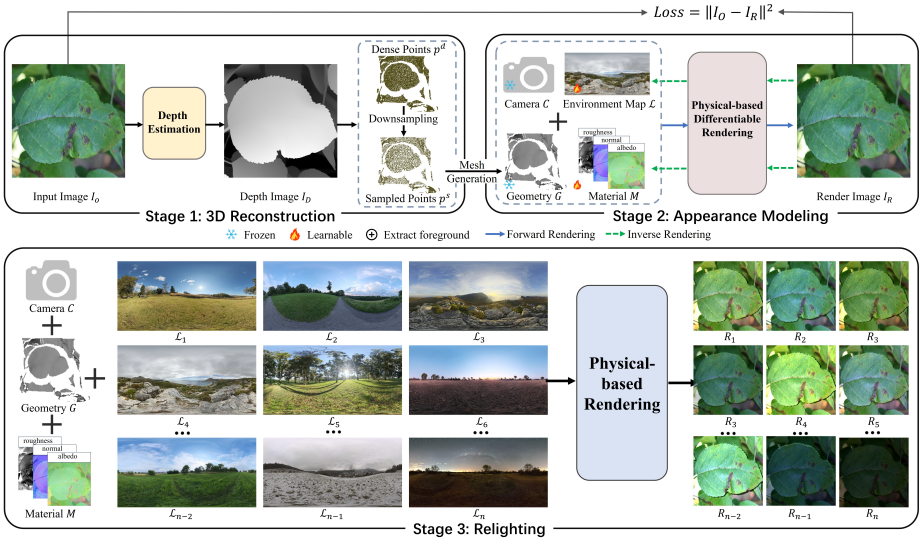

Fig. 1. The pipeline of the proposed framework. Given a single-view image, we first reconstruct the leaf geometry as a 3D mesh by generating a point cloud and performing surface reconstruction. Next, physics-based inverse rendering is employed to jointly optimize the leaf's material properties and scene illumination. Finally, the reconstructed scene is relit under HDR environment maps to achieve photorealistic appearance synthesis.

2 Related Work

2.1 Inverse Rendering

Inverse rendering aims to decompose the visual appearance of an image into its underlying geometric structure, material properties, and illumination [12, 13, 14, 15], and has become a foundational problem in both computer vision and graphics. Early work such as [16] introduced the first single-image inverse rendering framework under unknown environmental illumination. With the emergence of differentiable renderers, inverse rendering has been successfully applied to various domains, including human face reconstruction [17, 18] and photorealistic indoor scene modeling [19, 20]. Despite these advancements, the application of inverse rendering in agriculture remains largely unexplored. In this study, we adapt and extend differentiable inverse rendering techniques to the context of crop leaf image enhancement.

2.2 Generative Adversarial Networks (GANs)

GANs consist of two core components: a generator and a discriminator. The generator aims to produce images that closely resemble real samples, while the discriminator evaluates whether the input images are real or generated. The generator's

training objective is to maximize the probability of the discriminator making incorrect predictions, and through adversarial training, both components optimize each other. As adversarial training progresses, the generator and discriminator eventually converge to a stable equilibrium [2]. GANs have been widely employed for synthesizing crop disease images. These studies [21, 22] modified the original GAN architecture, effectively generating more realistic images and improving the performance of leaf disease classification. AR-GAN extends traditional GAN-based image-to-image translation methods, producing visually realistic tomato leaves and enhancing disease recognition in data-scarce environments [23]. Despite their ability to generate high-fidelity images, GAN-based methods require large-scale training datasets and often underperform when trained on limited data. Furthermore, the joint training of the generator and discriminator is computationally intensive and time-consuming.

2.3 Diffusion Models

Diffusion models involve two essential processes: a forward diffusion process and a reverse denoising process. Inspired by non-equilibrium thermodynamics [24, 25], the forward process incrementally adds Gaussian noise to an image until it becomes nearly pure noise, while the reverse process progressively removes the noise to generate samples that approximate the original data distribution [26]. Leveraging their strong generative stability, diffusion models have recently emerged as powerful alternatives to GANs [7], particularly in the context of agricultural image synthesis. TomatoDIFF [27] was introduced for the semantic segmentation of tomatoes in complex environments. Stable Diffusion is a text-to-image generation model [28], which was adopted by Moreno [29] to synthesize weed images within tomato crops. However, due to substantial variation in lighting conditions among collected images, multiple models had to be trained separately for different image categories. Furthermore, Deng [30] introduced the use of ControlNet for weed detection, incorporating both bounding boxes and text prompts as control signals. Despite offering a higher degree of controllability, this approach lacked robustness in consistently generating images that strictly adhered to specified content and visual constraints.

3 Method

Our pipeline for generating synthetic images of leaves is illustrated in Fig. 1. In Section 3.1, we reconstruct 3D models of leaves. In Section 3.2, we refine leaf material parameters through physics-based differentiable rendering. Section 3.3 leverages HDR environment maps to simulate dynamic illumination conditions inherent to agricultural ecosystems, enabling scalable generation of diverse and photorealistic leaf images.

3.1 Geometry Modeling

Given a single-view leaf image I_O, we reconstruct its geometry as a triangular mesh G. Using a pre-trained Depth Anything V2 network [31], we first predict



Fig. 2. Qualitative comparison of inverse rendering and real images. (a) Input reference images; (b) Re-rendered results using estimated illumination and material parameters.

the depth map I_D. Each pixel (u, v) is then back-projected to a 3D point p, forming a dense point cloud $p^d = \{p_1^d, p_2^d, \ldots, p_n^d\}$. To reduce computation cost while preserving geometric features, we apply farthest point sampling to obtain a uniformly distributed subset $p^s = \{p_1^s, p_2^s, \ldots, p_m^s\}\,(m < n)$. Finally, we reconstruct the triangular mesh G using the Ball Pivoting algorithm.

3.2 Appearance Modeling

Camera Parameters Accurate camera parameter specification constitutes a critical prerequisite for precise scene inference in inverse rendering pipelines. Nevertheless, agricultural leaf datasets frequently lack camera metadata. We propose an automated camera estimation framework grounded in geometric normalization. Assuming the camera is an orthographic camera positioned directly above the leaf at coordinates (0, 0, 1) and oriented toward (0, 0, 0), our method first computes the minimum bounding box R_b surrounding each leaf, where W_b and H_b represent the width and height of R_b. The geometric center of R_b is translated to the image center, ensuring that the camera remains directly above the leaf. To guarantee that the entire leaf remains visible within the rendered frame, we adjust the orthographic scale based on the ratio between the normalized model space and its actual size in image pixels. The orthographic camera's scale s in Mitsuba specifies the half-height of the projection plane. When the leaf's geometric model is normalized to $[-n, n]^3$, the camera scale for a rendered image with resolution $W_r \times W_r$ is computed as:

$$C_{scale} = \frac{W_r}{\max(W_b, H_b)} \times n \tag{1}$$

In our implementation, we set $n = 0.5$. Following the above procedure, the estimated camera parameters C can be directly used as known inputs in the inverse rendering process.

Material Estimation Based on Differentiable Rendering Rendering can be defined as a complex function g that maps a set of scene parameters θ to output imagery, formally expressed as:

$$I_R = \mathrm{g}(\theta) = \mathrm{g}(G, M, L, C) \tag{2}$$

where G, M, L, C represent geometry, materials, illumination, and camera parameters. The function g involves solving rendering equations by various numerical methods such as ray tracing or path tracing. The rendering equation follows the physical law of energy conservation and was proposed by Kajiya [32] to be a mathematical model to describe light propagation and interaction in a scene:

$$L_o(\omega_o, x) = \int_{\Omega} L_i(\omega_i, x) f_r(\omega_o, \omega_i, x)(\omega_i \cdot n) d\omega_i \tag{3}$$

where x denotes surface position, n is the surface normal, Ω represents the hemisphere space in the direction of incident light, ω_i and ω_o are the directions of incoming and outgoing light, L_i and L_o denote the radiance corresponding to those directions, f_r is the Spatially Varying Bidirectional Reflectance Distribution Function describing how different materials reflect the incident light. Our methodology follows Eq. (3) and uses radiance to calculate pixel values of the rendered image. For material modeling, we implement the Disney microfacet BRDF [33] as f_r in the rendering equation:

$$f_r(\omega_i, \omega_o) = \frac{f \cdot d \cdot g}{4(\omega_i \cdot n)(\omega_o \cdot n)} \tag{4}$$

where f is the Fresnel reflection coefficient, d is the microfacet distribution function, and g is the geometric attenuation factor. We encode material parameters (albedo, roughness) as spatially varying texture maps UV-mapped onto leaf geometry, with refractive indices stored as scalar values. This representation enables efficient material estimation in complex scenes.

Loss Function Our goal is to calculate the gradient with respect to scene parameters and optimize the environment illumination L and materials M by minimizing the difference between the rendered image I_R and the reference image I_O:

$$\mathcal{L}_{IR} = \lambda_{RGB}\mathcal{L}_{RGB} + \lambda_M\mathcal{L}_M + \lambda_{TV}\mathcal{L}_{TV} \tag{5}$$

where $\mathcal{L}_{RGB} = \|I_O - \mathrm{g}(\theta)\|^2$ is the pixel-wise RGB loss, $\mathcal{L}_M$ is defined as the mean square error loss between the true mask and generated mask, $\mathcal{L}_{TV}$ is the image total variation loss, which is a commonly used regularization technique in image processing. λ_{RGB}, λ_M and λ_{TV} represent the coefficient weights of each loss term.

3.3 Relighting

Natural illumination predominantly originates from skylight and ambient sources, so we use HDR environment maps to simulate the appearance variations of leaves under real agricultural conditions. The incident radiance at a given surface point can be derived from the environment map by evaluating it along the corresponding reflection vector. Specifically, the incident radiance in the direction ω_i is given by:

$$L_i(\omega_i) = L_{env}(N(\omega_i)) \tag{6}$$

where $N(\omega_i)$ denotes the reflection direction of the incident vector ω_i about the surface normal, and L_{env} represents the radiance function defined by the environment map.

A set of diverse environment maps $\{E_i\}_{i=1}^{n}$ can be employed to simulate a variety of natural illumination conditions. Using ray tracing, we compute the intersection point x between each camera ray ω_i and the leaf geometry surface G to determine the corresponding surface interaction. Given the surface normal n at the intersection point, the material properties M, and the illumination defined by E, the outgoing radiance along the viewing direction ω_o is evaluated using the Eq. (3).

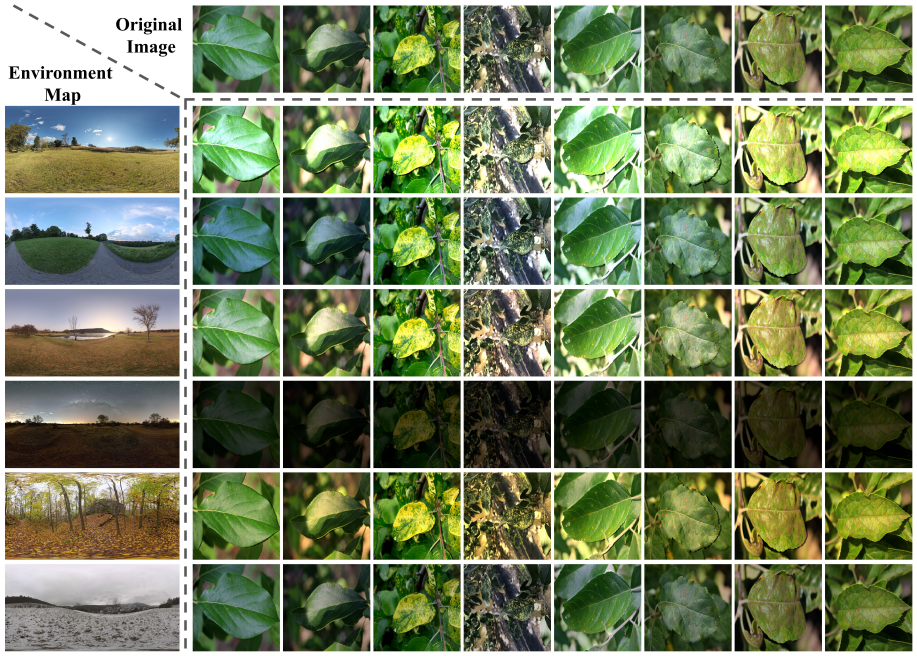

Fig. 3. Our method re-renders leaves under various illumination conditions using different environment maps. Specifically, the second to the seventh rows correspond to sunny, cloudy, evening, night, autumn, and winter lighting, respectively.

4 Experiments and Results

4.1 Dataset

To evaluate the effectiveness of our method, we conducted comparative experiments on the AppleLeaf9 dataset [34], which includes 14,582 images spanning healthy leaves and eight disease categories. Approximately 94% of the samples were collected in the wild, while the remaining 6% were obtained indoors under controlled settings, with an imbalanced distribution across categories. To simulate data scarcity scenarios, we selected four representative categories (healthy, mosaic, rust, and scab) and excluded all indoor-acquired samples. Through random

Table 1
Quantitative comparison of image quality metrics across different categories.

Categories	SSIM	PSNR	MSE
Healthy	0.9804	34.9075	0.0004
Mosaic	0.9598	31.7933	0.0008
Rust	0.9834	34.4767	0.0004
Scab	0.9836	35.0356	0.0003

sampling, we constructed a baseline dataset comprising 643 field-collected images: 169 healthy, 132 mosaic, 186 rust, and 156 scab, which was partitioned into 70% training-validation and 30% testing subsets. The training-validation set served as the foundation for generating synthetic images, with each synthesis method used to produce a 10-fold expansion of the dataset. All images were standardized to a resolution of 512×512.

4.2 Experiment Setup and Evaluation Metrics

We evaluated the effectiveness of our multi-illumination synthetic images by training four advanced classification models to distinguish between healthy and diseased leaves, including MobileNetV2 [35], DenseNet121 [36], ResNet50 [37], and VGG16 [38]. All models were trained under identical conditions on an NVIDIA RTX 3090 GPU for 30 epochs, with a batch size of 16, a learning rate of 0.001, and a momentum of 0.9. Classification performance was evaluated using four standard metrics: recognition accuracy, precision, recall, and F1 score.

4.3 Evaluation of Inverse Rendering and Relighting Results

Our method demonstrates the robustness of inverse rendering results. As shown in Fig. 2, the synthetic image generated with optimized material and illumination parameters is nearly indistinguishable from the real reference image. Table 1 provides quantitative evidence of this fidelity through standard image quality metrics (SSIM, PSNR, and MSE), showing consistently high scores across different categories. The qualitative results in Fig. 3 further validate our approach, where the scene is relit using multiple distinct HDR environment maps to simulate complex outdoor illumination scenarios, including direct sunlight, overcast skies, and other natural lighting conditions. These results indicate that the estimated leaf materials from the inverse rendering stage exhibit high fidelity and generalize well to novel lighting conditions, producing visually realistic renderings. The rendered leaf images preserve fine geometric and textural details and accurately capture the effects of varying illumination conditions.

4.4 Performance Evaluation of Classification Task

To more comprehensively validate the advantages of our approach, we compared it with several existing approaches. StyleGAN2-ADA [39] and FastGAN [40] are

Table 2
Performance comparison of classification models with different methods

Model	Method	Accuracy(%)	Precision(%)	Recall(%)	F1 Score(%)
MobileNetV2	None	77.29	79.50	77.31	77.61
	StyleGan2-ADA	95.17	95.40	95.30	95.20
	FastGan	96.62	96.62	96.91	96.73
	DreamBooth	91.30	91.54	91.55	91.47
	SVDiff	86.96	87.82	86.97	87.21
	Ours	**97.10**	**97.09**	**97.37**	**97.21**
DenseNet121	None	81.64	82.55	82.52	82.50
	StyleGan2-ADA	94.69	95.02	95.11	95.04
	FastGan	94.20	94.63	94.69	94.61
	DreamBooth	89.37	89.38	89.51	89.17
	SVDiff	93.72	93.97	93.59	93.72
	Ours	**95.65**	**95.84**	**95.78**	**95.72**
ResNet50	None	70.53	71.78	72.67	71.44
	StyleGan2-ADA	93.72	93.91	93.98	93.79
	FastGan	94.69	94.80	95.02	94.89
	DreamBooth	79.71	80.35	80.91	80.59
	SVDiff	91.30	92.01	91.70	91.80
	Ours	**96.62**	**96.50**	**96.83**	**96.64**
VGG16	None	77.29	78.82	78.10	78.25
	StyleGan2-ADA	96.62	96.61	96.76	96.65
	FastGan	94.20	94.23	94.56	94.04
	DreamBooth	81.64	87.59	80.78	82.19
	SVDiff	90.82	90.76	91.45	90.88
	Ours	**98.07**	**98.10**	**98.16**	**98.11**

GAN-based models that can be applied to few-shot image generation tasks. DreamBooth [41] and SVDiff [42] are diffusion-based fine-tuning methods that can acquire novel concepts from just 3–5 images, while maintaining diversity in the generated outputs under different lighting-related prompts.

Table 2 presents the classification performance comparison of four models on the original dataset and various synthesized datasets. The boldfaces show the best evaluation metric for each classification model. Using the original unenhanced dataset as the baseline, all image generation methods demonstrate significant performance improvements. Both StyleGAN2-ADA and FastGAN notably enhance classification accuracy, while diffusion-based methods DreamBooth and SVDiff show limited performance gains due to instability in structural consistency and quality control of generated images. In contrast, our proposed method achieves substantially greater advancements. Taking VGG16 as an example, its test accuracy on the DreamBooth enhanced dataset reaches 86.96%, representing only a 4.35% improvement over the baseline. However, with our approach, the accuracy surges to 98.07%, marking a 20.78% increase from the baseline and outperforming all comparative methods. Similar trends emerge across other models: MobileNetV2 achieves 97.10% accuracy with our method, DenseNet121 reaches 95.65%, and ResNet50 shows remarkable improvement from 70.53% on the baseline dataset to 96.62%. Across all four classification models, our approach achieves better performance in all metrics. Fig. 4 illustrates the confusion matrices of VGG16 across different methods. The classification network trained on original data struggles to distinguish between Healthy and Scab disease categories, whereas our method effectively resolves this issue, with only 4 misclassified samples in the test set. More results can be found in the supplementary materials.

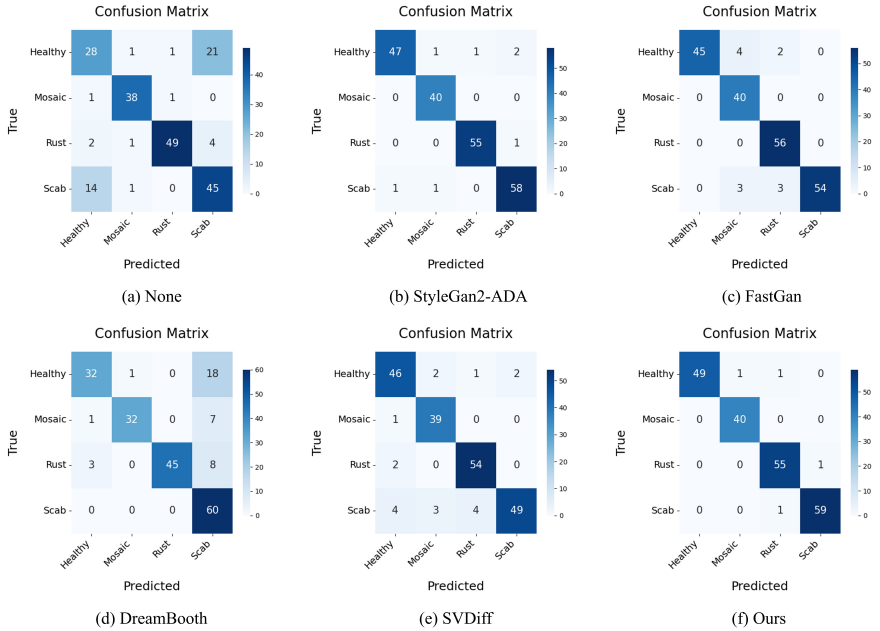

Fig. 4. Confusion matrices of the VGG16 classification model with different data augmentation methods.

Experimental results demonstrate that our proposed method enhances the illumination diversity of the leaves, enabling deep neural networks to better extract leaf features during training. The quantitative improvements across multiple architectures and metrics confirm the effectiveness of our approach in addressing critical challenges in real-world agricultural vision tasks.

5 Discussion and Conclusion

Collecting a large number of crop leaf images under diverse illumination conditions is a complex and costly task. Although contemporary generative models are capable of synthesizing high-fidelity image datasets, these models still rely on a substantial amount of agricultural data for training and face difficulties in generating leaf images under varying lighting conditions. In this study, we propose a simple yet efficient new method that learns the material parameters and lighting information of the leaves through inverse rendering, and then uses HDR environment maps for relighting to simulate the visual changes of the leaves under different illumination conditions. Our method demonstrates significant value in realistic smart agriculture tasks, particularly in the analysis of diseased leaves.

References

1. Sarkar, C., Gupta, D., Gupta, U., Hazarika, B.B.: Leaf disease detection using machine learning and deep learning: Review and challenges. Applied Soft Computing **145**, 110,534 (2023)
2. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
3. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys **56**(4), 1–39 (2023)
4. Lu, Y., Chen, D., Olaniyi, E., Huang, Y.: Generative adversarial networks (gans) for image augmentation in agriculture: A systematic review. Computers and Electronics in Agriculture **200**, 107,208 (2022)
5. Liu, B., Tan, C., Li, S., He, J., Wang, H.: A data augmentation method based on generative adversarial networks for grape leaf disease identification. IEEE Access **8**, 102,188–102,198 (2020)
6. Abbas, A., Jain, S., Gour, M., Vankudothu, S.: Tomato plant disease detection using transfer learning with c-gan synthetic images. Computers and Electronics in Agriculture **187**, 106,279 (2021)
7. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
8. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., Bharath, A.A.: Generative adversarial networks: An overview. IEEE signal processing magazine **35**(1), 53–65 (2018)
9. Li, Y., Guo, J., Qiu, H., Chen, F., Zhang, J.: Denoising diffusion probabilistic models and transfer learning for citrus disease diagnosis. Frontiers in plant science **14**, 1267,810 (2023)
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
11. Chen, D., Qi, X., Zheng, Y., Lu, Y., Huang, Y., Li, Z.: Synthetic data augmentation by diffusion probabilistic models to enhance weed recognition. Computers and Electronics in Agriculture **216**, 108,517 (2024)
12. Dong, Y., Chen, G., Peers, P., Zhang, J., Tong, X.: Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. ACM Transactions on Graphics (TOG) **33**(6), 1–12 (2014)
13. Sang, S., Chandraker, M.: Single-shot neural relighting and svbrdf estimation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16, pp. 85–101. Springer (2020)
14. Zhang, Y., Sun, J., He, X., Fu, H., Jia, R., Zhou, X.: Modeling indirect illumination for inverse rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18,643–18,652 (2022)
15. Boss, M., Braun, R., Jampani, V., Barron, J.T., Liu, C., Lensch, H.: Nerd: Neural reflectance decomposition from image collections. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12,684–12,694 (2021)
16. Li, Z., Xu, Z., Ramamoorthi, R., Sunkavalli, K., Chandraker, M.: Learning to reconstruct shape and spatially-varying reflectance from a single image. ACM Transactions on Graphics (TOG) **37**(6), 1–11 (2018)
17. Aldrian, O., Smith, W.A.: Inverse rendering of faces with a 3d morphable model. IEEE transactions on pattern analysis and machine intelligence **35**(5), 1080–1093 (2012)
18. Sengupta, S., Kanazawa, A., Castillo, C.D., Jacobs, D.W.: Sfsnet: Learning shape, reflectance and illuminance of facesin the wild'. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6296–6305 (2018)
19. Sengupta, S., Gu, J., Kim, K., Liu, G., Jacobs, D.W., Kautz, J.: Neural inverse rendering of an indoor scene from a single image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8598–8607 (2019)
20. Li, Z., Shafiei, M., Ramamoorthi, R., Sunkavalli, K., Chandraker, M.: Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2475–2484 (2020)

21. Yilma, G., Belay, S., Qin, Z., Gedamu, K., Ayalew, M.: Plant disease classification using two pathway encoder gan data generation. In: 2020 17th international computer conference on wavelet active media technology and information processing (ICCWAMTIP), pp. 67–72. IEEE (2020)
22. Zhang, L., Zhou, G., Lu, C., Chen, A., Wang, Y., Li, L., Cai, W.: Mmdgan: A fusion data augmentation method for tomato-leaf disease identification. Applied Soft Computing **123**, 108,969 (2022)
23. Nazki, H., Yoon, S., Fuentes, A., Park, D.S.: Unsupervised image translation using adversarial networks for improved plant disease recognition. Computers and Electronics in Agriculture **168**, 105,117 (2020)
24. Jarzynski, C.: Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. Physical Review E **56**(5), 5018 (1997)
25. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning, pp. 2256–2265. PMLR (2015)
26. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
27. Ivanovska, M., Štruc, V., Perš, J.: Tomatodiff: On–plant tomato segmentation with denoising diffusion models. In: 2023 18th International Conference on Machine Vision and Applications (MVA), pp. 1–6. IEEE (2023)
28. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10,684–10,695 (2022)
29. Moreno, H., Gómez, A., Altares-López, S., Ribeiro, A., Andújar, D.: Analysis of stable diffusion-derived fake weeds performance for training convolutional neural networks. Computers and Electronics in Agriculture **214**, 108,324 (2023)
30. Deng, B., Lu, Y.: Weed image augmentation by controlnet-added stable diffusion. In: Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications II, vol. 13035, pp. 163–175. SPIE (2024)
31. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. arXiv preprint arXiv:2406.09414 (2024)
32. Kajiya, J.T.: The rendering equation. In: Proceedings of the 13th annual conference on Computer graphics and interactive techniques, pp. 143–150 (1986)
33. Burley, B., Studios, W.D.A.: Physically-based shading at disney. In: Acm Siggraph, vol. 2012, pp. 1–7. vol. 2012 (2012)
34. Yang, Q., Duan, S., Wang, L.: Efficient identification of apple leaf diseases in the wild using convolutional neural networks. Agronomy **12**(11) (2022). DOI 10.3390/agronomy12112784. URL `https://www.mdpi.com/2073-4395/12/11/2784`
35. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4510–4520 (2018)
36. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708 (2017)
37. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
38. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
39. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. Advances in neural information processing systems **33**, 12,104–12,114 (2020)
40. Liu, B., Zhu, Y., Song, K., Elgammal, A.: Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In: iclr (2021)
41. Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., Aberman, K.: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 22,500–22,510 (2023)
42. Han, L., Li, Y., Zhang, H., Milanfar, P., Metaxas, D., Yang, F.: Svdiff: Compact parameter space for diffusion fine-tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7323–7334 (2023)