

Evaluation of One-image 3D Reconstruction for Plant Model Generation

Zihe Gao

`zach.gao@nottingham.ac.uk`

University of Nottingham

Zane Hartley

University of Nottingham

Andrew French

University of Nottingham

Research Article

Keywords: 3D plant reconstruction, image-conditioned generation, diffusion models, plant digitisation, PlantDreamer dataset, single-image reconstruction

Posted Date: October 6th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7577309/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Methods on January 13th, 2026. See the published version at <https://doi.org/10.1186/s13007-025-01482-6>.

Evaluation of One-image 3D Reconstruction for Plant Model Generation

Zihe Gao^{1*}, Zane Hartley¹ and Andrew French¹

^{1*}School of Computer Science, University of Nottingham, Wollaton Road, Nottingham, NG8 1BB, UK.

*Corresponding author(s). E-mail(s): zach.gao@nottingham.ac.uk;
Contributing authors: zane.hartley@nottingham.ac.uk;
andrew.p.french@nottingham.ac.uk;

Abstract

Generating accurate and visually realistic 3D models of plants from single-view images is crucial yet remains challenging due to plants’ intricate geometry and frequent occlusions. This capability matters because it supplements current plant datasets and enables non-destructive, high-throughput phenotyping for crop breeding and precision agriculture. More broadly, 3D reconstruction is particularly important because plant morphology is inherently three-dimensional, while 2D representations miss occluded leaves, branching geometry, and volumetric traits. However, plants present unique challenges compared to common rigid objects, and most current generative methods have not been systematically tested in this domain, leaving a gap in understanding their reliability for realistic plant reconstruction. This study systematically evaluates six advanced generative techniques—Hunyuan3D 2.0, Trellis (Structured 3D Latents), One2345++, InstantMesh, Direct3D and Unique3D—using the existing PlantDreamer dataset. Specifically, this research reconstructs mesh models from images of Bean plants and quantitatively assesses each method’s performance against ground-truth scans using Chamfer Distance, Normal Consistency, F-Score, PSNR, LPIPS, and CLIP Score. The paper also presents qualitative results of Kale and Mint plants. The results indicate that Hunyuan3D 2.0 achieves superior performance overall, suggesting its effectiveness in capturing complex plant structures. This work provides valuable insights into strengths and limitations of contemporary 3D generative approaches, guiding future improvements in realistic plant digitisation.

Keywords: 3D plant reconstruction, image-conditioned generation, diffusion models, plant digitisation, PlantDreamer dataset, single-image reconstruction

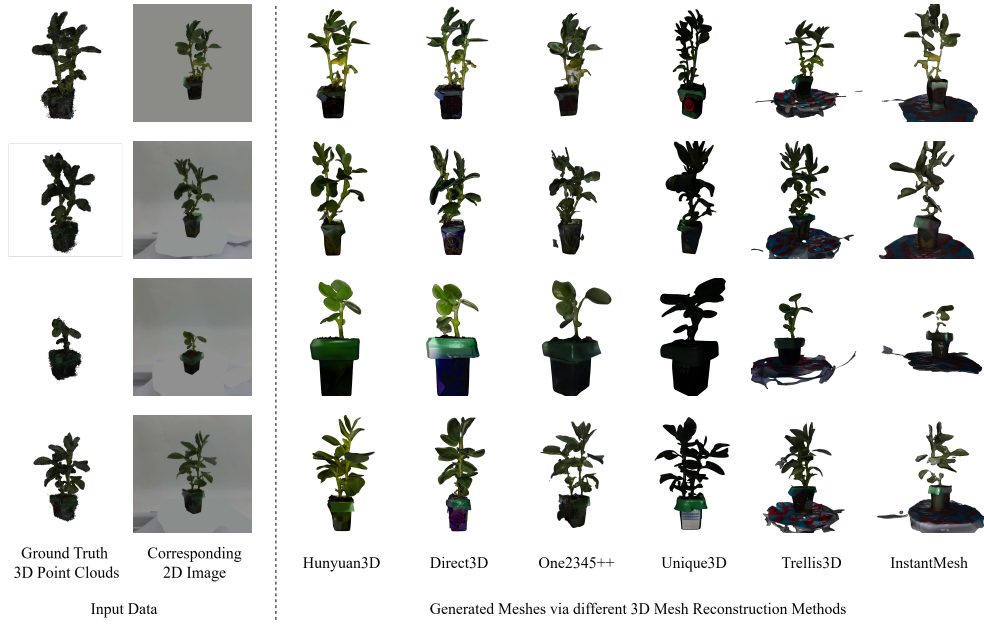


Fig. 1: Overview of this evaluation research. On the left are examples of input data, including ground truth 3D plant point clouds and corresponding 2D frontal-view images. The right side shows examples of generated meshes from the given 2D image via different reconstruction methods.

1 Introduction

The precise and realistic reconstruction of three-dimensional (3D) plant structures from limited-view images has become increasingly critical within plant sciences, impacting research fields such as digital plant phenotyping, crop breeding, ecological monitoring, and precision agriculture. Accurate 3D plant modelling allows researchers and practitioners to non-invasively analyse plant morphology [1], measure phenotypic traits, and monitor plant health under various environmental conditions, leading to greatly enhanced productivity, sustainability, and resilience in agriculture and ecology [2].

Despite its critical importance, the reliable reconstruction of realistic plant geometries from single or limited-view images remains particularly challenging [3]. Plants, characterised by intricate branching patterns, complex leaf arrangements, thin stems, and high degrees of self-occlusion, represent one of the most structurally challenging classes of objects to reconstruct accurately. Traditional 3D reconstruction approaches, which are originally designed for rigid, artificial objects such as furniture or vehicles, generally fail to handle the inherent structural complexities and irregularities of plants, frequently resulting in incomplete, distorted, or overly simplified geometries [4]. Consequently, specialised approaches tailored explicitly for plants are urgently needed.

1.1 Background

Recent advancements in generative artificial intelligence, particularly image-conditioned generative models, represent a fundamental shift, enabling reconstruction from only a single or few images. These methods leverage powerful deep-learning techniques, including diffusion models and structured latent spaces, to synthesise realistic 3D structures directly from limited-view imagery [5].

Notably, models like Hunyuan3D 2.0 [6] have introduced sophisticated diffusion pipelines that effectively generate detailed shapes and textures, while Trellis3D [7] leverages structured latent representations (SLAT) to achieve spatial coherence across diverse outputs. Additionally, rapid-generation models such as InstantMesh [8] and One2345++ [9] provide new avenues for real-time or interactive phenotyping applications, though often at the expense of geometric accuracy and detailed fidelity. Emerging approaches such as Unique3D [10] and Direct3D [11] further aim to refine reconstruction consistency and inference speed, yet their suitability and performance in plant-specific contexts require careful evaluation.

Although these generative methods have shown strong performance on standard computer vision benchmarks, results on non-plant datasets do not indicate how well they transfer to plants. Existing evaluations largely rely on rigid or synthetic objects, which lack the intricate branching, thin leaves, and self-occlusions characteristic of real plants. As a result, it remains unclear whether current methods can reliably reconstruct realistic plant structures, highlighting the need for systematic evaluation in this domain [4].

1.2 Motivation and Aims of This Study

Given the limitations highlighted above, there is an urgent need for comparative evaluations of advanced generative models explicitly focused on realistic plant reconstruction, specifically 3D mesh creation for plants. Most existing evaluations in the literature have primarily targeted rigid, synthetic, or simplified geometries, leaving significant knowledge gaps regarding performance and applicability in botanical scenarios.

The present study aims to directly address this research gap by rigorously evaluating six advanced image-conditioned generative models using the recently developed PlantDreamer dataset [12], which contains high-quality point clouds and corresponding single-view images of diverse plant species (Bean, Kale, Mint). We do this by investigating each method’s capabilities and limitations in accurately capturing the intricate geometries and complex textures inherent in real bean plant structures, as shown in Fig. 1.

The key research objectives are as follows:

1. Conduct a detailed comparative analysis of the selected generative methods (Hunyuan3D 2.0, Trellis3D, One2345++, InstantMesh, Unique3D, and Direct3D) specifically for realistic plant reconstruction.
2. Quantify reconstruction accuracy, visual realism, and semantic consistency using well-established evaluation metrics (Chamfer Distance, Normal Consistency, IoU, F-Score, PSNR, LPIPS, and adapted CLIP Score).

3. Identify practical strengths and limitations of each generative method to inform future development and real-world plant phenotyping applications.

The findings from this study provide critical insights into the capabilities and limitations of existing 3D generative models in realistic plant reconstruction. They also serve as an essential reference for selecting appropriate methods based on application-specific priorities, such as high accuracy, computational efficiency, or real-time performance. Ultimately, the results contribute directly to advancing digital plant phenotyping technologies and pave the way for future methodological innovations tailored explicitly for botanical research and precision agriculture.

The remainder of the paper is structured as follows: Chapter 2 reviews relevant literature, focusing on existing image-conditioned generative techniques, native 3D generative approaches, and plant-specific datasets. Chapter 3 describes the detailed methodology of this study, including dataset characteristics, experimental workflow, and evaluation metrics. Chapter 4 presents the experimental results, both quantitatively and qualitatively. In Chapter 5, this paper discusses the implications of the findings in the context of existing plant reconstruction research, highlighting methodological limitations and practical recommendations. Finally, Chapter 6 summarises the main conclusions and identifies promising future research directions within plant-specific 3D reconstruction methodologies.

2 Related Work

2.1 3D Plant Reconstruction in Plant Sciences

In plant science research, accurately capturing plant structure in three dimensions (3D) plays a crucial role in digital phenotyping, crop monitoring, plant growth analysis, and ecological research. Traditional manual measurement methods for plant traits are labour-intensive, subjective, and often destructive [13]. Thus, automated and non-invasive methods of 3D reconstruction from imaging data are gaining increasing importance.

Early plant 3D reconstruction approaches relied on multi-view stereo photogrammetry or laser scanning techniques (e.g., terrestrial laser scanning or LiDAR [4]). These methods produce highly accurate and detailed 3D representations but typically require extensive equipment setup, controlled conditions, multiple viewpoints, and considerable manual processing [13]. More importantly, the complex structural characteristics of plants—such as thin, delicate leaves, overlapping branches, and severe occlusions—often present challenges for traditional photogrammetric or laser-based techniques, frequently resulting in incomplete or distorted reconstructions. Any multi-view or scanning approach can also be hampered by the plant moving in-between captures, which would be a common problem in an outside environment.

Recent developments in computer vision, especially deep-learning-based generative models, have shown significant potential to address these challenges (for example, RGB cameras are more cost-effective and easier to access [14]). Image-conditioned generative methods that use powerful deep neural networks trained on large-scale

datasets promise automated, scalable, and accurate reconstructions directly from limited or even *single-view* imagery, offering new opportunities for precision agriculture and automated phenotyping applications.

2.2 Image-conditioned 3D Generation Methods via 2D Priors

Image-conditioned 3D generation methods via 2D priors leverage strong prior knowledge learned from large-scale image datasets to predict accurate and realistic 3D structures from limited-view images. Early approaches typically employed explicit geometric representations such as voxels [14] or point clouds [15]. However, recent advances have favoured implicit neural representations and diffusion-based generative approaches due to their superior detail fidelity and flexibility [16].

DreamFusion [17] marked a significant advancement in this area by introducing Score Distillation Sampling (SDS) [17], a technique where a pre-trained text-to-image diffusion model is used to optimise a 3D representation so that its renders are scored highly by the diffusion model for a given prompt. In this case, the 3D representation is a Neural Radiance Field (NeRF) [18], an implicit model that maps 3D coordinates and viewing directions to colour and density, allowing objects or scenes to be represented as continuous volumetric fields that can be rendered from novel views. While originally demonstrated in text-to-3D tasks, the underlying principles of SDS are applicable to single-image-to-3D reconstruction, enabling the generation of detailed textures and realistic geometries.

Zero123 [19] further improved multi-view consistency by synthesising additional viewpoints from a single input image, enhancing reconstruction accuracy across multiple perspectives. Similarly, One-2-3-45 [20] and its improved variant One2345++ [9] adopted a two-step pipeline—first synthesising multiple views, then performing implicit surface reconstruction via diffusion methods—to provide rapid, though sometimes less precise, reconstructions suitable for interactive applications. However, evaluations specifically targeting plant reconstruction scenarios remain limited, and performance typically declines with increasing geometric complexity, such as dense foliage or intricate branching patterns.

2.3 Native 3D Generative Methods

In contrast to image-conditioned models that heavily rely on 2D priors, native 3D generative methods directly operate in a dedicated 3D representation space. Recently developed models, including Hunyuan3D 2.0 [6] and Trellis3D [7], represent important advancements in native 3D generative approaches, demonstrating improved reconstruction quality and versatility.

Hunyuan3D 2.0 employs a sophisticated two-stage pipeline involving diffusion transformer-based geometry generation and subsequent high-resolution texture synthesis via diffusion painting. This approach has proven particularly effective for producing detailed and realistic reconstructions of complex geometries and textures, showing potential for intricate botanical structures.

Trellis3D introduced Structured 3D Latents (SLAT), enabling flexible decoding into multiple 3D output formats (meshes, NeRFs, Gaussian splats [16]). Its structured

latent representation significantly improves spatial coherence and flexibility, offering advantages in diverse plant modelling contexts.

Rapid generation methods, such as InstantMesh [8], achieve remarkable inference speeds through differentiable iso-surface extraction from learned implicit representations. These methods are especially promising for real-time interactive phenotyping applications, though often at the cost of geometric precision and fidelity.

Newer native methods, such as Unique3D [10] and Direct3D [11], further refine implicit-based reconstructions and direct mesh generation methods, respectively, targeting improved multi-view consistency and computational efficiency. However, comprehensive plant-focused evaluations for these emerging methods remain scarce, motivating the necessity of rigorous comparative assessments specifically tailored to realistic plant scenarios.

2.4 Available Plant-Specific Datasets

The availability of suitable datasets remains a critical limitation for realistic plant modelling research. Traditional 3D datasets, such as ShapeNet [21] or ModelNet [22], primarily comprise synthetic objects or rigid man-made structures, failing to represent plant complexity accurately. Datasets specifically developed for plant phenotyping, such as PhenoBench [23] and CVPPP datasets [24], often focus on segmentation or classification tasks rather than full-plant 3D reconstruction.

Recently, specialised datasets such as PlantNet [25] and PlantDreamer [12] have emerged, offering accurate and precise scans of real plants accompanied by corresponding image data. Notably, the PlantDreamer dataset, developed at the University of Nottingham, provides carefully acquired and controlled point-cloud scans paired with single-view images of various plant species, including Beans, Kale, and Mint. This dataset represents a significant advancement, enabling precise quantitative and qualitative assessments of 3D generative models for plant reconstruction tasks.

3 Methods

3.1 Aim, Design, and Experimental Workflow

The primary aim of this study was to systematically evaluate the performance of several state-of-the-art image-conditioned 3D generative models specifically in the context of realistic plant reconstruction from single-view images. To achieve this aim, this research conducted a comparative analysis involving six advanced generative models: Hunyuan3D 2.0 [6], Trellis3D (Structured 3D Latents, SLAT) [7], One2345++ [9], InstantMesh [8], Unique3D [10], and Direct3D [11]. These models were selected based on their prominence, diversity of underlying methodologies, and practical suitability for plant reconstruction tasks.

The experimental workflow consisted of four main stages shown as Fig 2: ground truth model and image preparation, generative model inference, model pre-processing and alignment, and quantitative and qualitative evaluations. Each stage will be introduced in the following sections of this chapter.

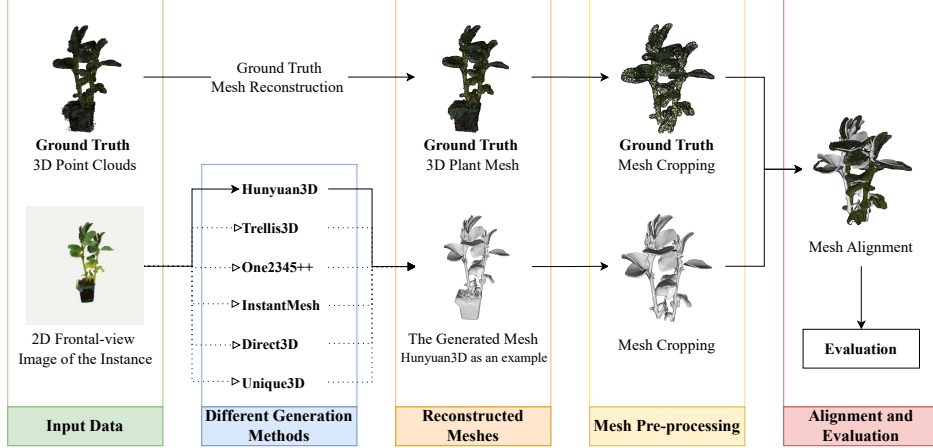


Fig. 2: Experimental pipeline for 3D plant model generation and evaluation. The top path represents the Ground Truth data. The original 3D plant point cloud is first reconstructed into a mesh representation, followed by standard pre-processing steps (cropping and alignment) to provide a consistent reference for evaluation. The bottom path illustrates the generation pipeline. A 2D frontal-view image of the same plant instance is given as input to different 3D generation methods. These methods output 3D plant meshes, which are subsequently reconstructed, cropped, and aligned to ensure comparability. Finally, the generated meshes are quantitatively and qualitatively evaluated against the processed Ground Truth mesh. For clarity, the figure shows Hunyuan3D as an example in the evaluation stage.

3.2 Dataset Characteristics and Plant Material

This study used the PlantDreamer [12] dataset, a specialised dataset recently developed by researchers at the University of Nottingham, explicitly aimed at facilitating rigorous quantitative and qualitative assessments of 3D plant reconstruction methods. The PlantDreamer dataset includes accurate 3D scans provided as dense point-cloud models, complemented by extensive multi-view RGB image data. Specifically, for each individual plant instance, the dataset provides over one hundred RGB images captured from multiple viewpoints around the plant, along with corresponding segmented versions of these images (with the background removed), and specially processed images where the base under the pot is digitally painted out.

In this experimental protocol, ten individual Bean plant instances from the dataset were arbitrarily selected as shown in Fig 3, considering their clear structural complexity, like clear stems and leaves, and intermediate level of geometric detail, which provided a suitable basis for systematic comparison. For each Bean plant instance, these experiments selected the single-view RGB image captured from the frontal viewpoint of the plant as the standard input for subsequent 3D reconstruction processes. Since some methods (Trellis3D and InstantMesh) to be evaluated lacked robust, integrated image-segmentation or background-removal capabilities, for these two methods,



Fig. 3: 10 individual Bean plant instances from the PlantDreamer dataset are selected as Ground Truth data, and their corresponding 2D frontal-view images are input data for 3D mesh reconstruction process.

this research substituted the original RGB images with corresponding segmented images (background removed) to ensure optimal reconstruction results and fair comparisons across models, as shown in Fig. 1. To get the fairest results of evaluation of different methods, the final evaluation process is conducted after cropping pots and bases of the meshes, and only generated meshes of plants themselves are kept for comparison as shown in Fig. 2.

To prepare ground-truth reference models for accurate evaluation, the original 3D point-cloud data provided by PlantDreamer was converted into precise ground-truth mesh models using the Alpha Shapes reconstruction algorithm [26]. This method effectively preserves detailed morphological features of the plants, ensuring an accurate representation suitable for rigorous comparative analyses. Other plant models, such as 6 mint models and 3 kale models from PlantDreamer dataset, are also generated for qualitative evaluation.

3.3 Evaluated Generative Methods and Computational Platforms

This research evaluated six state-of-the-art generative methods, each chosen for its innovative approach and distinct strengths in reconstructing 3D geometry and textures from single-view images.

Hunyuan3D 2.0 [6] employs a sophisticated two-stage generative pipeline. Initially, geometry is generated through a diffusion transformer network specifically optimised for capturing detailed structural information. Subsequently, a high-resolution diffusion-based painting stage produces realistic surface textures, enhancing visual realism and accuracy. Hunyuan3D 2.0 is particularly noted for its detailed output quality, making it highly promising for reconstructing intricate plant geometries.

Trellis3D [7] introduces Structured 3D Latents (SLAT), leveraging a unified sparse structured latent representation capable of decoding into multiple versatile output formats, including meshes, neural radiance fields (NeRFs), and Gaussian splats. Trellis3D’s latent structure supports consistent spatial reconstruction, positioning it as a versatile choice for complex botanical structures.

One2345++ [9] uses an implicit neural representation based on signed distance functions (SDFs), optimized for rapid, single-step inference. Despite its high computational efficiency, preliminary evaluations have indicated limitations in accurately reconstructing complex plant geometries, especially those with intricate branching or high levels of occlusion.

InstantMesh [8] represents an ultra-fast reconstruction method emphasising real-time interactive performance. It combines differentiable iso-surface extraction techniques with learned implicit representations, resulting in rapid generation suitable for immediate visualisation. However, previous evaluations [7] have noted reduced geometric precision compared to slower, more computationally intensive methods.

In addition, this paper included preliminary evaluations of two newly proposed methods, Unique3D [10] and Direct3D [11]. Unique3D employs advanced implicit neural fields designed specifically to enhance multi-view consistency and spatial accuracy. Direct3D adopts a direct mesh inference approach optimised explicitly for computational speed, robustness, and efficient geometry extraction. These two models were evaluated to identify their potential contributions and suitability for plant-specific reconstruction tasks, complementing the assessment of previously established generative techniques.

Due to the high computational resource demands, particularly GPU memory requirements, associated with contemporary 3D generative models, this research leveraged cloud-based demo platforms publicly available online for conducting these experiments. These platforms allowed direct use of pre-deployed models, thus eliminating the need for extensive local computational infrastructure. For Hunyuan3D 2.0, Tencent’s official Hunyuan3D demo platform was used, accessible through the Tencent Hunyuan3D website [27]. Trellis3D reconstructions were performed via the publicly-accessible Hugging Face Space provided by the Trellis community [28]. The One2345++ method was accessed through the SudoAI platform [29]. InstantMesh reconstructions were generated using the official InstantMesh Hugging Face Space provided by TencentARC [30]. For Direct3D evaluations, the Neural4D official online

studio was used [31], and Unique3D reconstructions were conducted through the AIUni platform [32].

Each Bean plant instance image (only 1 original RGB or segmented image, depending on the method) was individually uploaded and processed on the appropriate computational platform to produce corresponding reconstructed 3D meshes. The models generated from these platforms were then downloaded in standard mesh formats (typically GLB) for subsequent local processing and analysis.

3.4 Pre-processing and Alignment of Generated and Ground-truth Models

Upon generating the 3D models from the single-view input images, additional pre-processing steps were undertaken to ensure fair and accurate evaluation comparisons with the ground-truth meshes. Specifically, the pot and base components of both ground-truth and generated meshes were cropped out, leaving only the plant structures themselves. This cropping step was essential to eliminate extraneous variables, such as pot geometry and base supports, which are unrelated to the plant reconstruction quality assessment.

A crucial pre-processing step performed prior to the quantitative evaluation involved the accurate spatial alignment of each pair of corresponding generated and ground-truth meshes. Non-alignment would mean the metrics used to calculate the distance between the meshes would be unreliable. Model alignment is an inherently challenging task in 3D reconstruction studies, particularly when the structural similarity between the generated and ground-truth models is moderate or low. Due to these challenges, this research employed a semi-automated alignment procedure involving initial approximate alignment based on automatically computed centroids and bounding-box orientations for scaling, followed by careful manual refinement and the Iterative Closest Point (ICP) algorithm [33]. ICP is a widely used algorithm for rigid registration, which iteratively refines the alignment of two 3D shapes by minimising the distance between corresponding points on their surfaces. This combination was adopted to achieve as precise spatial correspondence as possible for each pair of meshes. The details, challenges, and implications of the alignment procedure are discussed further in the Discussion chapter 5.4 of this paper, highlighting alignment as an important topic for future methodological development.

3.5 Quantitative Evaluation Metrics

To systematically evaluate and compare model performance, here we selected established quantitative metrics commonly employed in 3D reconstruction research, ensuring comprehensive coverage of different facets of what makes a model "good": geometric accuracy, visual fidelity, and semantic consistency.

For *geometric accuracy* evaluation, we computed the Chamfer Distance, Normal Consistency, IoU and F-Score. Chamfer Distance [34] measures the average nearest-neighbour distance between points on the generated mesh and ground truth, with lower values indicating better geometric accuracy. Normal Consistency [35] quantifies how accurately generated surface normals align with those of the ground truth,

where higher values indicate better reconstruction quality. The F-Score [36] combines completeness and precision of reconstructed meshes based on a predefined threshold distance (set at 1% of the bounding box diagonal), with higher scores signifying more accurate and complete meshes.

To assess *visual fidelity*, we used Peak Signal-to-Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS). PSNR [37] provides a pixel-level similarity measure between rendered images of the generated and ground-truth meshes; higher PSNR values reflect superior visual similarity. LPIPS [38] quantifies perceptual visual similarity from the perspective of human observers, with lower scores indicating higher perceived realism and image quality.

To measure *semantic consistency* between generated and ground-truth renderings, we used an adapted form of the Contrastive Language-Image Pre-training (CLIP) Score [39]. Unlike the conventional CLIP Score typically used for evaluating similarity between images and textual captions, this research modified this metric to compute the cosine similarity between image embeddings directly derived from the ground-truth and generated model renderings. This adapted CLIP Score thus quantified the semantic alignment and perceptual similarity specifically between two corresponding images.

3.6 Statistical Analysis and Computational Details

Quantitative metrics were calculated individually for each Bean plant instance across all six evaluated generative methods. We then computed mean values and standard deviations for each metric across the entire set of ten Bean plant samples per model.

All mesh cropping, scaling, alignment, rendering, and metric computations were conducted locally using standardised computational packages and Python libraries (e.g., Open3D V0.19.0, Trimesh V4.6.4, NumPy V2.1.3, SciPy V1.15.2). Mesh rendering was performed using consistent camera viewpoints and standardised lighting conditions to ensure comparability. Finally, all qualitative analyses and graphical visualisations were generated in Python, ensuring a clear and reproducible presentation of experimental results.

4 Results

4.1 Quantitative Evaluation Results

The quantitative evaluation results for the six advanced image-conditioned generative models—Hunyuan3D 2.0, Trellis3D (Structured 3D Latents, SLAT), One2345++, InstantMesh, Direct3D, and Unique3D—are summarised in Tables 1 and 2, based on ten Bean plant instances from the PlantDreamer dataset. Geometry-focused metrics (Table 1) capture surface accuracy and volumetric completeness, while appearance- and semantics-oriented measures (Table 2) reflect visual realism and how well the generated meshes correspond to the conditioning images. Standard deviations are also reported to indicate the stability of each method across plant instances.

In addition to the mean and standard deviation values reported in Table 1 and 2, Fig. 4 illustrates the full distribution of Chamfer Distance and F-Score across 10 Bean

Table 1: Result table of **geometry metrics** for all methods. Mean of 10 instances, St. Dev. in brackets. CD for Champfer Distance, NC for Normal Consistency and IoU for Intersection over Union. The highest scores for each metric are in bold and the second-highest are in italic.

Methods	Geometry Metrics			
	CD ↓	NC ↑	IoU ↑	F-Score ↑
Hunyuan3D	0.00515 (0.000934)	<i>0.3691 (0.0511)</i>	0.07212 (0.0322)	0.8597 (0.0508)
Trellis	0.00626 (0.000877)	0.3351 (0.0254)	0.06534 (0.0249)	0.7920 (0.0506)
One2345++	0.00593 (0.001170)	0.3481 (0.0674)	0.04785 (0.0116)	0.8059 (0.0683)
InstantMesh	0.00795 (0.002250)	0.3715 (0.0980)	0.04181 (0.0183)	0.7040 (0.1144)
Direct3D	<i>0.00562 (0.001104)</i>	0.3646 (0.0562)	<i>0.07074 (0.0188)</i>	<i>0.8236 (0.0706)</i>
Unique3D	0.00603 (0.001337)	0.3228 (0.0361)	0.06086 (0.0256)	0.8129 (0.0626)

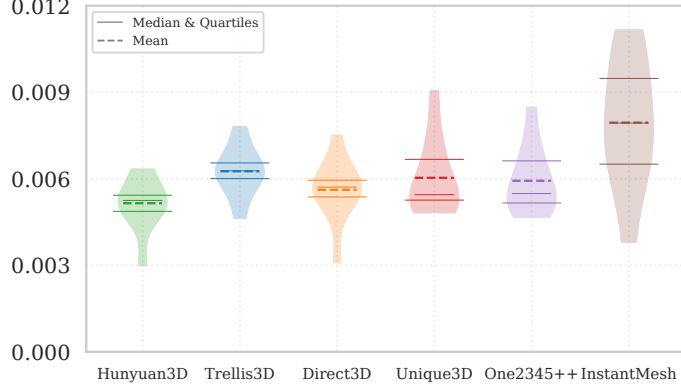
Table 2: Result table of Visual Fidelity and Semantic Consistency for all methods. Mean of 10 instances, St. Dev. in brackets. The highest scores for each metric are in bold and the second-highest are in italic.

Methods	Visual Fidelity		Semantic Consistency
	PSNR ↑	LPIPS ↓	CLIP ↑
Hunyuan3D	6.4972 (0.3048)	0.4469 (0.0618)	0.8564 (0.0213)
Trellis	6.7710 (0.4638)	0.4722 (0.0536)	0.7886 (0.0577)
One2345++	<i>6.5442 (0.2289)</i>	<i>0.4409 (0.0736)</i>	0.8288 (0.0567)
InstantMesh	6.4478 (0.2924)	0.4768 (0.0593)	0.7394 (0.0704)
Direct3D	6.5166 (0.3039)	0.4359 (0.0655)	<i>0.8305 (0.0342)</i>
Unique3D	6.3589 (0.2927)	0.4531 (0.0674)	0.7869 (0.0695)

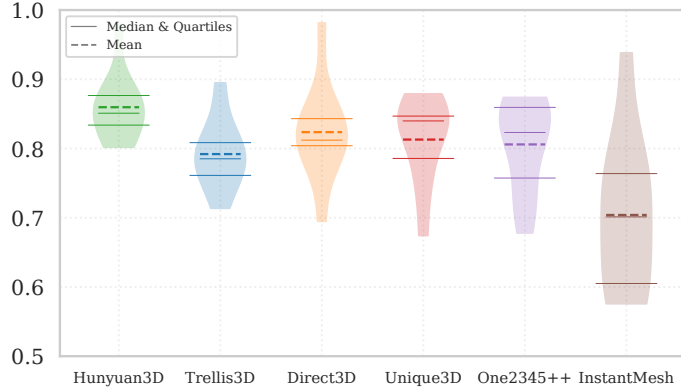
instances for each method. These violin plots highlight both the central tendency and variability, showing that Hunyuan3D consistently achieves the lowest Chamfer Distance and highest F-Score, whereas InstantMesh exhibits larger variance and generally weaker performance.

In terms of geometric accuracy, **Hunyuan3D** demonstrated the best overall performance, achieving the lowest Chamfer Distance (0.00515 ± 0.00093) and the highest F-Score (0.8597 ± 0.0508), indicating excellent overall geometric precision and completeness. Its Normal Consistency (0.3691 ± 0.0511) was slightly lower than InstantMesh, but still a strong second place among all evaluated methods. Additionally, Hunyuan3D attained the highest semantic alignment with the adapted CLIP Score (0.8564 ± 0.0213), underscoring its superior semantic consistency with the input image content.

Direct3D also exhibited strong geometric performance, achieving the second-lowest Chamfer Distance (0.00562 ± 0.00110), accompanied by robust Normal Consistency (0.3646 ± 0.0562) and IoU (0.07074 ± 0.0188), indicating its effectiveness in capturing spatial structures. Its semantic consistency (CLIP Score: $0.8305 \pm$



(a) Chamfer Distance (lower is better).



(b) F-Score (higher is better).

Fig. 4: Distribution of quantitative geometry metrics across methods. (a) Chamfer Distance (lower is better). (b) F-Score (higher is better). Each violin plot shows the distribution across 10 Bean instances, with the central dashed line indicating the mean and solid lines indicating the quartiles and median.

0.0342) and visual fidelity (LPIPS: 0.4359 ± 0.0655 , PSNR: 6.5166 ± 0.3039) also placed it among the top-performing methods in this evaluation.

Beyond Hunyuan3D 2.0 and Direct3D, the remaining four methods showed mixed and generally more limited performance. One2345++ achieved reasonable geometry overall but consistently struggled with volumetric completeness, leading to lower IoU scores. It nevertheless produced perceptually sharper renders, as indicated by its strong performance on LPIPS, making it visually convincing despite underlying structural errors. Trellis3D offered balanced but unspectacular results: while its geometric measures were only moderate, it stood out for delivering the highest PSNR values, suggesting that at the pixel level its renders aligned well with the ground truth, albeit



Fig. 5: Example images of resulting models. For each species, two example instances are shown with generated meshes by all six methods.

with reduced perceptual realism. InstantMesh was the fastest method evaluated and excelled in producing smooth, well-aligned surfaces (reflected in high normal consistency), but these came at the expense of accurate volumetric reconstruction and fine structural detail, limiting its usefulness for quantitative plant analysis. Finally, Unique3D produced coherent global plant shapes and acceptable geometry but tended to underperform in visual and semantic alignment compared to the stronger methods, placing it in the lower-middle tier of overall performance.

4.2 Qualitative Visual Comparisons

In addition to quantitative analysis, qualitative evaluations were carried out to assess visual performance and quality, and for this we use the other plants to show generalisation. Fig. 5 illustrates representative examples of generated plant meshes alongside the corresponding input images for bean plant instances and kale and mint instances, visually demonstrating the differences in reconstruction quality among the methods.

Qualitative comparisons confirmed Hunyuan3D’s superior capability in capturing intricate plant details, such as accurate leaf shapes, clear vein structures, and realistic branching patterns. The generated meshes from Hunyuan3D consistently exhibited close correspondence with the ground-truth geometry, preserving fine-scale details even in complex areas. Despite occasional minor inaccuracies in spatial alignment, Hunyuan3D outputs appeared highly realistic and faithful to the input images, especially for Beans and Kale.

Trellis3D’s generated meshes were visually coherent and captured the general morphology of the plant instances effectively. Nonetheless, detailed qualitative analysis revealed limitations in fine structural accuracy, like mint examples as shown in Fig. 5. Specifically, leaf edges and delicate branching structures were sometimes oversimplified or inadequately represented. While visually acceptable at moderate resolutions, close inspections revealed lower fidelity in intricate plant regions compared to Hunyuan3D.

Meshes generated by One2345++ displayed more deficiencies. Although certain generated models appeared visually acceptable from afar, closer scrutiny revealed substantial structural irregularities, including distorted branching structures, incomplete stem and leaf geometries, and significant spatial inaccuracies as shown in Fig. 5. These visual issues corresponded directly with quantitative findings, particularly lower IoU scores, indicating difficulties in reconstructing coherent 3D plant structures.

InstantMesh reconstructions presented the greatest degree of visual abstraction and simplification among all evaluated models. Generated meshes typically lacked detailed geometric features, such as accurate leaf shapes or branching structures, resulting in generalised forms and simplified geometries. However, InstantMesh excelled in generating visually smooth and coherent surface normals, albeit at the cost of significant detail loss, aligning closely with its quantitative performance characteristics.

Overall, qualitative evaluation reinforced the quantitative findings, highlighting Hunyuan3D as the most visually accurate and semantically consistent method. Trellis3D exhibited acceptable visual results but was limited in capturing fine structural details, while InstantMesh and One2345++ presented significant qualitative shortcomings, reflecting their lower quantitative performance metrics.

5 Discussions

5.1 Summary of Principal Findings

This study set out to determine how well contemporary image-conditioned 3D generative models recover realistic plant geometry from a single photograph, using a unified protocol on the PlantDreamer Bean subset, further verified by qualitative evaluation on mint and kale examples. Three conclusions emerge clearly from the results. First, Hunyuan3D delivered the strongest overall accuracy and semantic agreement with the input image, as reflected by the best Chamfer Distance and F-Score and the highest adapted CLIP score, while maintaining competitive surface normal agreement. Second, Direct3D formed a robust second tier, balancing strong geometry with good visual and semantic fidelity, suggesting that direct mesh-oriented pipelines can approach the accuracy of heavier diffusion-based systems. Third, methods designed primarily for speed or single-step inference—InstantMesh and, to a lesser extent,

One-2-3-45++—sacrificed geometric completeness and volumetric overlap, even when their surfaces appeared smooth or their perceptual scores were acceptable. Trellis3D occupied a middle ground: its PSNR was highest, indicating faithful view-wise appearance under this rendering protocol, but its geometry and perceptual similarity lagged behind the best models.

5.2 Model-specific Behaviour in the Context of Plant Morphology

The quantitative and qualitative analyses aligned with the structural demands of botanical targets. Hunyuan3D’s two-stage pipeline, separating shape diffusion from texture synthesis, consistently preserved leaf outlines, stem thickness, and overall leaf arrangement, and produced meshes that visually tracked the input photograph more closely than the other systems. Nevertheless, the method exhibited occasional spatial inconsistencies when confronted with dense, self-occluding foliage of the sort this study observes more often in Mint than in Bean; this manifested as local misplacements of leaf surfaces or slight branch drift during alignment, which in turn depressed normal consistency and IoU in those cases.

Trellis3D benefited from its structured latent space in terms of global spatial coherence: reconstructions were frequently “complete plants” with plausible topology, even when fine details were softened. The method’s top PSNR suggests its renders align well at the pixel level with the ground-truth views under a fixed camera, yet its higher LPIPS implies that high-frequency perceptual details and textural micro-structure are not as faithfully captured as by Hunyuan3D or Direct3D.

One-2-3-45++ achieved reasonable Chamfer and perceptual scores but struggled most with intricate structures. Qualitatively, leaf clusters often merged, and thin elements were sometimes under-reconstructed, leading to low IoU and a relatively variable F-Score. These issues are consistent with a single-step implicit pipeline that does not explicitly regularise for thin, highly occluded botanical parts.

InstantMesh excelled in surface smoothness and produced the highest normal consistency on average, yet its geometric overlap and completeness were the weakest. This combination—good normals with poor IoU and F-Score—reflects simplified, smoothed surfaces that lack the fine, articulated geometry critical for plants. In practice, the method’s speed is compelling for interactive scenarios, but its reconstructions under-represent thin leaves and inter-leaf gaps.

5.3 Interpreting Metric Disagreements on Botanical Targets

Several metric pairings illuminate plant-specific behaviour. The simultaneous lead of Trellis3D on PSNR versus its weaker LPIPS indicates that pixel-wise similarity under this renderer does not guarantee human-perceived fidelity for foliage textures; LPIPS, being feature-based, penalises the loss of high-frequency fine patterns and edge details that PSNR overlooks. InstantMesh’s strong normal consistency with poor IoU reiterates that smooth, well-oriented surfaces can still miss volumetric occupancy in thin regions. Across methods, absolute IoU values were low compared to typical rigid-object benchmarks. This is expected: IoU is sensitive to voxelisation and penalises thin

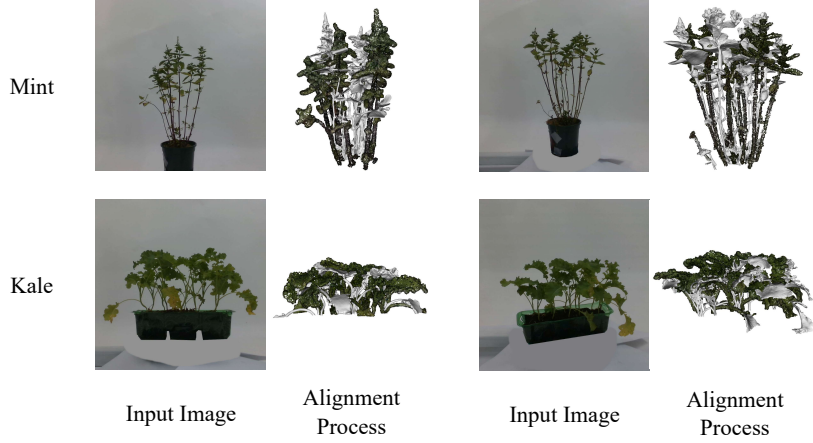


Fig. 6: Example images of the alignment process for a generated mesh and corresponding ground truth.

structures disproportionately; a leaf that is one to two voxels thick in a discretised volume is easy to miss even when its surface is well aligned. These observations argue for plant-aware evaluation that complements surface metrics (Chamfer, normal consistency, F-Score) with structure-aware measures such as skeleton similarity, branch graph edit distance, or leaf-level detection/segmentation accuracy in future studies.

The adapted CLIP score, computed as cosine similarity between image embeddings of generated and ground-truth renders, correlated well with the subjective sense of “does this look like the same plant from this view?”. Unlike the usual text–image use of CLIP, the image–image variant captures semantic and perceptual agreement without privileging pixel identity. Hunyuan3D’s lead on this metric mirrors its qualitative alignment to the input photograph and supports using adapted CLIP as a practical semantic proxy when botanical annotations are scarce.

5.4 Effects of Pre-processing and Alignment

Accurate evaluation of plant reconstructions hinges not only on the generative model itself but also on the quality of input preparation and output alignment. Input image preparation proved consequential. When models lacked robust internal matting, feeding segmented images (background removed) to the models reduced spurious conditioning from the pot, base, and studio backdrop, and produced measurably cleaner geometry around the crown and lower stem. Pre-processed images in which the base region was painted out also helped stabilise methods sensitive to background edges. On the target side, cropping pots and bases from both generated and ground-truth meshes before scoring eliminated confounds and ensured that scores reflected plant reconstruction rather than the fidelity of ancillary objects.

Mesh alignment was a non-trivial source of variance. Semi-automatic alignment, followed by manual refinement, was adequate for most Hunyuan3D and Direct3D outputs, which tended to be close to the reference pose. Conversely, methods with larger geometric deviations, notably InstantMesh and some One-2-3-45++ reconstructions, required substantial manual intervention. Trellis3D and Unique3D generally aligned well at the global scale but often exhibited moderate local mismatches, as shown in Fig. 6, necessitating occasional manual refinements, though to a lesser degree than InstantMesh or One-2-3-45++. Because Chamfer, IoU, and F-Score are all sensitive to residual misalignment, improving registration robustness—e.g., combining global feature-based initialisation with robust ICP variants that down-weight outliers—would further reduce evaluation noise, particularly for thin structures whose local correspondences are fragile. The need for manual refinement also underscores a practical point for plant scientists: deploying single-image 3D generation in phenotyping pipelines will require either reliable automatic alignment or direct metric learning on view-rendered images that is less pose-sensitive.

Overall, these findings highlight that both input pre-processing and alignment accuracy substantially affect reconstruction quality and metric reliability. The reliance on semi-manual alignment currently limits scalability, underscoring the need for robust automatic registration methods tailored to botanical geometry, or alternatively, evaluation protocols based on view-rendered images that are inherently less pose-dependent.

5.5 Implications for Plant Science Applications

From the perspective of digital phenotyping and botany, method choice should align with the downstream task. Where morphological measurements demand accurate leaf boundaries, petiole thickness, and plausible global topology—e.g., leaf count, leaf area estimation, or branching order—Hunyuan3D is the most suitable among the tested models, with Direct3D a close alternative when runtime or deployment simplicity is important. For interactive visualisation, rapid prototyping, or educational AR applications where speed dominates and fine-scale geometry is secondary, InstantMesh provides a practical baseline, albeit with the recognised limitations in completeness. Trellis3D’s spatial coherence suggests potential as a backbone for pipelines that will subsequently refine details. Unique3D’s coherent global structure but lower semantic alignment indicates it could benefit from tighter image-conditioned guidance or plant-specific priors.

5.6 Limitations and Opportunities for Methodological Improvement

Three limitations qualify the present conclusions. First, the main analysis was conducted on ten Bean instances; while beans capture relevant thin-structure challenges, broader botanical diversity—particularly denser crowns such as Mint and more heterogeneous phyllotaxy such as Kale—will stress generalisation. Second, the reliance on public cloud implementations introduces unknowns: some services may apply undocumented defaults, lightweight finetunes where the models could have fewer parameters,

or mesh post-processing to improve stability. Although the inputs and used platforms were standardised, exact parity with open-source checkpoints cannot be guaranteed. Third, the alignment protocol, while careful, still required manual refinement in difficult cases and remains a source of variance for thin structures. Finally, the absolute IoU values should be interpreted cautiously given the known bias of voxel occupancy toward thick parts; the inclusion of surface-based metrics mitigates but does not eliminate this effect.

The error modes observed point to concrete research directions. Incorporating plant-aware priors—such as branch skeleton constraints, leaf blade thickness regularisation, and self-occlusion reasoning—into shape diffusion or direct mesh decoders could improve reconstruction in dense crowns. A hybrid strategy that pairs a spatially coherent backbone (e.g., SLAT-style latents) with a detail-oriented refinement module (e.g., diffusion upsampler) may combine Trellis3D’s completeness with Hunyuan3D’s detail. On the data side, multi-view conditioning—even with synthetic views generated by robust view-diffusion—should reduce the “Janus” artefacts observed in single-image pipelines. Finally, evaluation can be strengthened by adding structure-aware metrics (skeleton overlap, leaf instance matching, branch angle distributions) and by adopting alignment-invariant, render-based measures that complement surface distances. Overall, it is worth noting that plant reconstruction is a particularly challenging task for these methods, and although some do not perform well here, they may perform very well on objects with different topologies and appearances.

In summary, single-image 3D plant reconstruction has reached a point where high-fidelity geometry is achievable for moderately complex specimens under controlled input preparation, with Hunyuan3D and Direct3D currently offering the best accuracy–stability trade-off. At the same time, systematically addressing thin-structure completeness, occlusion reasoning, and alignment sensitivity will be key to translating these advances into robust tools for large-scale plant phenotyping and allied applications.

6 Conclusions

This study systematically evaluated six state-of-the-art single-image 3D generative models on the PlantDreamer dataset under a unified protocol. By comparing methods that span diffusion pipelines, structured 3D latents, implicit decoders, and direct mesh inference, we assessed their ability to reconstruct realistic plant structures from photographs.

Three main conclusions emerged. Hunyuan3D achieved the most accurate and semantically faithful reconstructions across geometry and semantic metrics. Direct3D provided a strong balance of accuracy and efficiency, showing that lighter pipelines can rival diffusion-based systems. Faster approaches such as InstantMesh and One-2-3-45++ delivered smoother but less complete geometry, while Trellis3D and Unique3D produced coherent global structures but missed fine foliage detail. Qualitative inspection reinforced these findings and highlighted the trade-offs between accuracy, efficiency, and detail across current approaches.

For plant-science applications, Hunyuan3D is the most reliable choice when accurate morphology is required, while Direct3D offers a practical alternative when efficiency matters. InstantMesh may still be suitable for interactive or educational uses, provided its limitations are acknowledged. The limitations of this study point to future needs: incorporating plant-aware priors, hybridising spatial and detail-oriented approaches, and developing structure-aware evaluation protocols.

In sum, single-image 3D plant generation is now capable of producing high-fidelity reconstructions for moderately complex specimens under controlled input preparation. Among current options, Hunyuan3D and Direct3D offer the best trade-offs for accuracy-driven plant analysis, while faster pipelines remain attractive for interactive uses. By codifying a plant-centric protocol, highlighting error modes tied to botanical structure, and identifying metrics that better reflect plant morphology, this work provides a practical baseline and a roadmap for turning image-conditioned 3D generation into a reliable tool for digital plant phenotyping, crop science, and ecological research.

Acknowledgements

Funding

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) with Zane Hartley as the Principal Investigator; and the University of Nottingham PhD studentship scheme. Zane Hartley was supported by the EPSRC Doctoral Prize Fellowship [EP/R513283/1]. Zihe Gao was funded by a University of Nottingham PhD studentship. The publication fee was covered by the University of Nottingham’s central open access funding for research derived from UKRI-funded grants. We thank our colleagues and supervisors for their helpful feedback during the research process.

Data availability

The data supporting this publication will be made available through the University of Nottingham Research Data Repository. A persistent identifier (DOI) will be included here upon acceptance: *[placeholder DOI – to be updated]*.

Declarations

Ethics approval and consent to participate

Not applicable.

Competing interests

The authors declare no competing interests.

References

- [1] Stuart, L.A.G., Wells, D.M., Atkinson, J.A., Castle-Green, S., Walker, J., Pound, M.P.: High-fidelity wheat plant reconstruction using 3D Gaussian splatting and neural radiance fields **14**, 022 <https://doi.org/10.1093/gigascience/giaf022>

- [2] Gibbs, J.A., Pound, M., French, A.P., Wells, D.M., Murchie, E., Pridmore, T.: Plant Phenotyping: An Active Vision Cell for Three-Dimensional Plant Shoot Reconstruction **178**(2), 524–534 <https://doi.org/10.1104/pp.18.00664>
- [3] Pound, M.P., French, A.P., Murchie, E.H., Pridmore, T.P.: Automated Recovery of Three-Dimensional Models of Plant Shoots from Multiple Color Images **166**(4), 1688–1698 <https://doi.org/10.1104/pp.114.248971>
- [4] Okura, F.: 3D modeling and reconstruction of plants and trees: A cross-cutting review across computer graphics, vision, and plant phenotyping **72**(1), 31–47 <https://doi.org/10.1270/jsbbs.21074> 36045890
- [5] Long, X., Guo, Y.-C., Lin, C., Liu, Y., Dou, Z., Liu, L., Ma, Y., Zhang, S.-H., Habermann, M., Theobalt, C., Wang, W.: Wonder3D: Single Image to 3D Using Cross-Domain Diffusion. <https://doi.org/10.48550/arXiv.2310.15008>
- [6] Zhao, Z., Lai, Z., Lin, Q., Zhao, Y., Liu, H., Yang, S., Feng, Y., Yang, M., Zhang, S., Yang, X., Shi, H., Liu, S., Wu, J., Lian, Y., Yang, F., Tang, R., He, Z., Wang, X., Liu, J., Zuo, X., Chen, Z., Lei, B., Weng, H., Xu, J., Zhu, Y., Liu, X., Xu, L., Hu, C., Yang, S., Zhang, S., Liu, Y., Huang, T., Wang, L., Zhang, J., Chen, M., Dong, L., Jia, Y., Cai, Y., Yu, J., Tang, Y., Zhang, H., Ye, Z., He, P., Wu, R., Zhang, C., Tan, Y., Xiao, J., Tao, Y., Zhu, J., Xue, J., Liu, K., Zhao, C., Wu, X., Hu, Z., Qin, L., Peng, J., Li, Z., Chen, M., Zhang, X., Niu, L., Wang, P., Wang, Y., Kuang, H., Fan, Z., Zheng, X., Zhuang, W., He, Y., Liu, T., Yang, Y., Wang, D., Liu, Y., Jiang, J., Huang, J., Guo, C.: Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. <https://doi.org/10.48550/arXiv.2501.12202>
- [7] Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., Yang, J.: Structured 3D Latents for Scalable and Versatile 3D Generation. <https://doi.org/10.48550/arXiv.2412.01506>
- [8] Xu, J., Cheng, W., Gao, Y., Wang, X., Gao, S., Shan, Y.: InstantMesh: Efficient 3D Mesh Generation from a Single Image with Sparse-view Large Reconstruction Models. <https://doi.org/10.48550/arXiv.2404.07191>
- [9] Liu, M., Shi, R., Chen, L., Zhang, Z., Xu, C., Wei, X., Chen, H., Zeng, C., Gu, J., Su, H.: One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation And 3D Diffusion. <https://doi.org/10.48550/arXiv.2311.07885>
- [10] Wu, K., Liu, F., Cai, Z., Yan, R., Wang, H., Hu, Y., Duan, Y., Ma, K.: Unique3D: High-Quality and Efficient 3D Mesh Generation from A Single Image. <https://doi.org/10.48550/arXiv.2405.20343>
- [11] Wu, S., Lin, Y., Zhang, F., Zeng, Y., Xu, J., Torr, P., Cao, X., Yao, Y.: Direct3D: Scalable Image-to-3D Generation Via 3D Latent Diffusion Transformer. <https://doi.org/10.48550/arXiv.2405.20343>

[//doi.org/10.48550/arXiv.2405.14832](https://doi.org/10.48550/arXiv.2405.14832)

- [12] Hartley, Z.K.J., Stuart, L.A.G., French, A.P., Pound, M.P.: PlantDreamer: Achieving Realistic 3D Plant Models with Diffusion-Guided Gaussian Splatting. <https://doi.org/10.48550/arXiv.2505.15528>
- [13] Murphy, K.M., Ludwig, E., Gutierrez, J., Gehan, M.A.: Deep Learning in Image-Based Plant Phenotyping **75**(1), 771–795 <https://doi.org/10.1146/annurev-arplant-070523-042828> 38382904
- [14] Hartley, Z.K.J., Jackson, A.S., Pound, M., French, A.P.: GANana: Unsupervised Domain Adaptation for Volumetric Regression of Fruit **2021** <https://doi.org/10.34133/2021/9874597>
- [15] Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. <https://doi.org/10.48550/arXiv.1901.05103>
- [16] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. <https://doi.org/10.48550/arXiv.2308.04079>
- [17] Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: DreamFusion: Text-to-3D Using 2D Diffusion. <https://doi.org/10.48550/arXiv.2209.14988>
- [18] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. <https://doi.org/10.48550/arXiv.2003.08934>
- [19] Liu, R., Wu, R., Hoorick, B.V., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot One Image to 3D Object. <https://doi.org/10.48550/arXiv.2303.11328>
- [20] Liu, M., Xu, C., Jin, H., Chen, L., T, M.V., Xu, Z., Su, H.: One-2-3-45: Any Single Image to 3D Mesh in 45 Seconds Without Per-Shape Optimization. <https://doi.org/10.48550/arXiv.2306.16928>
- [21] Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet: An Information-Rich 3D Model Repository. <https://doi.org/10.48550/arXiv.1512.03012>
- [22] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3D ShapeNets: A Deep Representation for Volumetric Shapes. <https://doi.org/10.48550/arXiv.1406.5670>
- [23] Weyler, J., Magistri, F., Marks, E., Chong, Y.L., Sodano, M., Roggiolani, G.,

- Chebrolu, N., Stachniss, C., Behley, J.: PhenoBench: A Large Dataset and Benchmarks for Semantic Image Interpretation in the Agricultural Domain **46**(12), 9583–9594 <https://doi.org/10.1109/TPAMI.2024.3419548>
- [24] Minervini, M., Fischbach, A., Scharr, H., Tsafaris, S.A.: Finely-grained annotated datasets for image-based plant phenotyping **81**, 80–89 <https://doi.org/10.1016/j.patrec.2015.10.013>
- [25] Garcin, C., Joly, A., Bonnet, P., Servajean, M., Salmon, J.: Pl@ntNet-300K Image Dataset. <https://doi.org/10.5281/zenodo.4726653>
- [26] Edelsbrunner, H., Kirkpatrick, D., Seidel, R.: On the shape of a set of points in the plane **29**(4), 551–559 <https://doi.org/10.1109/TIT.1983.1056714>
- [27] Tencent: Tencent Hunyuan3D. <https://3d.hunyuan.tencent.com>
- [28] TRELLIS - a Hugging Face Space by Trellis-Community. <https://huggingface.co/spaces/trellis-community/TRELLIS>
- [29] sudoAI. <https://www.sudo.ai/image-to-3d>
- [30] InstantMesh - a Hugging Face Space by TencentARC. <https://huggingface.co/spaces/TencentARC/InstantMesh>
- [31] Neural4D. <https://www.neural4d.com/studio/image-to-3d>
- [32] Aiuni. <https://aiuni.ai/>
- [33] Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes **14**(2), 239–256 <https://doi.org/10.1109/34.121791>
- [34] Wu, T., Pan, L., Zhang, J., Wang, T., Liu, Z., Lin, D.: Density-Aware Chamfer Distance as a Comprehensive Metric for Point Cloud Completion. <https://doi.org/10.48550/arXiv.2111.12702>
- [35] Yang, Z., Wang, P., Xu, W., Zhao, L., Nevatia, R.: Unsupervised Learning of Geometry with Edge-aware Depth-Normal Consistency. <https://doi.org/10.48550/arXiv.1711.03665>
- [36] Christen, P., Hand, D.J., Kirielle, N.: A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives **56**(3), 73–17324 <https://doi.org/10.1145/3606367>
- [37] Horé, A., Ziou, D.: Image Quality Metrics: PSNR vs. SSIM. In: 2010 20th International Conference on Pattern Recognition, pp. 2366–2369. <https://doi.org/10.1109/ICPR.2010.579>
- [38] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The Unreasonable

Effectiveness of Deep Features as a Perceptual Metric. <https://doi.org/10.48550/arXiv.1801.03924>

- [39] Hessel, J., Holtzman, A., Forbes, M., Bras, R.L., Choi, Y.: CLIPScore: A Reference-free Evaluation Metric for Image Captioning. <https://doi.org/10.48550/arXiv.2104.08718>