CropCraft: Complete Structural Characterization of Crop Plants From Images

Albert J. Zhai[1] Xinlei Wang[1] Kaiyuan Li[1] Zhao Jiang[1] Junxiong Zhou[2]
Sheng Wang[1] Zhenong Jin[2] Kaiyu Guan[1] Shenlong Wang[1]
[1] University of Illinois Urbana-Champaign [2] University of Minnesota Twin Cities

Abstract

The ability to automatically build 3D digital twins of plants from images has countless applications in agriculture, environmental science, robotics, and other fields. However, current 3D reconstruction methods fail to recover complete shapes of plants due to heavy occlusion and complex geometries. In this work, we present a novel method for 3D modeling of agricultural crops based on optimizing a parametric model of plant morphology via inverse procedural modeling. Our method first estimates depth maps by fitting a neural radiance field and then optimizes a specialized loss to estimate morphological parameters that result in consistent depth renderings. The resulting 3D model is complete and biologically plausible. We validate our method on a dataset of real images of agricultural fields, and demonstrate that the reconstructed canopies can be used for a variety of monitoring and simulation applications. Project page: https://ajzhai.github.io/CropCraft

1. Introduction

Plants are ubiquitous objects that appear all around the world and serve as the foundation for agriculture, which underpins our civilization's growth and survival. The ability to automatically build 3D digital twins of plants from images has countless applications in agriculture, environmental science, robotics, and other fields. In particular, the development of such methods in the context of agriculture will enable automatic, large-scale monitoring of crops. The collected data can provide decision support for farmers, aid carbon budgeting for decision-makers, support the development of new agricultural techniques, and inform the design of new genotypes [10, 48, 66]. All of these advances will contribute to increasing crop productivity, alleviating the rising food crisis of today's world [4, 36].

However, 3D reconstruction of plants remains to be a challenging vision problem. Many plants, including most of those found in agriculture, are composed of complex arrangements of thin leaves and branches that heavily occlude one another. These low visibility conditions cause existing reconstruction pipelines to fall short. Multi-view reconstruction methods such as those based on neural rendering [27, 34, 53, 57, 62] or multi-view stereo [46, 58, 61, 65] do not reconstruct the invisible regions of the scene, leading to incomplete plant shapes and extraneous geometry. Learning-based methods [13, 17, 29] can overcome this issue but require large amounts of ground-truth 3D data for training, which is extremely difficult to collect for dense vegetation due to the same occlusion reasons.

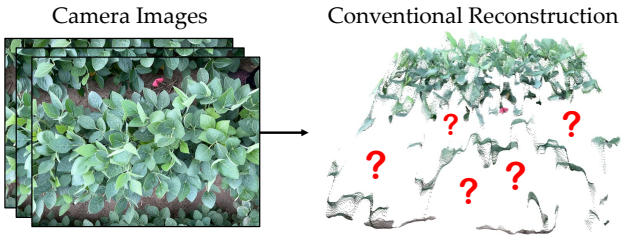

Figure 1. **Occlusion makes complete plant reconstruction challenging.** Current 3D reconstruction methods (e.g. VGGT [56]) miss a large portion of the crop canopy due to heavy occlusion, making them unsuitable for monitoring and analysis applications. We propose a novel method that combines neural rendering and procedural modeling to output a complete, interpretable, and biologically plausible 3D mesh model of the crop canopy.

On the other hand, there is a large body of work that has found success in modeling 3D plant shapes via procedural generation [25, 31, 40, 41, 50, 51]. These procedural models are grounded in scientific knowledge and are carefully designed to produce biologically plausible plant shapes that consist of anatomically complete arrangements of plant organs with realistic shapes. However, these models generally require human input to set their parameters, and the task of automatically generating plants that accurately represent plant instances observed in the real world ("inverse procedural modeling") remains to be difficult.

In this work, we present a novel method for 3D modeling of agricultural crops based on optimizing the parameters of a procedural plant morphology model. Our method combines the flexibility of data-driven neural reconstruction methods with the robust foundational knowledge in procedural models. To this end, we first use neural radiance field

(NeRF) techniques [34, 53] to estimate the geometry of the visible surfaces in the scene. We then apply heuristics to localize the planting rows and calculate a *row-aligned camera pose* from which to render depth maps. Next, we render depth maps from both the NeRF and a virtual scene generated by our procedural model, and use Bayesian optimization to minimize a specially designed loss function with respect to the procedural model's parameters.

The design of the loss function and the parameterization of the procedural model are crucial for obtaining a 3D canopy model that is useful for field-level analysis applications. Importantly, for such applications, the fine-grained positions of each *individual* leaf are mostly irrelevant, as long as the *aggregate shape characteristics* of the crop canopy are faithful. We thus define our loss on *histogram statistics* of the depth map and optimize a highly *compact* set of parameters to ensure that the key canopy characteristics for determining crop productivity are accurately captured without being distracted by fitting irrelevant details. Thanks to the strong priors imposed by the procedural model, this process jointly optimizes the *complete* plant shape, including portions not visible in the input images.

To validate our approach, we collect a multi-view dataset of real soybean and maize fields, paired with manual measurements of leaf area and leaf angle. We show that the proposed method can successfully reconstruct realistic crop canopies across different growth stages and estimate key canopy structure variables more accurately than baselines. These variables are tightly linked to crop yield and are used extensively in crop science. We also show that the reconstructed 3D canopies can be directly fed into radiative transfer modeling software to provide accurate predictions of photosynthesis rates. The results highlight the potential for monitoring crop productivity directly from camera images instead of costly flux tower equipment. Our contributions are summarized below:

- We present a novel approach for reconstructing complete 3D morphological models of large-scale crop plant fields from a collection of images.
- We provide a framework for image-based growth and photosynthesis quantification, paving the way for scalable yield prediction and carbon exchange monitoring.

2. Related Work

3D Reconstruction. The task of reconstructing the 3D geometry of a scene given images is a longstanding problem in computer vision. Most reconstruction pipelines start from Structure-from-Motion (SfM), which infers camera parameters for the input images through joint optimization with detected keypoints [45, 54]. Afterwards, 3D geometry can be estimated based on photometric consistency across input views. Multi-view stereo (MVS) methods attempt to match correspondences across images and then triangulate the 3D coordinates of the points [46, 58, 61, 65]. Another approach for 3D reconstruction is based on fitting a neural radiance field (NeRF) to the scene [27, 28, 34, 53, 57, 62, 63]. NeRF is a high-fidelity and compact 3D scene representation that consists of a color field and a density field, both parameterized by neural networks and fitted to the input images via differentiable volume rendering [34]. After training, depth maps can be obtained by volume rendering with the density field. NeRF-based methods can model view-dependent effects and tend to give better surface geometry than MVS. More recently, new methods based on 3D Gaussians have been proposed [14, 19, 24], achieving higher visual fidelity and efficiency than NeRF, but often with less accurate geometry. In our approach, we use an off-the-shelf NeRF [53] to reconstruct visible surfaces of plants, which has been explored before [2, 18, 43, 49]. However, due to the inevitable occlusion in agricultural scenes, this is inadequate for full canopy reconstruction. To solve this issue, we turn to procedural models, which incorporate scientific knowledge and define a space of complete plant shapes that we can constrain our solution to.

Inverse Procedural Modeling. Inverse procedural modeling (IPM) refers to the problem of finding a procedure for generating a 3D representation of a given object or scene. The key design choices for IPM pipelines are the choice of procedural model and the method for optimizing the parameters of the model. IPM has been applied for various input categories, with works focusing on building interiors [5, 11, 20, 21, 32], facades [8, 33, 59], driving scenes [7, 22], and forestry [15, 25, 26, 30, 35, 51]. The procedural models range from scene grammars [12, 21, 22] to constructive solid geometry trees [64] to L-systems [37, 40], which are especially popular for modeling botanical trees. To fit the model, a structural similarity measure is defined based on geometric and or semantic agreement. Depending on the type of model, which may require estimating procedural rules and or parameters [1], various search-based or dynamic programming-based optimization methods may be used to maximize the similarity measure. Unlike the reconstruction methods described above, inverse procedural modeling methods tend to focus less on photometrically exact reconstructions, aiming to obtain structurally similar approximations useful for simulation applications.

To the best of our knowledge, our work is the first to use IPM to reconstruct agricultural crops in the field and the first to integrate NeRFs into the IPM pipeline. Compared to individual plants [30, 66], modeling dense crop canopies poses a new set of challenges due to the high levels of occlusion. IPM methods for trees [26, 51] are inspirational but do not adequately constrain the plant topology for smaller crops. We propose a specialized IPM approach to estimate 3D crop meshes that can be used for plant phenotyping, visualization, and simulation of biophysical processes.

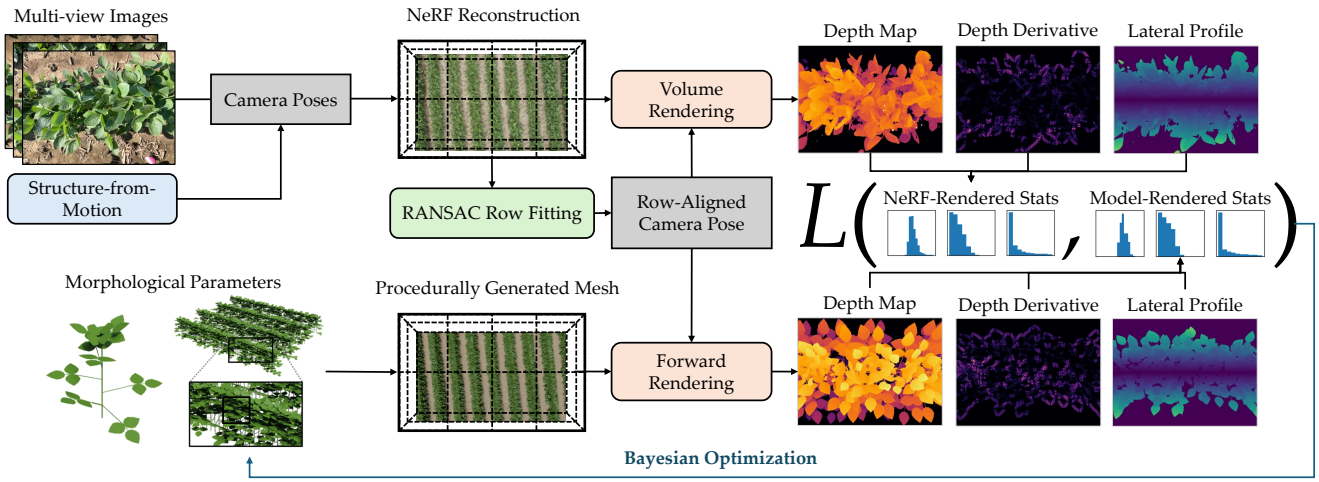

Figure 2. **Overview of our method.** We aim to estimate the parameters for a procedural generation model to generate a shape that matches the observed images. First, we use standard structure-from-motion and neural radiance field (NeRF) techniques to reconstruct the visible geometry of the scene. We then apply geometric heuristics to calculate a camera pose aligned with the planting rows of the crops. This pose is used to render depth maps from both the NeRF and the procedural model. We define a specialized loss function based on histogram statistics of the depth maps and minimize it with respect to the morphological parameters using Bayesian optimization.

3. Inverse Procedural Modeling for Crops

Our proposed method takes as input a set of images of a field of crops, and outputs a set of parameters that can be fed into a procedural generation model to produce a 3D mesh of the crops in the images. An overview of our method is provided in Fig. 2. By constraining the space of possible 3D shapes to the output space of the procedural model, we can ensure that the reconstructed plants are complete and have realistic leaf shapes and branch topology. Of course, we also need our reconstruction to match what we can observe in the input images. We achieve this by searching for parameters that minimize a specialized loss function based on depth maps estimated from the input images and corresponding depth maps rendered from the reconstruction.

The layout for the rest of this section as follows. In Sec. 3.1, we describe the procedural generation models we use for plant morphology. In Sec. 3.2, we describe how we use a neural radiance field (NeRF) [34] and geometric heuristics to obtain row-aligned depth maps from the input images. In Sec. 3.3, we describe our loss function and in Sec. 3.4, we describe our optimization algorithm.

3.1. Procedural Morphology Model

At the core of our method is a procedural morphology model that defines a primitive for the leaf shape and the possible topologies for the plant structure. Since this varies drastically across different species, these models should be species-dependent. In this work, we focus on soybean (*Glycine max*) and maize (*Zea mays*), the two most-grown crops in the United States [55].

Accurate crop modeling requires appropriate species-based shape constraints. Our procedural generation models are heavily based on existing work in the crop morphology literature. For soybean, we build upon the mCanopy model by Song et al. [50], and for maize, we build upon the coupled maize model by Qian et al. [41]. The mCanopy model consists of nodes of trifoliate leaves attached via a petiole to either the main plant stem or a branch stem. The maize model consists of large curved leaves growing from a single main stem. We introduce our own parameterization for each model, as illustrated in Fig. 3. The parameterization is designed to be flexible enough to capture the primary shape variations across instances of each species while remaining as low-dimensional as possible (5 parameters for soybean, 4 for maize). Note that the procedural generation process is in general stochastic; e.g. the model samples leaf angles according to a probability distribution. We refer readers to the supplementary for additional details about each model.

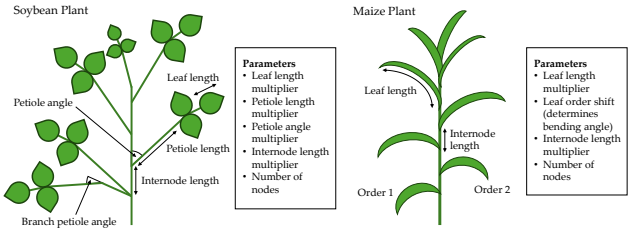

Figure 3. **Procedural plant morphology models.** We adopt procedural generation models from existing soybean morphology [50] and maize morphology [41] models. This (stylized) illustration shows the parameters that we allow to be optimized to model variations across individual instances of each species.

3.2. Pose Alignment for Depth Rendering

The procedural generation model defines a space of 3D shapes that can represent plants of various stages of maturity, growing conditions, and cultivars. In order to search within this space for the shape that best matches the plants in input observations, we define a specialized loss function based on depth maps estimated from the input images and depth maps rendered using the proposed shape.

Note that, in order to properly compare depth maps between the real scene and the procedurally generated scene, the procedural plants should be placed in the same positions and orientations (relative to the camera) as in the real scene. However, it is practically impossible to determine the precise base position of every individual plant under heavy occlusion. Fortunately, it is also *unnecessary* to do so, since our applications demand characterizing the canopy overall, not each individual plant. Thus, we handle pose alignment by using geometric heuristics to estimate plant *row* locations, while carefully designing our loss function to be invariant to individual plant locations *within* each row.

We assume that our crop plants are planted in rows with a known planting density (typically true). To estimate the row locations, we first use an off-the-shelf NeRF (Nerfacto [53]) to obtain a 3D point cloud of the visible surfaces. We render the depth for each input view and unproject each pixel to a 3D point. Then, we randomly sample a fixed number of points within a bounding box for the scene and perform voxel downsampling. Next, we segment the points into ground/plant using a threshold on their color in LAB space. We fit a plane to the ground points using RANSAC [44] and take an upper slice of the plant points based on their distance to the ground plane. Finally, we fit a line to each row by sequentially running RANSAC and removing the inliers for each fitted line. Additional details for the row-fitting procedure are provided in the supplementary.

Given the ground plane and rows, we define a standardized camera to render from at a predetermined height above the center of the row with the most inliers, facing straight downwards with its x-axis aligned with the row. This design will allow our loss to be invariant to the exact placement of plants along the x-axis. The camera is used to render both the observation depth map $\mathbf{I}_{\text{obs}}$ from the NeRF and the depth map $\mathbf{I}_{\text{pred}}$ induced by the procedurally generated scene. Note that depth estimators such as VGGT [56] are unsuitable here since we are rendering depth from a novel (non-input) view. We also use a color threshold and a bounding box to acquire foreground (plant) masks $\mathbf{M}_{\text{obs}}$ and $\mathbf{M}_{\text{pred}}$ corresponding to each depth map.

3.3. Loss Function

One of the key motivations for estimating the 3D structure of agricultural crops is to enable photosynthesis simulations that provide accurate predictions of primary productivity and other ecological variables. Recall that we are interested in these variables at a large scale, at the level of patches or fields, not at the level of individual plants. Thus, the reconstruction of individual leaf locations is not necessary as long as certain statistics of the aggregated crop canopy are accurate. With this in mind, we design a loss function around histogram statistics of the rendered depth maps, aiming to recover the key shape properties necessary for photosynthesis simulation while avoiding overfitting to irrelevant aspects of the scene geometry.

Depth Profile Term. The first term of the loss function is the L2 distance between histograms of foreground depth values in the observed and predicted depth maps. Formally,

$$L_{\text{depth}} = ||\mathbf{h}_{\text{obs}} - \mathbf{h}_{\text{pred}}||_2^2, \tag{1}$$

where $\mathbf{h}_{\text{obs}}$ and $\mathbf{h}_{\text{pred}}$ are histograms of the foreground depth values in $\mathbf{I}_{\text{obs}}$ and $\mathbf{I}_{\text{pred}}$ respectively. Unlike the depth maps themselves, the histograms are invariant to translation along the ground plane, while still containing information about the crop height.

Lateral Profile Term. The lateral profile term L_{lateral} is essentially the same as the depth profile term but for y-coordinate absolute values instead of z-coordinates. This term helps to constrain the extent to which the canopy spreads out on the ground, which the depth profile is mostly invariant to. We do not include another term for the x-coordinate because the x-axis is aligned with the planting rows, making the distribution close to uniform.

Depth Derivative Term. To capture more information about surface normals, we include another L2 loss L_{sobel} on (histograms of) magnitudes of Sobel derivatives of the depth maps. We sum absolute values of the Sobel derivatives in both directions with a kernel size of 3.

Mask Area Term. The final term is a squared-error on the foreground mask area:

$$L_{\text{mask}} = (||\mathbf{M}_{\text{obs}}||_1 - ||\mathbf{M}_{\text{pred}}||_1)^2. \tag{2}$$

Our final loss is a weighted combination of the above terms:

$$L = L_{\text{depth}} + \lambda_{\text{lateral}} L_{\text{lateral}} + \lambda_{\text{sobel}} L_{\text{sobel}} + \lambda_{\text{mask}} L_{\text{mask}}. \tag{3}$$

3.4. Bayesian Optimization

Since the procedural generation model directly adds new mesh faces to create plants with different topology, the transformation from parameters to generated shape is not differentiable. Thus, it is difficult to minimize the loss function with respect to the parameters using gradient-based

Figure 4. **Qualitative reconstruction results (soybean)**. We validate the reconstruction quality of our method on images from real agricultural fields. From left to right: example images from the multi-view data, row-aligned NeRF-rendered depth, procedural model depth, rendered visualization of the procedural mesh. By matching statistics of the depth, our method is able to estimate the key shape parameters needed to characterize the growth of the plants throughout the growing season, consistently producing realistic reconstructions.

optimization methods. Instead, we employ Bayesian optimization, a black-box optimization method that is commonly used for hyperparameter tuning [47]. Bayesian optimization creates a surrogate for the objective function using Gaussian process regression, and then decides where to sample next by optimizing an acquisition function that combines uncertainty and the expected objective value [9]. We perform Bayesian optimization with a Matern kernel and the simple expected-improvement (EI) acquisition function [9] to estimate the procedural generation parameters that minimize our loss function. We find that, in many scenes, there may be multiple solutions with similar loss values. In order to make our method more robust to different random initialization seeds, we run the optimization 10 times (in parallel) and average the solutions, although this is not strictly necessary. An analysis of the distribution of the optimization results is provided in the supplementary.

4. Experiments

Dataset. To validate our approach, we collected a dataset of multi-view images in real crop fields in the U.S. Midwest, paired with manual measurements of key morphological variables. For soybean, we collected at 3 geographical locations at 6 different time points throughout the growing season, for a total of 18 scenes. The images were captured covering a $2\,\mathrm{m}\times2\,\mathrm{m}$ area using the Polycam app on an iPad Pro, around 50 per scene. The paired manual measurements consist of leaf areas and leaf angles (angle between surface normal and zenith direction). Leaf area was measured using a LAI-2200 scanner, and leaf angle was measured using a protractor. For maize, we collected 5 different time points in one location, with each scene containing around 500 images from a DJI Mavic UAV flying at a height of $13\,\mathrm{m}$. The images were aligned using Agisoft Metashape. Here, only

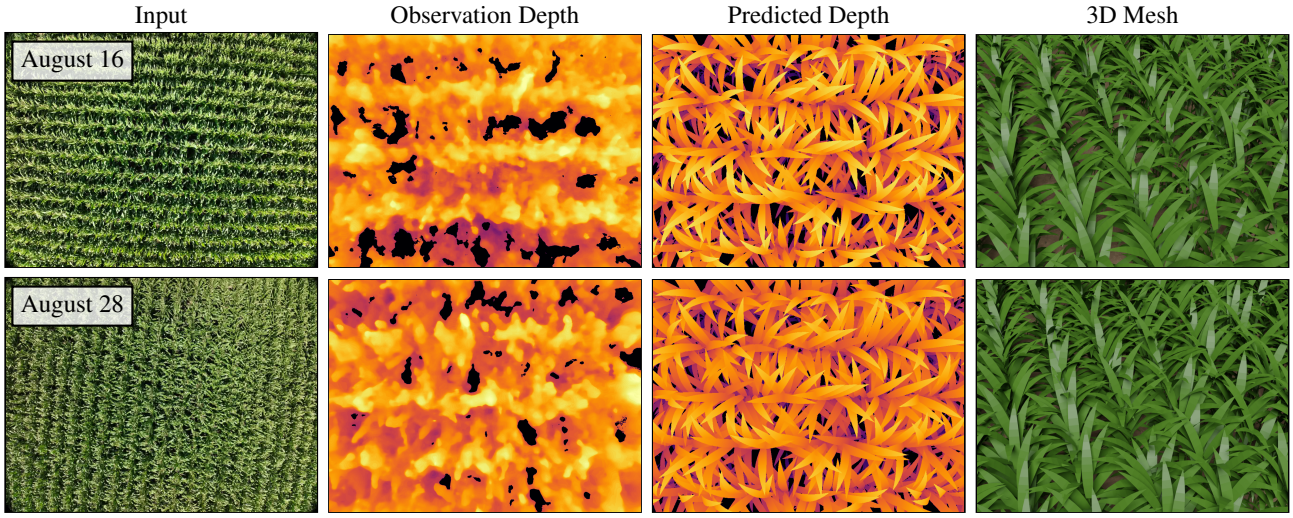

Figure 5. **Qualitative reconstruction results (maize)**. We show that our method can be applied to model maize as well as soybean. From left to right: example images from the drone-captured multi-view data, row-aligned NeRF-rendered depth, procedural model depth, rendered visualization of the procedural mesh. The resulting reconstructions are complete and anatomically realistic despite heavy occlusion.

leaf area was measured, not leaf angles. Both the soybean and maize data are available through our project page.

Metrics. Our evaluation metrics (and field measurements) are centered around the commonly considered canopy structure variables of leaf area index (LAI) and the leaf angle distribution. LAI is defined to be the total surface area of leaves per unit of ground area and is widely used in productivity models, climate models, and methods to estimate other vegetative surface properties [38]. Similarly, leaf angle is regarded as a key component of plant ecological strategy with significant impact on land surface properties such as carbon flux, surface temperature and spectral signature [60]. To assess how accurately our reconstructions capture these key traits, we consider the following metrics: **LAI Error (LAIE)** is the root mean squared error (RMSE) between the predicted and ground-truth (manually measured) LAI. **LAI Percent Error (LAIPE)** is the absolute percent error between the predicted and ground-truth LAI. **Angle Mean Error (AME)** is the RMSE between the mean of the predicted leaf angles and the mean of the ground-truth leaf angles. **Angle Standard Deviation Error (ASDE)** is the RMSE between the standard deviation of the predicted leaf angles and that of the ground-truth leaf angles. We also evaluate the accuracy of the estimated stem structure and topology in the supplementary, using synthetic scenes instead of real scenes due to the difficulty of obtaining corresponding ground-truth measurements.

Baselines. We compare our reconstruction method with a variety of baselines: **Poisson** refers to applying Poisson surface reconstruction [23] on the point cloud extracted from NeRF, and then thresholding in 3D to remove the background geometry. All of the remaining mesh faces are then considered as leaf faces. **MLP** refers to using a learned multilayer perceptron (MLP) to predict procedural generation parameters given the histograms, mask area, and view height of the observation depth map. The MLP is trained on 20K pairs of histogram statistics and model parameters, synthesized by running the procedural generation model and rendering depth maps. **Trust-Region** is the same as our proposed method but uses a (constrained) trust-region method [6] for optimization instead of Bayesian optimization. This requires an estimate of the gradient of the loss function, which is obtained via 2-point finite difference estimation [52]. **Random** refers to uniformly randomly sampling values for the procedural generation parameters.

Implementation Details. For soybean, we train Nerfacto [53] with default parameters for 20K iterations. For maize, we set the near-plane and far-plane to 0.05 and 6.0 respectively, the MLP width to 128, and the orientation method to "vertical". We defer details about the RANSAC row-fitting and depth rendering to the supplementary. For our loss function, we use $\lambda_{\text{lateral}} = 2, \lambda_{\text{sobel}} = 4, \lambda_{\text{mask}} = 1$ for soybean and $\lambda_{\text{lateral}} = 1, \lambda_{\text{sobel}} = 0, \lambda_{\text{mask}} = 1$ for maize. This is because the low NeRF quality in the maize scenes makes the depth derivatives unreliable. For Bayesian optimization, we use scikit-optimize [16] and run for 500 iterations, with 200 iterations for random initialization. Details about histogram bins and baseline implementations can be found in the supplementary.

4.1. Canopy Shape Modeling

Qualitative Results. We show example results at different stages of soybean growth for the same geographic location in Fig. 4. We observe that our method produces

Table 1. **Canopy reconstruction results.** We highlight the **best** and second best values.

	Soybean				Maize	
	LAIE (↓)	LAIPE (↓)	AME (↓)	ASDE (↓)	LAIE (↓)	LAIPE (↓)
Poisson	1.11	0.57	31.60	4.10	1.63	0.42
MLP	0.92	0.24	29.79	**3.91**	2.79	0.88
Trust-Region	1.37	0.28	**10.34**	7.10	1.71	0.53
Random	3.38	2.84	27.92	11.83	2.42	0.75
Ours	**0.69**	**0.15**	12.07	7.39	**0.97**	**0.26**

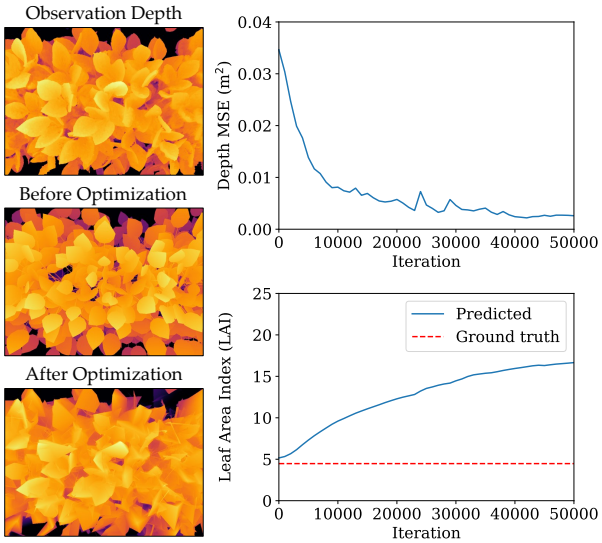

Figure 6. **Leaf-level optimization does not improve overall structure accuracy**. We experiment with using differentiable rendering to directly optimize mesh vertices with respect to the MSE between the rendered and observed depth. Although the depth MSE decreases, the induced LAI becomes increasingly inaccurate, demonstrating the unsuitability of such vertex-level optimization.

complete reconstructions that capture how the shape of the canopy evolves over time, from small seedlings to large bushes. Unlike most existing reconstruction methods, our method is able to estimate the invisible (occluded) portions of the plants as well as the visible portions. The final mesh is also easily decomposable (at both the plant-level and part-level) and is guaranteed to be biologically reasonable thanks to the constraints imposed by the procedural model.

We also show example results for maize in Fig. 5. Due to windy conditions, larger scene scale, and noisy camera pose estimation, the quality of the NeRF is significantly lower in the maize data compared to soybean. However, our method is still able to produce reasonable canopy reconstructions using those observations.

Quantitative Results. We provide quantitative results for the canopy reconstruction in Tab. 1. In all metrics except ASDE, we find that our proposed method achieves the highest performance. The angle standard deviation may be estimated less accurately than the other variables because it does not have as strong of an impact on the loss function.

The Poisson reconstruction fails because it completely ignores occluded regions and also frequently stitches nearby leaves together, creating extraneous faces. The MLP gives reasonable predictions in terms of soybean leaf area, but tends to give highly erroneous leaf angles as a result of the domain gap between the synthetic histograms it was trained on and the actual histograms from the NeRF depth maps. This is exacerbated with the maize data, where the NeRF renderings are blurrier. The trust-region optimization produces leaf angles comparable with our (Bayesian optimization) method, but does not perform as well for leaf area, most likely falling into local minima or making erroneous steps due to the stochasticity of the procedural model.

Naive geometric optimization. We emphasize that our goal is *not* exact leaf-level reconstruction of the crop canopy. Instead, our task demands complete crop canopies that capture key shape statistics of the actual leaf distribution, such that downstream applications like growth monitoring or photosynthesis prediction can be conducted accurately. Our model parameterizations and loss function terms are designed to capture precisely these key shape variations, and not overfit to irrelevant geometric details of the input observations. To illustrate this point, we conduct an experiment using differentiable rendering to directly optimize the vertices of our output mesh to further fit the observed depth map. The optimization is done in PyTorch3D [42] using SGD with learning rate 0.01 and momentum 0.9 on an MSE loss. The results are shown in Fig. 6. Although the pixel-wise depth error can be made very low, important variables such as the total leaf area become *less* accurate, which is reflected in the LAI [38]. This justifies the need for shape constraints and distribution-level optimization.

4.2. Ablation Studies

We conduct an ablation study on the components of our loss function in Tab. 2. We observe that each of the components contribute to producing a more accurate reconstruction, some significantly more than others. The depth derivative loss has a pronounced beneficial effect on the leaf area metrics. This is because the other statistics may have trou-

Table 2. **Loss function ablations.** We highlight the **best** and second best values.

Loss Components				Soybean			
L_{depth}	L_{lateral}	L_{sobel}	L_{mask}	LAIE (↓)	LAIPE (↓)	AME (↓)	ASDE (↓)
✓				1.62	0.52	17.00	9.26
	✓	✓	✓	**0.67**	0.19	12.08	**6.81**
✓		✓	✓	0.71	0.16	10.37	7.69
✓	✓		✓	1.94	0.49	**9.73**	8.15
✓	✓	✓		0.78	0.19	12.30	8.03
✓	✓	✓	✓	0.69	**0.15**	12.07	7.39

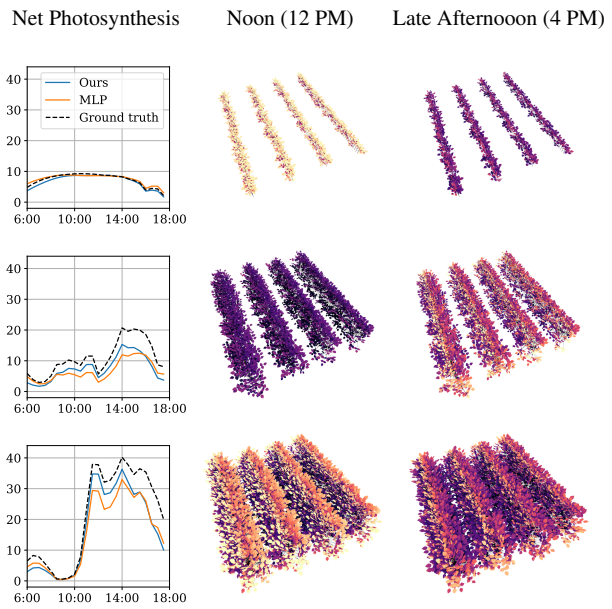

Figure 7. **Photosynthesis simulation results**. We perform simulations using Helios [3] on soybean canopies reconstructed by our method. The left column shows a timeseries of the net photosynthesis rate for the crop canopy over the course of a day, in units of $\mu\text{molCO}_2/\text{m}^2/\text{s}$. The other columns show a mesh visualization where leaf faces are colored by the rate of photosynthesis over that face (brighter = higher rate). The results demonstrate the potential of using our pipeline for monitoring crop productivity.

ble constraining the leaf size when the canopy has covered most of the ground, while the depth derivative distribution will be strongly shifted upwards with smaller leaves compared to larger ones. We note that removing some of the other terms one at a time does not actually significantly harm the final performance. The hard constraints imposed by the procedural model and the soft constraints imposed by the remaining terms are able to force a similar solution. We decide to include all the terms in our final method due to the improved LAIPE, which we prefer over LAIE as it is not biased towards the later time points with larger plants.

We also test our method with different procedural models to investigate the importance of having realistic shape priors. The results can be found in the supplementary, along with a discussion of limitations.

4.3. Photosynthesis Simulation

The outputs of our method can be directly used with radiative transfer models to predict photosynthesis rates, which directly impacts crop productivity. We showcase this using Helios [3], a state-of-the-art biophysical modeling framework. Helios can predict the net photosynthesis rate for every leaf mesh face at any point in time given a set of environmental variables measured by a flux tower with sensors for temperature, humidity, radiation, etc. We show visualizations of the per-face photosynthesis rate as well as timeseries of the aggregate rate across the entire canopy in Fig. 7. Ground-truth values for the photosynthesis rate are usually calculated at a landscape-level using eddy covariance data from a flux tower [39]. Since we do not have paired flux tower measurements for our dataset (collected in 2023), we use measurements from a soybean field in the previous year (2022) with a similar location and climate.

The predictions appear to accurately represent the net photosynthesis over time as well as its spatial variations throughout the canopy, validating our pipeline's potential for facilitating large-scale monitoring of crop productivity.

5. Conclusion

We presented a novel method for 3D modeling of agricultural crops that combines neural rendering with inverse procedural modeling to produce complete plant shapes despite heavy occlusion. Our method first constructs a neural radiance field (NeRF) and then optimizes a procedural model to be consistent with the NeRF in a visibility-aware manner. We validate our method on real-world agricultural fields and show that it can reconstruct realistic crop canopies across a variety of growth stages. We also show that it enables realistic simulations of crop photosynthesis using a radiative transfer model. The method could potentially be improved in the future by incorporating plant growth priors for temporal consistency, or optimizing the model in a coarse-to-fine manner to capture more shape details.

Acknowledgments. This work is supported by NSF Awards #2331878, #2340254, #2312102, #2414227, and #2404385, an Intel research gift, the NCSA Faculty Fellowship, and the UIUC Agroecosystem Sustainability Center.

References

[1] Daniel G Aliaga, İlke Demir, Bedrich Benes, and Michael Wand. Inverse procedural modeling of 3d models for virtual worlds. In *ACM SIGGRAPH 2016 Courses*, pages 1–316. 2016. 2

[2] Muhammad Arbab Arshad, Talukder Jubery, James Afful, Anushrut Jignasu, Aditya Balu, Baskar Ganapathysubramanian, Soumik Sarkar, and Adarsh Krishnamurthy. Evaluating nerfs for 3d plant geometry reconstruction in field conditions. *arXiv preprint arXiv:2402.10344*, 2024. 2

[3] Brian N Bailey. Helios: A scalable 3d plant and environmental biophysical modeling framework. *Frontiers in Plant Science*, 10:1185, 2019. 8

[4] Ramesh Chand. The global food crisis: causes, severity and outlook. *Economic and Political Weekly*, pages 115–122, 2008. 1

[5] Jiacheng Chen, Chen Liu, Jiaye Wu, and Yasutaka Furukawa. Floor-sp: Inverse cad for floorplans by sequential room-wise shortest path. In *ICCV*, 2019. 2

[6] Andrew R Conn, Nicholas IM Gould, and Philippe L Toint. *Trust region methods*. SIAM, 2000. 6

[7] Jeevan Devaranjan, Amlan Kar, and Sanja Fidler. Meta-sim2: Unsupervised learning of scene structure for synthetic data generation. In *ECCV*, 2020. 2

[8] Lubin Fan, Przemyslaw Musialski, Ligang Liu, and Peter Wonka. Structure completion for facade layouts. *ACM Transactions on Graphics*, 33:1 – 11, 2014. 2

[9] Peter I Frazier. A tutorial on bayesian optimization. *arXiv preprint arXiv:1807.02811*, 2018. 5

[10] Pierre Friedlingstein, Michael O'sullivan, Matthew W Jones, Robbie M Andrew, Dorothee CE Bakker, Judith Hauck, Peter Landschützer, Corinne Le Quéré, Ingrid T Luijkx, Glen P Peters, et al. Global carbon budget 2023. *Earth System Science Data*, 15(12):5301–5369, 2023. 1

[11] Yasutaka Furukawa, Brian Curless, Steven M Seitz, and Richard Szeliski. Reconstructing building interiors from images. In *ICCV*, 2009. 2

[12] Aditya Ganeshan, R Kenny Jones, and Daniel Ritchie. Improving unsupervised visual program inference with code rewriting families. In *ICCV*, pages 15791–15801, 2023. 2

[13] Ruiqi Gao, Aleksander Holynski, Philipp Henzler, Arthur Brussee, Ricardo Martin Brualla, Pratul Srinivasan, Jonathan Barron, and Ben Poole. Cat3d: Create anything in 3d with multi-view diffusion models. *NeurIPS*, 37:75468–75494, 2024. 1

[14] Antoine Guédon and Vincent Lepetit. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *arXiv preprint arXiv:2311.12775*, 2023. 2

[15] Jianwei Guo, Shibiao Xu, Dong-Ming Yan, Zhanglin Cheng, Marc Jaeger, and Xiaopeng Zhang. Realistic procedural plant modeling from multiple view images. *IEEE transactions on visualization and computer graphics*, 26(2):1372–1384, 2018. 2

[16] Tim Head, Manoj Kumar, Holger Nahrstaedt, Gilles Louppe, and Iaroslav Shcherbatyi. scikit-optimize/scikit-optimize: v0. 8.1. *Zenodo*, 2020. 6

[17] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. Lrm: Large reconstruction model for single image to 3d. *arXiv preprint arXiv:2311.04400*, 2023. 1

[18] Kewei Hu, Ying Wei, Yaoqiang Pan, Hanwen Kang, and Chao Chen. High-fidelity 3d reconstruction of plants using neural radiance field. *arXiv preprint arXiv:2311.04154*, 2023. 2

[19] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *SIGGRAPH*, pages 1–11, 2024. 2

[20] Satoshi Ikehata, Hang Yang, and Yasutaka Furukawa. Structured indoor modeling. In *ICCV*, 2015. 2

[21] Hamid Izadinia, Qi Shan, and Steven M. Seitz. Im2cad. In *CVPR*, 2017. 2

[22] Amlan Kar, Aayush Prakash, Ming-Yu Liu, Eric Cameracci, Justin Yuan, Matt Rusiniak, David Acuna, Antonio Torralba, and Sanja Fidler. Meta-sim: Learning to generate synthetic datasets. In *ICCV*, pages 4551–4560, 2019. 2

[23] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, page 0, 2006. 6

[24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42 (4), 2023. 2

[25] Bosheng Li, Jacek Kałużny, Jonathan Klein, Dominik L Michels, Wojtek Pałubicki, Bedrich Benes, and Sören Pirk. Learning to reconstruct botanical trees from single images. *ACM Transactions on Graphics (TOG)*, 40(6):1–15, 2021. 1, 2

[26] Yuan Li, Zhihao Liu, Bedrich Benes, Xiaopeng Zhang, and Jianwei Guo. Svdtree: Semantic voxel diffusion for single image tree reconstruction. In *CVPR*, 2024. 2

[27] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-fidelity neural surface reconstruction. In *CVPR*, 2023. 1, 2

[28] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundle-adjusting neural radiance fields. In *ICCV*, 2021. 2

[29] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *CVPR*, 2023. 1

[30] Zhihao Liu, Zhanglin Cheng, and Naoto Yokoya. Neural hierarchical decomposition for single image plant modeling. In *CVPR*, pages 733–742, 2025. 2

[31] Steven Longay, Adam Runions, Frédéric Boudon, and Przemyslaw Prusinkiewicz. Treesketch: Interactive procedural modeling of trees on a tablet. In *SBIM@ Expressive*, pages 107–120. Citeseer, 2012. 1

[32] Kevis-Kokitsi Maninis, Stefan Popov, Matthias Nießner, and Vittorio Ferrari. Vid2cad: Cad model alignment using multi-view constraints from videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1320–1327, 2023. 2

[33] Andelo Martinovic and Luc Van Gool. Bayesian grammar learning for inverse procedural modeling. In *CVPR*, pages 201–208, 2013. 2

[34] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 2, 3

[35] Till Niese, Sören Pirk, Matthias Albrecht, Bedrich Benes, and Oliver Deussen. Procedural urban forestry. *ACM Transactions on Graphics (TOG)*, 41(2):1–18, 2022. 2

[36] World Health Organization et al. *The state of food security and nutrition in the world 2022: Repurposing food and agricultural policies to make healthy diets more affordable*. Food & Agriculture Org., 2022. 1

[37] Wojciech Palubicki, Kipp Horel, Steven Longay, Adam Runions, Brendan Lane, Radomír Měch, and Przemyslaw Prusinkiewicz. Self-organizing tree models for image synthesis. *ACM Transactions On Graphics (TOG)*, 28(3):1–10, 2009. 2

[38] Geoffrey G Parker. Tamm review: Leaf area index (lai) is both a determinant and a consequence of important processes in vegetation canopies. *Forest Ecology and Management*, 477:118496, 2020. 6, 7

[39] Gilberto Pastorello, Carlo Trotta, Eleonora Canfora, Housen Chu, Danielle Christianson, You-Wei Cheah, Cristina Poindexter, Jiquan Chen, Abdelrahman Elbashandy, Marty Humphrey, et al. The fluxnet2015 dataset and the oneflux processing pipeline for eddy covariance data. *Scientific data*, 7(1):225, 2020. 8

[40] Przemyslaw Prusinkiewicz, Jim Hanan, and Radomír Měch. An l-system-based plant modeling language. In *International workshop on applications of graph transformations with industrial relevance*, pages 395–410. Springer, 1999. 1, 2

[41] Binxiang Qian, Wenjiang Huang, Donghui Xie, Huichun Ye, Anting Guo, Yuhao Pan, Yin Jin, Qiaoyun Xie, Quanjun Jiao, Biyao Zhang, et al. Coupled maize model: A 4d maize growth model based on growing degree days. *Computers and Electronics in Agriculture*, 212:108124, 2023. 1, 3

[42] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020. 7

[43] Farah Saeed, Jin Sun, Peggy Ozias-Akins, Ye Juliet Chu, and Changying Charlie Li. Peanutnerf: 3d radiance field for peanuts. In *CVPRW*, pages 6253–6262, 2023. 2

[44] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient ransac for point-cloud shape detection. In *Computer graphics forum*, pages 214–226. Wiley Online Library, 2007. 4

[45] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 2

[46] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006. 1, 2

[47] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P Adams, and Nando De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015. 5

[48] Rebecca A Slattery and Donald R Ort. Perspectives on improving light distribution and light use efficiency in crop canopies. *Plant Physiology*, 185(1):34–48, 2021. 1

[49] Claus Smitt, Michael Halstead, Patrick Zimmer, Thomas Läbe, Esra Guclu, Cyrill Stachniss, and Chris McCool. Pagnerf: Towards fast and efficient end-to-end panoptic 3d representations for agricultural robotics. *IEEE Robotics and Automation Letters*, 9(1):907–914, 2023. 2

[50] Qingfeng Song, Venkatraman Srinivasan, Steve P Long, and Xin-Guang Zhu. Decomposition analysis on soybean productivity increase under elevated co2 using 3-d canopy model reveals synergestic effects of co2 and light in photosynthesis. *Annals of botany*, 126(4):601–614, 2020. 1, 3

[51] Ondrej Stava, Sören Pirk, Julian Kratt, Baoquan Chen, Radomír Měch, Oliver Deussen, and Bedrich Benes. Inverse procedural modelling of trees. In *Computer Graphics Forum*, pages 118–131. Wiley Online Library, 2014. 1, 2

[52] John C Strikwerda. *Finite difference schemes and partial differential equations*. SIAM, 2004. 6

[53] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, et al. Nerfstudio: A modular framework for neural radiance field development. In *ACM Transactions on Graphics (SIGGRAPH)*, 2023. 1, 2, 4, 6

[54] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*, pages 298–372. Springer, 2000. 2

[55] USDA National Agricultural Statistics Service. NASS-quick stats. *UNAS Service (Ed.)*, 2021. 3

[56] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *CVPR*, pages 5294–5306, 2025. 1, 4

[57] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *NeurIPS*, 2021. 1, 2

[58] Zizhuang Wei, Qingtian Zhu, Chen Min, Yisong Chen, and Guoping Wang. Aa-rmvsnet: Adaptive aggregation recurrent multi-view stereo network. In *CVPR*, 2021. 1, 2

[59] Fuzhang Wu, Dong Ming Yan, Weiming Dong, Xiaopeng Zhang, and Peter Wonka. Inverse procedural modeling of facade layouts. *ACM Transactions on Graphics*, 33(4):121, 2014. 2

[60] Xi Yang, Rong Li, Andrew Jablonski, Atticus Stovall, Jongmin Kim, Koong Yi, Yixin Ma, Daniel Beverly, Richard Phillips, Kim Novick, et al. Leaf angle as a leaf and canopy trait: Rejuvenating its role in ecology with new technology. *Ecology Letters*, 2023. 6

[61] Yao Yao, Zixin Luo, Shiwei Li, Jingyang Zhang, Yufan Ren, Lei Zhou, Tian Fang, and Long Quan. Blendedmvs: A large-

scale dataset for generalized multi-view stereo networks. In *CVPR*, 2020. 1, 2

[62] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. In *NeurIPS*, 2021. 1, 2

[63] Lior Yariv, Peter Hedman, Christian Reiser, Dor Verbin, Pratul P Srinivasan, Richard Szeliski, Jonathan T Barron, and Ben Mildenhall. Bakedsdf: Meshing neural sdfs for real-time view synthesis. *arXiv preprint arXiv:2302.14859*, 2023. 2

[64] Fenggen Yu, Qimin Chen, Maham Tanveer, Ali Mahdavi Amiri, and Hao Zhang. D^2csg: Unsupervised learning of compact csg trees with dual complements and dropouts. *NeurIPS*, 36, 2024. 2

[65] Jingyang Zhang, Shiwei Li, Zixin Luo, Tian Fang, and Yao Yao. Vis-mvsnet: Visibility-aware multi-view stereo network. *International Journal of Computer Vision*, 131(1): 199–214, 2023. 1, 2

[66] Rongsheng Zhu, Kai Sun, Zhuangzhuang Yan, Xuehui Yan, Jianglin Yu, Jia Shi, Zhenbang Hu, Hongwei Jiang, Dawei Xin, Zhanguo Zhang, et al. Analysing the phenotype development of soybean plants using low-cost 3d reconstruction. *Scientific Reports*, 10(1):7055, 2020. 1, 2