Deep implicit surface reconstruction of 3D plant geometry from point cloud

Anjana Deva Prasad **Anushrut Jignasu** **Zaki Jubery** **Soumik Sarkar**
Baskar Ganapathysubramanian **Aditya Balu** **Adarsh Krishnamurthy†**

Iowa State University
†adarsh@iastate.edu

Abstract

Reconstructing the geometry of crops from 3D point cloud data is useful for a variety of plant phenotyping applications. Due to very thin and slender segments, obtaining accurate surface geometry representations from the 3D point cloud data is challenging. Further, defects (noise) and holes (sparsity or occlusion) in the point cloud data might be errors in the reconstructed plant structures. While the reconstruction of a surface from an input point cloud has been studied for decades, recent work on deep learning frameworks that learn neural implicit representations have shown significant promise in accurately reconstructing 3D data, especially under noisy and sparse sampling conditions. However, these approaches have not yet been deployed for slender members. In this work, we explore neural implicit representations to reconstruct the surfaces of fully developed maize plants using data acquired from Terrestrial Laser Scanners (TLS). We compare several neural implicit approaches with more traditional methods of surface reconstruction. We also analyze the robustness of these neural implicit methods for 3D plant data reconstruction. We finally utilize the predicted surface to infer structural features from the data. This approach paves the way for detailed flow/-transport simulations of agricultural domains from 3D point cloud data.

Introduction

Plant phenomics has garnered significant interest in recent years due to increasing demand for understanding and unraveling the relationships between phenotype, genotype, management, and the environment, particularly using machine learning approaches (Singh et al. 2016; Guo et al. 2021). Such quantitative relationships can be used in a variety of ways—genome-wide association studies, ideotype design, precision management, climate-resilient breeding—to provide insight into the agricultural decision-making process (Zhou et al. 2021; Singh et al. 2021). Extracting 3D plant traits is particularly useful as they provide additional nuanced information about growth/phyllotaxy, light interception, and information that can be directly incorporated into detailed crop models and agricultural simulations. However, extracting 3D traits is significantly challenging due to the data size (from 3D point cloud data) or computational complexity (from structure from motion approaches). Here, we explore a generalized approach to construct surface representations of full plants from noisy and sparse 3D point cloud measurements. The availability of such an approach will significantly simplify the workflow of pipelining 3D data to detailed agricultural simulators.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

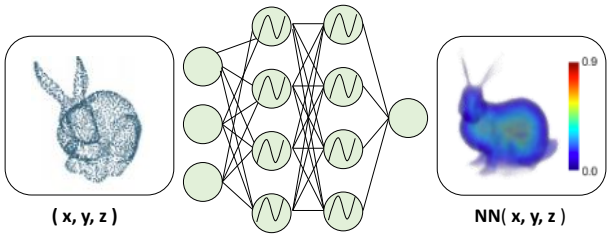

Figure 1: A neural implicit architecture for point cloud reconstruction. It takes in the input points (x,y,z) and predicts the implicit function value, representing the signed distance field for a given geometry.

A recent surge in 3D imaging devices has ushered in an abundance of point cloud data allowing for improved tracking of plant growth and plant traits. A point cloud is a common form of 3D data obtained using LIDAR, laser scanners, or multi-view images of an object. It represents several points (sampled over the surface of the object) in the 3D Euclidean space. Depending on the method of point cloud acquisition, additional attributes such as surface normals, point area, and point color are also acquired along with the point coordinates. In this paper, we use the raw point cloud for our analysis. Using the 3D point cloud data gives us a better understanding of plants than traditional 2D imaging techniques because of the natural availability of more information about their geometry. Particularly, reconstructing plant surfaces from point cloud data provides information such as leaf area, phyllotaxy, and volume of the plant (Pound et al. 2014). Further, we can directly incorporate such data structures into agricultural simulators that can model light interception (Bailey 2019), water transport (Kempthorne et al. 2014), as well as plant scale gas transport. Since most of these computational approaches require the 3D geometry to be represented using a watertight, 2-manifold surface representation, we propose to use implicit surface reconstruction approaches.

This paper focuses on the deep learning-based implicit surface reconstruction of very slender, complex, and flexible structures observed in plant shoots or crops such as maize and wheat. This problem (although a subset of the generic problem of surface reconstruction) brings in several natural challenges. First, the stem is often made of slender members in a plant, and the leaves are made of thin and (possibly non-manifold) surfaces, which makes the geometrical features of the plant very complicated and difficult to reconstruct. Second, plants are topologically complex, and to represent them using a manifold, watertight geometry is challenging and may require many surfaces stitched together carefully. Finally, depending on the method of point cloud acquisition, there may be only partial information and several imperfections such as noisy or misoriented normals.

While several traditional methods have tackled the problem of surface reconstruction for decades (Berger et al. 2017; Jacobson, Kavan, and Sorkine-Hornung 2013), most recent works use deep learning-based approaches to reconstruct different surface representations from input point clouds. These approaches can be broadly grouped into the following categories: (i) Representing shapes as volumetric functions (Choy et al. 2016; Häne, Tulsiani, and Malik 2017); (ii) Fitting a collection of geometric primitives such that the union of these shapes captures the underlying geometry (Gao et al. 2019; Sharma et al. 2019, 2020); (iii) Attempting to infer a mesh representation directly from the input point cloud (Sinha et al. 2017; Groueix et al. 2018); and (iv) Using deep learning to predict an implicit functional representation (called neural implicit representation) whose 0-level set represents the desired surface (Atzmon and Lipman 2020; Gropp et al. 2020; Williams et al. 2021; Sitzmann et al. 2020). Recently, there has been a surge in the use of neural implicit methods for point cloud reconstruction. Traditional representations such as voxels, meshes, and point cloud representations are not storage efficient because the resolution of the output is directly dependent on the complexity of the input. The quality of reconstructions obtained is also dependent on the output size constraints introduced by the feed-forward network. On the other hand, implicit representations are not coupled to the spatial resolution and are ideal for real-world data where the inputs are higher-dimensional signals that are memory intensive and can represent highly complex geometries.

In this work, we explore the application of implicit methods for surface reconstruction on 3D plant data. We focus specifically on recent network architectures for learning neural implicit representations. Our key contributions are:

1. Exploring several surface reconstruction approaches (traditional and neural representation-based) from 3D point cloud data of single plants.
2. Comparison of the reconstructed shape of a synthetically generated point cloud from a non-manifold triangular mesh of a plant to the base mesh, in the context of surface reconstruction of slender structures.
3. Surface reconstruction of real-life point cloud data of plants obtained using terrestrial laser scanners using the best reconstruction approach.

Methods

Classical methods for surface reconstruction fit a surface over dense input point clouds by extracting the information from oriented points clouds using Poisson's reconstruction or approximating implicit surface functions using radial basis functions (RBF) (Kazhdan, Bolitho, and Hoppe 2006; Carr et al. 2001). These methods work on a single shape and do not rely on an entire dataset for modeling. Deep learning-based approaches that learn explicit representations adopt various data-driven techniques for inferring surfaces from the input geometry (Sharma et al. 2019; Groueix et al. 2018). These methods rely on graph neural networks, or encoder-decoder architectures, to reduce point clouds into latent vectors that generate a complete shape from the priors learned by the deep learning framework (Mescheder et al. 2019). While such explicit representations have achieved significant success, they are still heavily coupled to the resolution and complexity of the input. We focus on more recent network architectures that learn implicit representations for every input. An neural implicit representation takes in the input coordinates in the euclidean space (x, y, z) and predicts an implicit functional representation of the object $NN(x, y, z)$ as shown in Fig. 1. One of the most common implicit representations used is the distance fields. Distance fields refer to the minimum distance value to the object boundary at a given point (x, y, z). While most of the neural implicit architectures use signed distance field as the implicit function, the strategies and loss functions used are different.

SAL: Sign Agnostic Learning

In Sign Agnostic Learning (Atzmon and Lipman 2020), the input is the coordinates, and the output is the signed distance field. For training the network, they sample a lot of points in the domain and compute the unsigned distance fields. Using a sign-agnostic loss function, they can obtain the signed distance fields. The sign-agnostic loss function optimizes the network's weights such that the zero-level set of the function is a surface approximating the input geometric data. Additionally, the authors make use of an unsigned distance measure and an unsigned similarity function. The former encourages the function learned by the network to resemble the unsigned distance, and the latter generates a local minimum of the loss where the function learned by the network is considered to be a *signed* function, such that its absolute value approximates the unsigned distance. The neural architecture consists of an 8-layer MLP (multi-layer perceptron) with 512 wide hidden layers and a single skip connection to the middle layer.

IGR : Implicit Geometric Regularization

Finding implicit surface representations directly from raw data is challenging. In prior methods, the learning stages of the network are heavily decoupled from the reconstruction stage, and as a result, the information learned about one kind of shape is not used to improve the reconstruction of other similar shapes. In addition to this, despite having an unsupervised isosurface reconstruction process, even in methods like Sign Agnostic Learning, the proposed loss functions use iterative sampling on the zero-level sets.

In Implicit Geometric Regularization, the surface is predicted without any loss term on the level set(Gropp et al. 2020). The reconstructed surface is generated using stochastic gradient optimization of a simple loss function modeled after the Eikonal equation to fit a multi-layer perceptron to the input point cloud.

NS : Neural Splines

In this method, the authors seek to bridge the gap between traditional reconstruction methods that use kernels and neural network-based methods by leveraging the deep connections between kernels and neural networks (Williams et al. 2021). They propose using random feature kernels that arise from infinitely wide shallow ReLU networks. They claim that kernels instead of neural networks circumvent the slow convergence times introduced by methods like SAL and IGR, which use ReLU activation functions, and SIREN, which uses sinusoidal activations for learning surface from the input point cloud.

SIREN: Sinusoidal Representation Network

In this method, the proposed network architectures learn implicit functions that capture the finer details hidden under natural inputs (Sitzmann et al. 2020). Sinusoidal Representation Networks (or SIRENs) use sine activation functions and learn a set of priors on the inputs and their derivatives for each input data point. SIRENs can also work on diverse input formats like images, audio, video, and even solving partial differential equations such as the Poisson's and Eikonal equation.

Mathematically, SIRENs explore a class of problems that can be represented as

$$F(x, \phi, \nabla_x \phi, \nabla_x^2 \phi, ..) = 0, \phi : x \rightarrow \phi(x) \quad (1)$$

SIREN attempts to find an implicit function ϕ that satisfies a group of M constraints that correlate the function and its derivatives to quantities $a(x)$. This can be expressed as: Compute $\phi(x)$ subject to constraints $C_m(a(x), \phi(x), \nabla\phi(x), ...) = 0, \forall x \in \Omega_m, m = 1, ...M$.

The network then tries to minimize a loss function that penalizes any deviation from the original constraints on the input domain Ω_m:

$$\mathcal{L} = \int_{\Omega} \sum_{m=1}^{M} \mathbf{1}_{\Omega_m}(\mathbf{x}) ||\mathcal{C}_{\Updownarrow}(a(\mathbf{x}), \phi(\mathbf{x}), \nabla\phi(\mathbf{x})...)|| d\mathbf{x}. \quad (2)$$

The network architecture for SIREN uses the sine function as a periodic activation function to generate neural implicit representations:

$$\phi(\mathbf{x}) = \mathbf{W}_n(\phi_{n-1}.\phi_{n-2}...\phi_0)(\mathbf{x}) + \mathbf{b}_n \quad (3)$$

where the input $\mathbf{x}_i$ maps to the output $\phi_i(\mathbf{x}_i)$ as $sin(\mathbf{W}_i\mathbf{x}_i + \mathbf{b}_i)$. ϕ_i represents each layer in the network and it comprises an affine transformation computed using a weight matrix W_i and biases b_i, after which a non-linear sine activation function is applied to each component of the resulting vector.

For our experiments, we leverage the potential of SIREN to represent shapes using learned signed distance functions

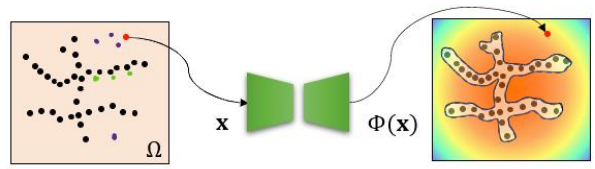

Figure 2: The input to SIREN is a collection of *on-surface* points from the input point cloud and local and global *off-surface* points. The black points represent the *on-surface* points. The green and violet points denote the local and global *off-surface* points, respectively.

(SDF). An SDF measures the distance of a point Ω of the point cloud sample space from the boundary points Ω_m. We train a network that maps the input point coordinates (x, y, z) to signed distance values. In addition to the input point cloud (*on-surface* points), we provide samples of local and global *off-surface* points as input to the network (see Fig. 2). For the local *off-surface* points, we use a distribution defined as the average of a uniform distribution and a sum of Gaussians centered at the point cloud, with a standard deviation equal to the distance of the kth nearest neighbor (k set to 50, by default). For the global off-surface points, we use a random uniform distribution. We used the same number of on-surface points, local and global off-surface points.

SIRENs offer the unique distinction of learning from complicated signals since the derivative of the sine activation function used is also a phase-shifted sine. This allows us to impose multiple constraints that amount to solving the Eikonal equation. The main constraint of the Eikonal boundary value equation is that the norm of spatial gradients is 1 everywhere. For training our network, the model $\phi(x)$ learns using points sampled from the surface of the point cloud where $\phi(x) = 0$. The loss function for our network is defined as follows:

$$\begin{aligned}\mathcal{L}_{\mathcal{SDF}} = &\int_{\Omega} ||\nabla_x \phi(x) - 1|| dx \\ &+ \int_{\Omega_0} ||\phi(x)|| + (1 - < \nabla_x \phi(x), n(x) >) dx \\ &+ \int_{\Omega \setminus \Omega_0} \psi(\phi(x)) dx. \end{aligned} \quad (4)$$

The second term indicates that for points on the surface, the SDF should be zero. It also makes sure the gradient of the SDF and the normal vector align. The third term in the loss function penalizes off-surface points to ensure SDF values are close to zero.

Dataset

In this paper, we perform two tasks: (i) a comparison of all the above-mentioned neural implicit representations along with some traditional surface reconstruction approaches. (ii) we then use the best surface reconstruction approach for obtaining neural implicit representations for real-life point clouds of the plant. To this end, we use two different datasets.

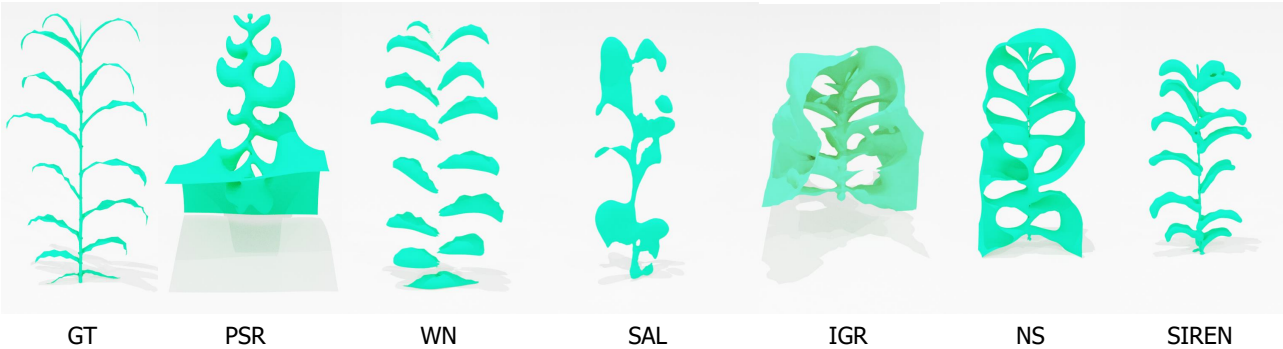

Figure 3: Results of different surface reconstruction methods. From left to right: Ground truth (GT), Poisson Surface reconstruction (PSR), Winding number (WN), Sign agnostic learning (SAL), Implicit geometric representation (IGR), Neural splines (NS), Sinusoidal representation network (SIREN).

First, using a non-manifold triangulated mesh of a slender 3D plant, we randomly sample points of arbitrary size to obtain the point cloud to compare different approaches. Next, we acquire real-life 3D point cloud data using the Faro Focus S350 Scanner. The scanner's angular resolution is 0.011 degrees, corresponding to a 1.5 mm point spacing over a 10 m scanning range. The scanner can acquire point clouds of up 700 million points (MP) at 1 million points per second. The plants were cultivated in the field and were harvested when they reached vegetative maturity and transported to an indoor room for scanning. The plants were oriented so that most of their surface was within the scanner's line of sight, and there were no overlaps between plants. Individual plants were cropped out from the scanned point clouds and stored in separate files using the distance filter and connected components algorithm.

Results

We explored several implicit surface reconstruction approaches, both traditional and neural representation-based on a single plant mesh generated synthetically. Fig. 3 shows the comparison of different methods to the ground truth. Here, the ground truth refers to the original non-manifold triangulated mesh used for sampling the point cloud. For the sake of comparison, we use 300,000 points in the point cloud and use the raw point coordinates only. Using jet fitting, we compute the normals and orient them consistently. Using these points and normals together, we perform surface reconstruction using several methods as shown in Fig. 3. PSR and WN refer to Poisson Surface Reconstruction and Winding number-based reconstruction. Both the approaches are not based on machine learning, whereas all the other methods use machine learning. We observe that among all the methods explored, SIREN provides the best reconstruction of the point cloud.

We make use of the one-sided Chamfer distance and the Hausdorff distance to quantitatively assess the surface reconstruction.

$$\mathcal{L}_{CD} = \sum_{\vec{P}_i \in \vec{P}} \min_{\vec{Q}_j \in \vec{Q}} ||\vec{P}_i - \vec{Q}_j||_2 \tag{5}$$

Table 1: Performance of the different reconstruction approaches: Poisson Surface reconstruction (PSR), Winding number (WN), Sign agnostic learning (SAL), Implicit geometric representation (IGR), Neural splines (NS), Sinusoidal representation network (SIREN).

Accuracy	CD	HD
PSR	54.07	206.27
WN	97.14	81.17
NS	13.43	59.47
SAL	89.09	167.43
IGR	88.98	166.89
SIREN	4.65	14.89

Table 2: Reconstruction time, including training and inference time for neural approaches.

Speed	Time (s)
PSR	54.07
WN	93449.40
NS	184.47
SAL	5635.44
IGR	3586.71
SIREN	444.26

$$\mathcal{L}_{HD} = \max_{\vec{P}_i \in \vec{P}} \left(\min_{\vec{Q}_j \in \vec{Q}} ||\vec{P}_i - \vec{Q}_j||_2 \right) \tag{6}$$

Here, $||\vec{P}_i - \vec{Q}_i||_2$ refers to the L_2 norm of the difference or the distance between the two points $\vec{P}_i$ and $\vec{Q}_i$. In Table 1, we compare the accuracy of each method to the original triangulated mesh. We found that among all the approaches, SIREN was able to capture the stem and leaf geometries in the plant most accurately, thus having the least reconstruction error. Given the maximum bounding box length of 320, the one-sided HD of 14.89 is less than 5% of the bounding box.

Figure 4: Point cloud reconstruction for three different plants with the input data (left) in green, and the SIREN reconstruction (right) in cyan.

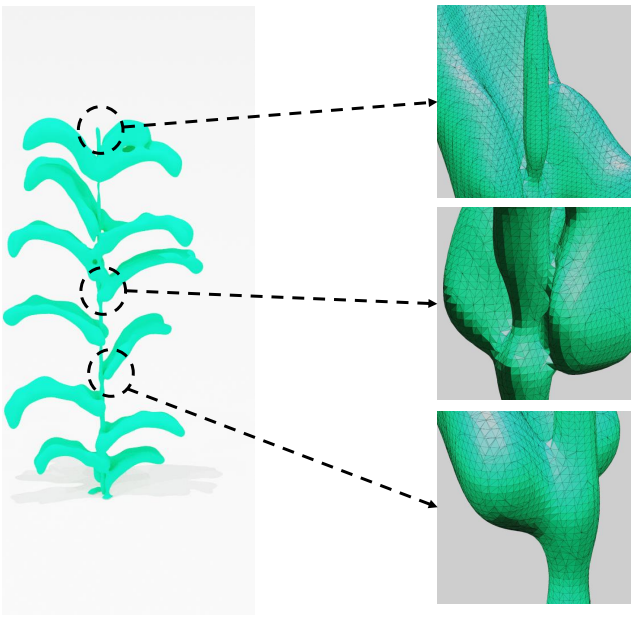

Figure 5: An Anecdotal example of the reconstructed mesh using SIREN. Note that, the surface obtained here is watertight, 2-manifold and closed.

Further, the time taken for complete reconstruction (including the training time and the inference time for obtaining the surface mesh) is also low for SIREN compared to other methods such as SAL or WN as shown in Table 2. While there are methods such as PR and NS that are faster, the accuracy of the reconstructed surface is much better with SIREN. Further, PR and NS have additional surfaces outside the central plant geometry, making the predicted surfaces non-manifold and not watertight. On the other hand, SIREN obtains a closed and watertight surface as shown in Fig. 5.

We also performed an ablation study with a different number of points sampled from the same non-manifold triangular mesh to understand the number of points required for good reconstruction. We observe that both the CD and HD reach low values with around 50,000 points. The CD does not improve much with more points, while the HD, being a more conservative metric, fluctuates with an increase in points.

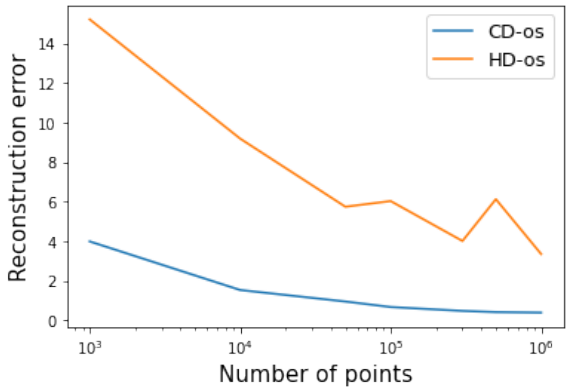

Figure 6: Reconstruction error (quantified using CD, one-sided and HD, one-sided) trends with increasing number of points.

Finally, we show the ability of our method to be able to reconstruct the shapes well for real-life point cloud of maize plants using SIREN in Fig. 4. We can obtain an accurate watertight reconstruction of the plant shapes.

Conclusions

We explored several approaches for surface reconstruction of slender plant structures from a 3D point cloud. We observed that SIREN accurately captured the details of a slender 3D plant. We also used SIREN to reconstruct the geometry of real-world plants obtained from a terrestrial laser scanner. While SIREN did provide good reconstructions of the input, it still does not reconstruct the thin geometry of the plant leaves. We believe we can resolve this issue by adding an extra term to the loss function that enforces the thinness of the leaves. We also plan to use the reconstructed plant surfaces to perform radiosity calculations and construct the skeleton of the plants to analyze different plant traits.

References

Atzmon, M.; and Lipman, Y. 2020. Sal: Sign agnostic learning of shapes from raw data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2565–2574.

Bailey, B. N. 2019. Helios: A Scalable 3D Plant and Environmental Biophysical Modeling Framework. *Frontiers in Plant Science*, 10: 1185.

Berger, M.; Tagliasacchi, A.; Seversky, L. M.; Alliez, P.; Guennebaud, G.; Levine, J. A.; Sharf, A.; and Silva, C. T. 2017. A Survey of Surface Reconstruction from Point Clouds. *Comput. Graph. Forum*, 36(1): 301–329.

Carr, J. C.; Beatson, R. K.; Cherrie, J. B.; Mitchell, T. J.; Fright, W. R.; McCallum, B. C.; and Evans, T. R. 2001. Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 67–76.

Choy, C. B.; Xu, D.; Gwak, J.; Chen, K.; and Savarese, S. 2016. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction. *CoRR*, abs/1604.00449.

Gao, J.; Tang, C.; Ganapathi-Subramanian, V.; Huang, J.; Su, H.; and Guibas, L. J. 2019. DeepSpline: Data-Driven Reconstruction of Parametric Curves and Surfaces. arXiv:1901.03781.

Gropp, A.; Yariv, L.; Haim, N.; Atzmon, M.; and Lipman, Y. 2020. Implicit geometric regularization for learning shapes. *arXiv preprint arXiv:2002.10099*.

Groueix, T.; Fisher, M.; Kim, V. G.; Russell, B. C.; and Aubry, M. 2018. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. arXiv:1802.05384.

Guo, W.; Carroll, M. E.; Singh, A.; Swetnam, T. L.; Merchant, N.; Sarkar, S.; Singh, A. K.; and Ganapathysubramanian, B. 2021. UAS-Based Plant Phenotyping for Research and Breeding Applications. *Plant Phenomics*, 2021.

Häne, C.; Tulsiani, S.; and Malik, J. 2017. Hierarchical Surface Prediction for 3D Object Reconstruction. *CoRR*, abs/1704.00710.

Jacobson, A.; Kavan, L.; and Sorkine-Hornung, O. 2013. Robust Inside-Outside Segmentation using Generalized Winding Numbers. *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH)*, 32(4): 33:1–33:12.

Kazhdan, M.; Bolitho, M.; and Hoppe, H. 2006. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7.

Kempthorne, D. M.; Turner, I. W.; Belward, J. A.; McCue, S. W.; Barry, M.; Young, J.; Dorr, G. J.; Hanan, J.; and Zabkiewicz, J. A. 2014. Surface reconstruction of wheat leaf morphology from three-dimensional scanned data. *Functional Plant Biology*, 42(5): 444–451.

Mescheder, L.; Oechsle, M.; Niemeyer, M.; Nowozin, S.; and Geiger, A. 2019. Occupancy Networks: Learning 3D Reconstruction in Function Space. arXiv:1812.03828.

Pound, M. P.; French, A. P.; Murchie, E. H.; and Pridmore, T. P. 2014. Surface reconstruction of plant shoots from multiple views. In *European Conference on Computer Vision*, 158–173. Springer.

Sharma, G.; Goyal, R.; Liu, D.; Kalogerakis, E.; and Maji, S. 2019. Neural Shape Parsers for Constructive Solid Geometry. arXiv:1912.11393.

Sharma, G.; Liu, D.; Kalogerakis, E.; Maji, S.; Chaudhuri, S.; and Mech, R. 2020. ParSeNet: A Parametric Surface Fitting Network for 3D Point Clouds. *CoRR*, abs/2003.12181.

Singh, A.; Ganapathysubramanian, B.; Singh, A. K.; and Sarkar, S. 2016. Machine learning for high-throughput stress phenotyping in plants. *Trends in plant science*, 21(2): 110–124.

Singh, A.; Jones, S.; Ganapathysubramanian, B.; Sarkar, S.; Mueller, D.; Sandhu, K.; and Nagasubramanian, K. 2021. Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends in Plant Science*, 26(1): 53–69.

Sinha, A.; Unmesh, A.; Huang, Q.; and Ramani, K. 2017. SurfNet: Generating 3D shape surfaces using deep residual networks. arXiv:1703.04079.

Sitzmann, V.; Martel, J.; Bergman, A.; Lindell, D.; and Wetzstein, G. 2020. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33.

Williams, F.; Trager, M.; Bruna, J.; and Zorin, D. 2021. Neural Splines: Fitting 3D Surfaces With Infinitely-Wide Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9949–9958.

Zhou, Y.; Kusmec, A.; Mirnezami, S. V.; Attigala, L.; Srinivasan, S.; Jubery, T. Z.; Schnable, J. C.; Salas-Fernandez, M. G.; Ganapathysubramanian, B.; and Schnable, P. S. 2021. Identification and utilization of genetic determinants of trait measurement errors in image-based, high-throughput phenotyping. *The Plant Cell*, 33(8): 2562–2582.