Large Scale Farm Scene Modeling from Remote Sensing Imagery

ZHIQI XIAO and HAO JIANG[*], Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, China
ZHIGANG DENG[†], Department of Computer Science, University of Houston, USA
RAN LI, WENWEI HAN, and ZHAOQI WANG, Institute of Computing Technology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, China

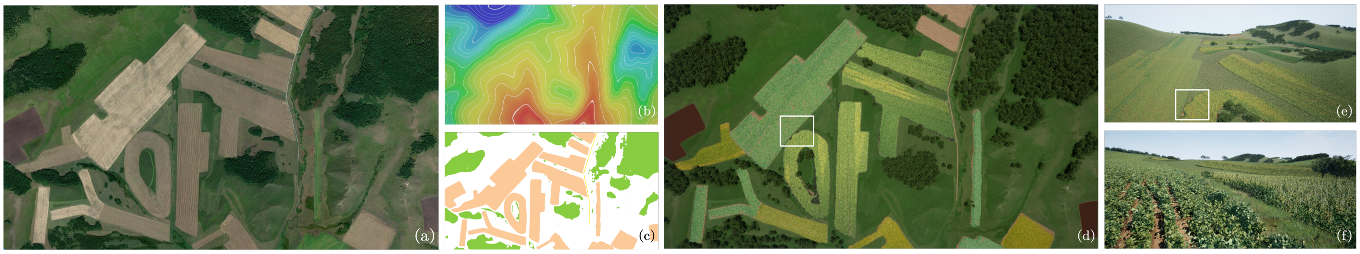


Fig. 1. (a) An input satellite image. (b) An input elevation map used for terrain modeling. (c) The semantic segmentation result obtained using our method. (d) The farm scene generated using our method combined with interactive editing, populated with parametric layout models representing fields, trees, roads, and grassland. The layout of the scene is derived from the input satellite image. (e) A zoomed-in view showcasing the rich, self-organized patterns of its agricultural landscape. (f) A further zoomed-in view highlighting the detailed variations in plant growth.

In this paper we propose a scalable framework for large-scale farm scene modeling that utilizes remote sensing data, specifically satellite images. Our approach begins by accurately extracting and categorizing the distributions of various scene elements from satellite images into four distinct layers: fields, trees, roads, and grasslands. For each layer, we introduce a set of controllable Parametric Layout Models (PLMs). These models are capable of learning layout parameters from satellite images, enabling them to generate complex, large-scale farm scenes that closely reproduce reality across multiple scales. Additionally, our framework provides intuitive control for users to adjust layout parameters to simulate different stages of crop growth and planting patterns. This adaptability makes our model an excellent tool for graphics and virtual reality applications. Experimental results demonstrate that our approach can rapidly generate a variety of realistic and highly detailed farm scenes with minimal inputs.

CCS Concepts: • **Computing methodologies** → **Mesh models**; **Shape representations**; **Virtual reality**.

[*] co-corresponding author
[†] co-corresponding author

Authors' Contact Information: Zhiqi Xiao, xiaozhiqi22b@ict.ac.cn; Hao Jiang, jianghao@ict.ac.cn, Institute of Computing Technology, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, China; Zhigang Deng, Department of Computer Science, University of Houston, USA, zdeng4@central.uh.edu; Ran Li, rleeiris@qq.com; Wenwei Han, asadirysue@gmail.com; Zhaoqi Wang, zqwang@ict.ac.cn, Institute of Computing Technology, Chinese Academy of Sciences, and University of Chinese Academy of Sciences, China.

ACM 1557-7368/2024/12-ART260
https://doi.org/10.1145/3687918

Additional Key Words and Phrases: Image-based modeling, procedural modeling, farm scene modeling, remote sensing images, parametric layout models

ACM Reference Format:
Zhiqi Xiao, Hao Jiang, Zhigang Deng, Ran Li, Wenwei Han, and Zhaoqi Wang. 2024. Large Scale Farm Scene Modeling from Remote Sensing Imagery. *ACM Trans. Graph.* 43, 6, Article 260 (December 2024), 12 pages. https://doi.org/10.1145/3687918

1 Introduction

Scene modeling has been used in a wide range of applications, including urban planning [Vanegas et al. 2009], building detection [Marmanis et al. 2018], and natural resource management [Salter et al. 2009], etc. In particular, farm scene modeling is poised to unlock significant potential in agricultural remote sensing, such as cultivated land detection and yield prediction [Liakos et al. 2018]. In addition, farm scene modeling can craft immersive virtual environments for games and VR, and such models can serve as a critical starting point for the simulation of various agricultural downstream tasks, such as training for unmanned agricultural machinery.

Manually modeling large-scale detailed farm scenes is non-trivial. Procedural modeling, bolstered by advances in inverse procedural modeling [Li et al. 2021; Martinovic and Van Gool 2013; Niese et al. 2022], offers a promising solution. Although many of them focus on specific elements, a holistic farm scene requires the consideration of diverse elements like fields, trees, and roads. In reality, each of these elements exhibits distinct characteristics. For example, the layout of fields and the orientation of roads are often influenced by planning and established practices, while the density of trees might be constrained by natural ecological factors. Given the distinct characteristics of these elements and the influence of human interventions, defining separate procedural generation rules for reconstruction is essential. However, as they are intricately intertwined, separating and modeling them individually presents a challenge.

Satellite images, widely available and capable of capturing vast agricultural landscapes, are an ideal source for farm scene modeling. However, directly using satellite images to generate realistic farm scenes is challenging due to the following inherent limitations of remote sensing. First, the presence of noise in remote sensing images, a consequence of fixed satellite viewpoints, diminished object resolutions, vague boundaries, and a shortage of high-quality annotations, significantly impedes region partitioning and the semantic extraction of distinct elements. Second, the representational capacity of farmland satellite images has intrinsic constraints. Specifically, a single pixel might encompass multiple crops, obscuring detailed information about crop growth.

In this work, we introduce a new farm scene modeling framework to create accurate and realistic farm scenes that adhere to actual distributions, by leveraging readily available satellite images. Inspired by structured planting practices in modern agriculture, we observed that fields and forest regions, while appearing as scattered patches in satellite images, often display consistent internal textural patterns. Cutting-edge deep learning techniques excel at distinguishing such distinct textures. Specifically, our framework accurately extracts the distribution of the scene from a satellite image and decomposes the scene into four layers: fields, trees, roads, and grasslands. A parametric layout model (PLM) is defined for each layer, learning its respective layout parameters from the satellite image. Our experimental results demonstrate the efficacy of the proposed method. Our method enables the efficient creation of 3D farm scenes that reproduce input satellite imagery with remarkable fidelity. Further, user-controlled parameter adjustments introduce flexibility in the level of detail and variations within the generated scenes.

The main contributions of our work are as follows: (1) To utilize remote sensing data for precise farm scene reconstruction, we develop a hierarchical virtual farm procedural modeling approach that robustly identifies the genuine scene distribution from satellite imagery. (2) In order to decipher and define a range of farm scene patterns, we design an array of PLMs that are capable of accurately reproducing realistic farm scenes at varying scales, including a field PLM that effectively models fields by capturing crop spatial distributions and phenological stages, thereby replicating self-organizing patterns observed in agricultural landscapes.

2 Related Work

Inverse Procedural Modeling. Many inverse procedural modeling methods have been proposed for individual categories, including terrain [Guérin et al. 2017; Zhao et al. 2019], buildings [Demir et al. 2016; Kelly et al. 2017; Martinovic and Van Gool 2013; Nishida et al. 2016; Vanegas et al. 2012a], man-made objects [Bokeloh et al. 2010], and material textures [Guerrero et al. 2022; Hu et al. 2019, 2022a,b; Lagae et al. 2009; Wei et al. 2008]. Each method is strictly confined to its specific category. Researchers also applied generic rules to accommodate various inputs, including leveraging natural language processing to encode a set of generative rules to learn design patterns from data [Talton et al. 2012], or using deep learning to learn the L-system representation of pixel images [Guo et al. 2020]. However, these strategies generally do not consider the decoupling and modeling of diverse categories present within a single input. This proves to be a significant challenge in intricate settings, such as farm landscapes, where various categories are often intertwined.

Agricultural Farmland Extraction. Farmland extraction plays a crucial role in remote sensing image analysis, with the objective of accurately and automatically delineating agricultural boundaries or crop types. Traditional methods [Cheng et al. 2020; Rydberg and Borgefors 2001; Wagner and Oppelt 2020; Yan and Roy 2014], including multispectral segmentation and gradient edge detection, and time-series data processing, often faced challenges with irregularly shaped farmlands and required non-trivial parameter tuning, resulting in potential over-segmentation [Hong et al. 2021]. Recently, deep learning has shown immense potential to extract agricultural plots from high-resolution remote sensing imagery [Fare Garnot and Landrieu 2021; Jong et al. 2022; Long et al. 2022; Masoud et al. 2019; Waldner and Diakogiannis 2020; Xia et al. 2018; Xu et al. 2023]. For example, Masoud et al. [2019] used a fully convolutional network (FCN) for farm boundary delineation and introduced a method to enhance spatial resolution. Persello et al. [2019] developed a strategy for smallholder farm boundary extraction by merging global and grouping algorithms. Researchers also applied deep convolutional networks [Waldner and Diakogiannis 2020] and temporal self-attention networks [Fare Garnot and Landrieu 2021] for more refined agricultural data extraction. These methods, however, face the challenge of generalization across diverse landscapes and farm types.

Urban and Plant Modeling. In the realm of scene reconstruction, parallels can be drawn between farm modeling and urban modeling, especially considering the prevalence of vegetation and roads. Several urban modeling studies [Aliaga et al. 2016, 2008; Chen et al. 2008; Parish and Müller 2001; Vanegas et al. 2012b] proposed road modeling methodologies based on generation. Nonetheless, the primary focus within the scope of farm scene reconstruction resides in the accurate replication of actual road locations, as opposed to the generation of new road networks.

On the vegetation front, in addition to detailed 3D modeling of individual trees [Liu et al. 2021], some research efforts focused on the spatial distribution of trees in urban scenes [Beneš et al. 2011; Niese et al. 2022], or modeled plants within ecological systems [Makowski et al. 2019; Pałubicki et al. 2022]. In particular, Niese et al. [2022] proposed the first method to place vegetation in large-scale urban landscapes using satellite images. The Procedural Placement Models (PPM) in their approach simplify complex vegetation patterns into a set of simple placement rules. However, directly applying PPM to farm scenes would present additional challenges. Influenced by various environmental factors, combined with distinct planting patterns, crops not only exhibit spatial self-organization patterns but also are in different phenological stages due to genetic and nutritional competition [Gao and Zhang 2021]. The focus of our work diverges significantly from these existing efforts, specifically accounting for the presence and distribution of crops and trees in natural outdoor settings. Unlike traditional style transfer methods that describe vegetation coverage, we employ semantic segmentation techniques to precisely capture masks for various elements.

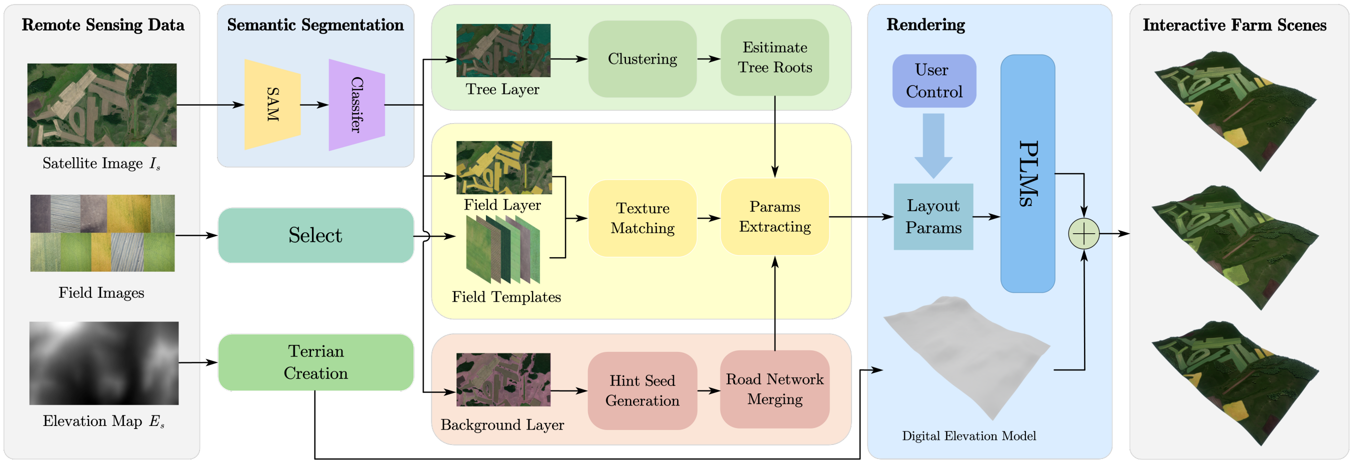


Fig. 2. Pipeline illustration of our farm scene modeling framework. First, the input satellite imagery undergoes region-based semantic segmentation, resulting in distinct classifications of farmland, tree, and background layers. Upon this foundation, we define PLMs for each layer, extracting necessary parameter values accordingly.

3 Method Overview

Our approach consists of the following two main modules. Figure 2 illustrates the pipeline of our framework.

Farm scene semantic segmentation. A two-step segmentation pipeline is specifically tailored to accurately capture the unique distributions associated with various farm scene elements. We dissect the input satellite image into discrete region masks, which are subsequently classified into distinct categories: field layers and tree layers. After that, we further refine the initial segmentation by subdividing these masks to isolate elements of interest. For the background layer, we first extract the road network structures. The remaining unclassified regions naturally assume to be grassland.

First, we employ the SAM segmentation framework [Kirillov et al. 2023] to dissect the input satellite image, denoted as $\mathcal{I}_s$, into discrete region masks. These masks are subsequently classified into distinct categories: field layers and tree layers. Following this initial segmentation, we further refine our results by subdividing these masks to isolate elements of interest. Additionally, we establish templates for a variety of crop types and stages of cultivation. We then deploy texture matching techniques for more refined subdivisions within the field layers. As for the background layer, we first extract the road network structures. The remaining unclassified regions naturally assume the role of grasslands.

Procedural farm scene generation. We customize PLMs for each layer, utilizing machine learning techniques to automatically extract the necessary parameter values for the PLMs. We define distinct layout rules for different field templates, thereby enabling our modeled fields to display macro-scale crop self-organization patterns. Meanwhile, through the incorporation of vigor parameters within the PLMs, we allow the simulation to reflect plant growth variations at the granular level.

4 Farm Scene Analysis and Segmentation

As described above, our farm scene segmentation is a two-stage process. The first stage focuses on the accurate delineation of boundaries, setting the stage for the second stage, where the semantic categorization of these delineated areas is determined. To this end, a detailed structured representation of the farm scene can be achieved, thereby enabling a deep analysis and reconstruction of different regions in the farmland landscape.

In light of advances in the SAM segmentation framework [Kirillov et al. 2023], which excels in boundary extraction, we leverage its powerful performance to extract boundary information. However, the semantic meanings of the extracted regions are not defined. To address this, we use neural networks to semantically interpret these patches. The intrinsic variability in the segmented regions means that these patches come in various sizes and shapes, as illustrated in Figure 3(b).

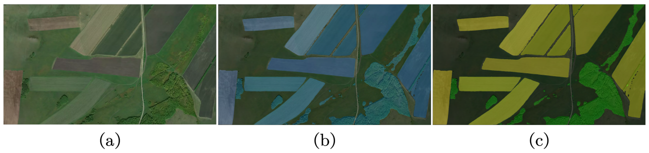


Fig. 3. Our two-stage segmentation process: an input satellite image (a), the preliminary segmentation result (b) using the SAM, and the final segmentation result (c) obtained by a trained classifier, where yellow areas represent fields and green areas represent trees.

The variability of patch sizes introduces a challenge: the discrepancy in input dimensions. To handle this inconsistency, we leverage the pre-trained ResNet34 model [He et al. 2016]. By applying adaptive average pooling to each patch region, we achieve the uniformity of dimensionality and extract standardized features from the regions. We then employ an MLP classifier, augmented with stacked convolutions during the feature processing phase. This not only ensures the consistency in the input dimensions, but also permits us to categorize spatial zones rooted in their inherent attributes. The entire division process can be represented as

$$\mathrm{Y} = c(f(\mathcal{I}_s \odot \mathcal{M}_{\mathrm{sam}})), \tag{1}$$

where Y represents the labels, f and c refer to the feature extractor and classifier respectively, and $\mathcal{M}_{\mathrm{sam}}$ represents the mask images obtained through segmentation. Through the $\odot$ operation, the regions corresponding to the masks are extracted. To this end, we divide the input satellite image into distinct categories, namely, fields, trees, and background, shown in Figure 3(c). Each category is treated with different methods, as described in the following sections.

Upon the determination of the subcategory for a region, appropriate machine learning methodologies are applied to extract the parameters indispensable to render the landscape. As a result, each PLM takes a triplet as input, described below.

$$\mathcal{L} = (\mathcal{R}_s, \mathcal{P}_l, \mathcal{M}_n)\,, \tag{2}$$

Where $\mathcal{R}_s$ represents a set of layout rules associated with a subcategory s, $\mathcal{P}_l$ denotes a collection of distinct layout parameters, and $\mathcal{M}_n$ denotes the regional mask used for the arrangement. In the following sections, we will describe the PLMs for fields, trees, and roads in detail. As for grassland areas, considering their complex and unordered textures, we merely implement a projection intensity parameter to regulate the addition of color offsets for the corresponding sections of the satellite image in the 3D scene's ground.

5 Procedural Field Crops

Our field modeling process begins by selecting a characteristic template for each farmland category from satellite imagery. For each crop type, we select a representative square region from satellite images containing that crop within farm areas. This template captures the most common texture characteristics observed in these farmlands, aiding in texture matching and thus enabling the precise identification of specific crop types in the fields. Once the templates are matched, we employ carefully designed parametric rules to generate a variety of farm landscapes, from macro to micro scales.

5.1 Subcategory Division

Given all the segmented areas that are classified as fields, it is essential to represent them using distinct texture representations specific to farmland. To accomplish this, we leverage texture analysis techniques to analyze and match the textures present within individual patch areas, which is called texture matching in this work. By employing statistical measures, we extract relevant texture features that encapsulate crucial information regarding grayscale co-occurrence statistics, LBP descriptors, color distribution statistics, and other discriminative texture attributes. Subsequently, these extracted features are utilized for region-specific computations and analysis.

We select six category-specific templates for different types of field: soybean, wheat, corn, sugarcane, harvested land, and bare soil, as illustrated in Figure 4. These templates serve as representative visual patterns that capture the unique characteristics of each category. Each scene utilizes a subset of these templates based on its specific characteristics. We leverage the extracted feature values from the segmented areas and compute the cosine similarity between these features and the template features. This allows us to assign each segmented region to the most matched template, thereby identifying the corresponding farmland subcategory.

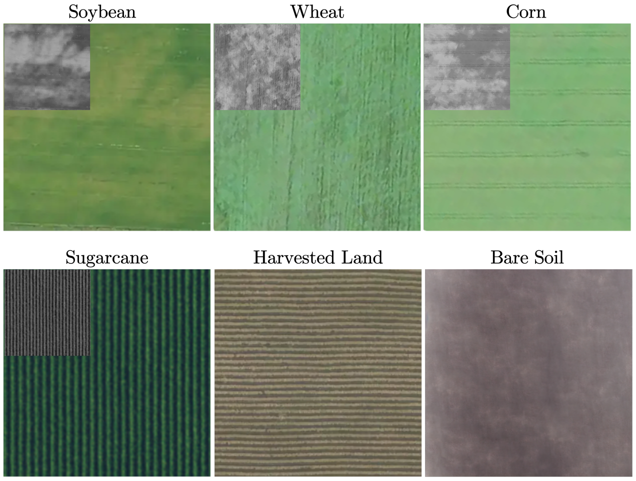


Fig. 4. The six category-specific templates. A vitality map is shown in the top-left corner. Note harvested land and bare soil templates do not have vitality maps.

5.2 Crop Layout

Pałubicki et al. [2022] proposed an ecosystem simulation method that models the self-organization patterns of plants in controlled environments. In agricultural settings, crop layout at the sowing stage is influenced by mechanical operations, while the distribution of nutrients and mutual shading among plants during growth are shaped by irrigation and terrain illumination. These result in crops being at different phenological stages and exhibiting self-organizing patterns in the field [Gao and Zhang 2021]. These manifest as distinctive striated textures in remote sensing images.

5.2.1 Directional field. We generate 2D strands from images as a guide for creating crop layout strands, thus simulating the natural distribution of crops in farmlands. Specifically, we first adopt the method in [Chai et al. 2012] to compute a 2D orientation map O_n from patches that are extracted using the farmland mask $\mathcal{M}_n$ from input satellite image $\mathcal{I}_s$, with an example illustrated in Figure 6(a). Then, non-maximum suppression is used to obtain a set of seed pixels, serving as a starting point for further strand sampling.

In the process of line segment tracing and sampling, we introduce two extensions to the original method in [Chai et al. 2012]. First, for the current sampling point p_i in an iteration, we extend an arc sampling approach to select the next endpoint, creating smooth strands. This is achieved by considering pixel points along the arc of a sector, centered around the given direction and extending α degrees on both sides of the center line, with a radius r. For each candidate point q_j along the arc, we extract the pixels passed through from the current point to the candidate point using the Bresenham algorithm [Pitteway and Watkinson 1980]. We then calculate a score based on the consistency of their orientations with the orientation at the current point as follows:

$$S(q_j) = \frac{1}{N}\sum_{k=1}^{N} I_{|\Delta\theta_{ijk}|\leq\lambda}\cdot\left(1-\frac{|\Delta\theta_{ijk}|}{2\cdot\alpha}\right), \tag{3}$$

where N is the total number of pixels traversed, $\Delta\theta_{ijk}$ is the directional difference between the current point p_i and the k-th pixel along the path to the candidate point q_j (as indicated in the direction field), and λ is an angular threshold. I is an indicator function that equals 1 when the inner condition holds, and 0 otherwise. Additionally, α accounts for the angular offset in the arc sampling, reducing bias towards paths with larger angular differences. Finally, we select the candidate point q_j with the highest score as the sampling point.

Second, we use curvature changes, instead of angular variations in [Chai et al. 2012], to determine when to terminate the sampling. The sampling process is terminated at a point where the discrepancy between its curvature and the average local curvature exceeds the designated threshold ϵ. We found this revised sampling strategy to be highly effective in all our experiments. For the results presented in this paper, the parameters $\lambda = \pi/16$, $\alpha = \pi/4$, and $\epsilon = 0.4$.

5.2.2 Generating dense strands. With the agricultural field boundary $\mathcal{B}_i$ extracted from the mask $\mathcal{M}_i$ and sparse 2D strands as input, our method generates dense strands representing well-distributed and logically-arranged crop planting areas in two steps. Initially, the sparse strands undergo a deformation process through iterative optimization, considering boundary constraints, local orientation constraints, and regularization terms. Subsequently, we introduce an iterative rule-based approach to generate dense strands within the agricultural mask, filling areas where the directions of the crop planting are less evident in the original image.

Strand deformation. Based on our observations on real-world farmlands, we derive two heuristic rules: First, crops planted close to contour boundaries tend to align more closely with the contours. Second, on a local scale, multiple rows of crops exhibit similar planting directions, resembling lines with similar curvatures. Therefore, in the optimization of strand deformation, we consider the following three constraints: (1) *Boundary constraints* to ensure that lines closer to the boundary contours are more aligned with the directions of the contours; (2) *Neighborhood constraints* to ensure that nearby lines with similar directions are more aligned with each other; (3) *Strand constraints* to ensure that line segments do not deviate excessively during deformation and maintain smooth curvatures without abrupt changes.

Then, we define energy functions for the above constraints as:

$$\begin{cases} E_B(p_i) = \beta\cdot(1-|d(p_i)\cdot d(\mathcal{B}(p_i))|),\\ E_N(p_i) = (1-\beta)\sum_{n\in N(p_i)}(1-|d(p_i)\cdot d(n)|),\\ E_R(p_i) = |p_i-\bar{p}_i|^2+\omega_1\,|p_{i-1}-2p_i+p_{i+1}|^2, \end{cases}$$

where β is used to balance the deformation tendencies of strands relative to their boundary distances. In our experiments, $\beta = \max(1-\frac{m}{l}, 0)$. Here, m represents the distance of a point p_i from the boundary, which can be calculated using grid-based image analysis as depicted in Figure 5(a); and the parameter l controls the influence range of the boundary constraint. $d(p)$ is the normalized direction vector of point p, $N(p)$ encompasses nearby points not on the same strand but with similar orientations to p, and $\mathcal{B}(p)$ identifies the nearest point on the boundary to p. The strands are then deformed by minimizing the following overall energy function:

$$E = \sum_i \omega_2(E_B(p_i)+E_N(p_i))+E_R(p_i), \tag{4}$$

using the Adam optimizer. After 50 iterations, this optimization results in strands that exhibit more rational and unified orientations. In our experiments, the weights were empirically set to $\omega_1 = 10$ and $\omega_2 = 50$. Figures 5(b) and 5(c) show the 2D strands before and after optimization, respectively.

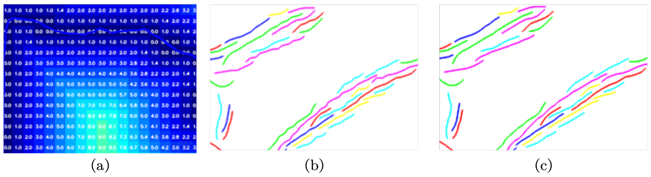


Fig. 5. 2D strands optimization: (a) visualizes the calculated distances from the boundary; (b) illustrates the 2D strands before optimization; and (c) shows the result after optimization.

Iterative sampling. The next step is to generate fully covered and smooth strands. The seed points generated by Chai et al. [2012] tend to be concentrated in areas with distinct textural features, often resulting in an incomplete coverage of the farmland. To address this, we introduce an iterative approach combined with rule-based generation for comprehensive sampling. Specifically, optimized strands are first placed in a reference set, followed by the addition of sampled strands at each round into the sampled set.

Our iterative sampling strategy repeats the following steps: (1) The round of sampling is concluded by selecting a reference line within the farmland mask. Alternatively, designating a starting coordinate position initiates subsequent steps. (2) Find the initial sampling direction in a window centered on this point: first consider the direction of the nearest point on the reference strand within range; otherwise consider the direction of the nearest point on the sampled strand within range. (3) Generate all initial points at equal intervals along two normal directions until they extend beyond the boundary or conflict with previously sampled strands. (4) Trace along the initial direction to generate all strands.

While tracing each strand, we adhere to the following rules: 1) The search direction at each time is the average of the directions of the nearest points on different reference strands nearby. This is to ensure similarity to previous search directions. 2) If there exist sampled strands nearby during tracing, their curvatures are considered to continue the tracing. 3) Terminate the tracing when a strand crosses a reference line segment with a large directional deviation or goes beyond the boundary.

Figure 6(b) shows the 2D strands generated through our iterative sampling, plausibly covering a farmland area. By referencing the actual crop planting spacing and interpolating it into a high-resolution mask, we can arrange crops to align their planting directions with the input satellite image. Figure 6(c) shows a close-up view of this alignment, showing the rows of crops in a highly realistic layout.

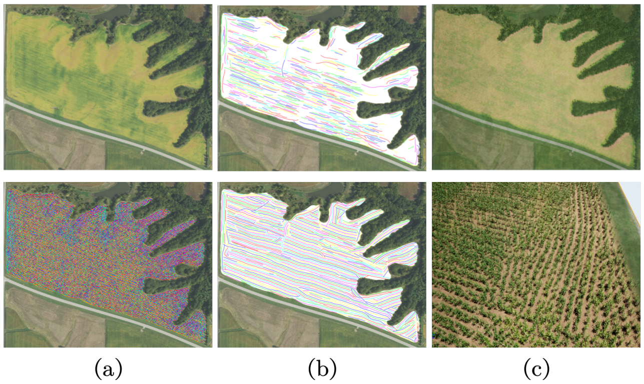


Fig. 6. (a): An initial 2D orientation map (bottom) is extracted using the farmland mask from an input satellite image (top). (b): Initial 2D strands (top) and sampled 2D strands by our approach (bottom). (c): Aerial view (top) and close-up view (bottom) of the reconstructed field.

5.3 Crop Types, Growth Stages, and Vitality

Visual attributes of farmlands are intrinsically linked to the types of crops cultivated and their respective growth stages. To quantify these characteristics, we introduce a set of parameters, notably the growth stage parameter γ, to delineate various stages of crop growth. An analysis of reference images enables us to determine the crop type α. Subsequently, inspired by the work of [Liu et al. 2017a] in distinguishing healthy plants from aging ones using color differentiation, we use a contrast color-to-grayscale algorithm [Grundland and Dodgson 2007] to transform the color information in the satellite image into grayscale values to derive a vitality map indicative of the crop's health. Examples of these vitality maps are provided in Figure 4. Considering the consistent color transitions displayed by crops throughout their growth cycles, the grayscale values serve as an effective metric for growth stage assessment.

Upon obtaining the vitality map, we sample the crop's planting position (x, y) to ascertain the crop's growth stage and compute its vitality parameter, denoted as $v = \text{Vitality}(x, y)$. The *Vitality* function, defined as the average of all pixel values within the crop coverage area, offers a quantitative depiction of the crop's vitality at that region. Figure 7 depicts the interplay between our growth stage parameter γ and vitality parameter v via the progression of the corn plant growth. As illustrated in this figure, by appropriately partitioning the growth stages for the crop and modulating the vitality parameter v, we can adeptly simulate the evolution of crop size and leaf count during various growth stages.

6 Environmental Elements Generation

6.1 Trees Placement

In real world agricultural scenarios, trees are often distributed along specific pathways, with varied distances between trees, especially in the vicinity of roads and fields (see Figure 8(a)). Considering the challenge of directly recognizing such arrangements from satellite images, we design procedural rules for the generation of such trees alongside roads and fields. These trees are placed based on a distance parameter h, which allows for variable intervals between trees. The challenge intensifies when dealing with tree masks classified as *clustered*. These masks, obtained via the region segmentation, are meant to cover forested regions, yet numerous areas within the mask still lack trees (see Figure 8(b)).

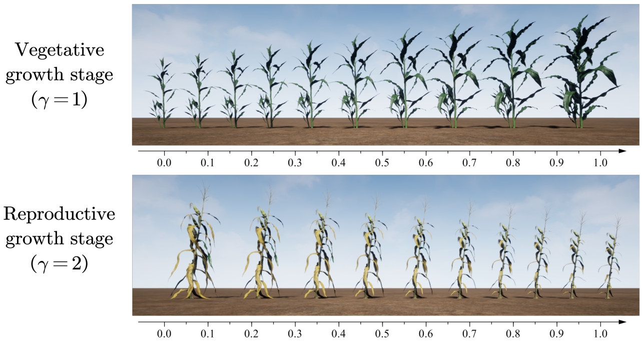

Vegetative growth stage ($\gamma = 1$)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Reproductive growth stage ($\gamma = 2$)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

Fig. 7. The two growth stages of corn controlled by vitality. The absolute v values are segmented into two intervals: the first interval (top) where $v \in (0, 0.3)$, and the second interval (bottom) where $v \in (0.3, 1)$. These intervals correspond to different growth stages. The X-axis represents the normalized vitality value v.

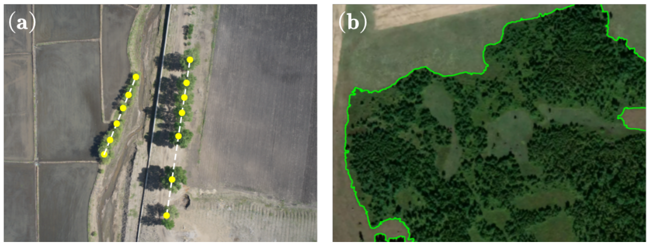


Fig. 8. Two tree distribution patterns. (a) Trees are in the vicinity of roads and fields; (b) substantial portions within a segmented forest region still lack trees.

To faithfully replicate the distribution of trees, we design a method to obtain the probabilistic distribution of trees from a satellite image, and subsequently scatter points in a 3D scene to arrange trees. Our method is inspired by the concept of statistical tree distribution [Emilien et al. 2015]. Specifically, we first employ the K-means clustering algorithm to partition the color space within the forest mask into k clusters. For each cluster, we calculate the pixel count and compute r_i, which represents the relative occurrence probability of color cluster c_i in the forest mask. To adjust the probability of the point distribution within each color region (e.g., assigning lower weights to background color regions), we propose a novel probability value $s_i = w_i r_i$, where w_i is the weight associated with color cluster c_i. Subsequently, we normalize all s_i values to ensure the sum of the probabilities across all color clusters is unity.

The tree distribution generation process encompasses two steps. Initially, for each of t trees, we draw a color cluster index i from a multinomial distribution, $\mathit{Multinomial}(S)$, where $S = s_1, s_2, \ldots, s_k$. This step is represented as:

$$i_{[j]} \sim \mathit{Multinomial}(S), \quad j = 1, \ldots, t, \tag{5}$$

The reason for using a multinomial distribution is that it allows trees to appear in multiple areas, reflecting real-world scenarios where each area or color cluster carries a varying degree of importance, as represented by their respective weights w_i.

Next, we randomly select a plant position (x, y) within the 3D scene region corresponding to the color cluster $c_{i_{[j]}}$, considering the bounding boxes of previously placed trees. Specifically, for every new tree, we calculate its bounding box and ensure that it does not overlap with the bounding boxes of any previously placed trees. This step can be described as:

$$(x, y) \sim \mathrm{Uniform}((x, y)|(x, y) \in C_{i_{[j]}}), \quad j = 1, \ldots, t, \tag{6}$$

where $C_{i_{[j]}}$ denotes the set of pixel positions corresponding to color cluster $c_{i_{[j]}}$, and t is the planned number of trees to be placed in the scene. Through this method, we can generate random tree positions that conform to the color distribution and weight probabilities. This is demonstrated by the results shown in Figure 9, using the regions in Figure 8(b) as an example.

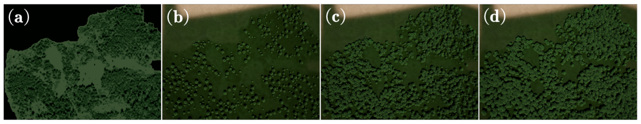


Fig. 9. (a) The result of color clustering. We set the ground color class weight w_1 to 0 and the shadow class weight w_2 to 0.1. The weights for other color classes representing trees, w_3 and w_4, are both set to 0.45. (b)-(d) show the results of progressively increasing the number of trees in the placement.

6.2 Road Network Generation

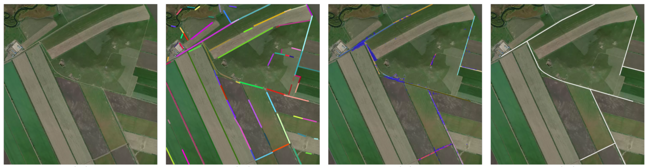
In satellite images, roads are often inconspicuous, and the availability of training data for road detection is limited. Moreover, many unpaved roads are not easily visible. As a result, extracting road masks directly from a satellite image is a challenging problem. Existing neural network structures for road mask extraction may suffer from discontinuities and misclassifications [Liu et al. 2017b]. Furthermore, to model road networks in a 3D scene, having just road masks is insufficient. Instead, we need input data that define the shapes and positions of the roads through a series of path points.

Recall in Section 4, we have obtained background masks that exclude field and tree areas. Our task is to extract a parametric road network within these background regions. To achieve this, we introduce a novel road network generation algorithm based on weak hints to extract the parameters of the road PLM.

Initially, from an input satellite image $\mathcal{I}_s$, we extract masks representing hint positions for roads using our trained DeeplabV3+ model [Chen et al. 2018]. Subsequently, we extract line segments that can potentially be roads, using the aforementioned 2D strand generation method (described in Section 5.2.1). Then, we cast rays from each segment's endpoints to determine whether they intersect with another segment or the rays casted from another segment. If the condition is met, we merge the initial segment with the nearest one. When the angle difference between two segments does not exceed a threshold (e.g., 5 degrees), we use PCA to merge them into a single line, categorizing the resulting points as *Linear*. By contrast, when the angle difference exceeds the threshold, we use the centripetal Catmull-Rom splines algorithm [Catmull and Rom 1974] to generate a natural transition curve between the segments, which results in *Curvy* type points. During this process, the tangent vector at each curvy point is computed through the derivative of the spline function. Additionally, if the intersection point falls within the middle of another line segment, the line segments are directly connected to that point, and the point is classified as a *Junction*. Once all segment endpoints have undergone ray casting and no new segments can be merged, we obtain the road network (see Figure 10(d)), where each point consists of three parameters: coordinates, tangent vector, and point type. Due to the ambiguity of roads in satellite images, our method is more effective for main roads, which may result in overlooking potential roads between fields when generating hints. However, interactive manual specification is also supported through the interactive features of the SAM segmentation framework [Kirillov et al. 2023], allowing users to specify roads such as main roads or dirt paths, and generate roads at different levels of detail based on the desired output. Algorithm 1 describes our road network generation algorithm.

ALGORITHM 1: Road Network Generation Algorithm

```
Input  : A satellite image $\mathcal{I}_s$
Output : Road networks
Extract potential road line segments;
Initialize clustering with the longest line segment;
while there are unclustered line segments do
    Cluster line segments with angle deviation $\leq$ 5 degrees and projection distance below threshold;
end
Compute principal axes of the clusters using PCA;
Generate road mask seed hints using DeepLabV3+ or SAM;
Extend search from hint positions;
while there are mergable line segments do
    Form junctions or use PCA or Catmull-Rom splines to merge line segments;
    Compute the tangent vector at each point on the curve;
end
Build the road networks;
```



(a) (b) (c) (d)

Fig. 10. An example of road network generation. (a) An input satellite image; (b) the extracted potential road segments; (c) the generated road seed hints, as illustrated by blue masks; and (d) the final generated road network.

7 Interactive Editing

Interactive editing allows users to flexibly adjust and control various factors, including crop types, planting densities, and crop growth stages. This flexibility ensures that users can conveniently customize and control the modeled farm scene in various ways, such as changing labels and adjusting parameter values. Furthermore, alterations can be made to plants themselves, such as modifying the growth pattern of leaves or adjusting the proportions of different morphological models. Our system enables efficient generation of various farmland scenes. Figure 11(a)-(f) shows the results of user editing by our system.

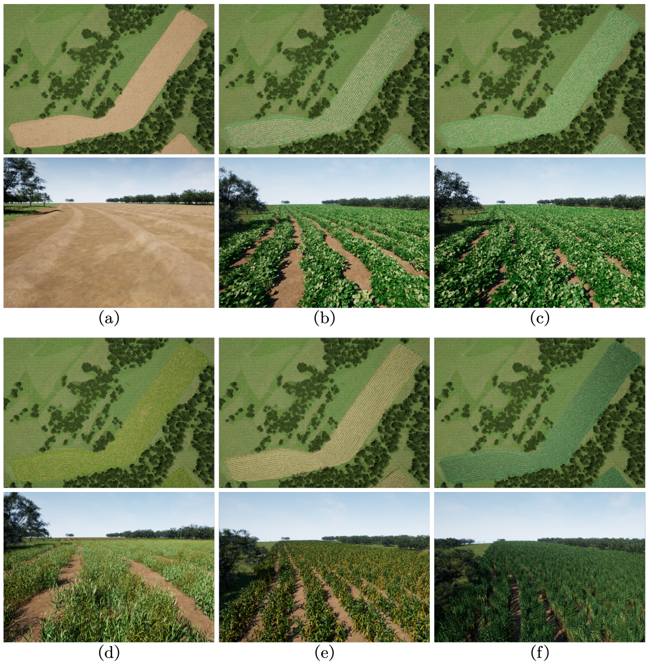


Fig. 11. User-controllable field editing: the bird's eye view (top), and a close-up view (bottom). (a) The reconstructed harvested field; (b) a field replaced with a soybean template; (c) a soybean field with an adjusted density; (d-f) a field replaced with wheat, corn, and sugarcane templates, respectively.

Users can easily add new crop types. For example, when adding cotton, we first label 50 cotton fields to create a supplementary dataset for fine-tuning the MLP classifier. Next, we select a representative template from satellite images of cotton fields for precise texture matching. We then incorporate two models for the boll-opening and budding stages of cotton respectively, establishing corresponding vitality thresholds. By assessing plant vitality, our method selects the appropriate models for different growth stages. Figure 12 shows the result of cotton modeling. This showcases the extensibility of our template-based approach, demonstrating its adaptability to new crops' unique growth patterns and visual characteristics.

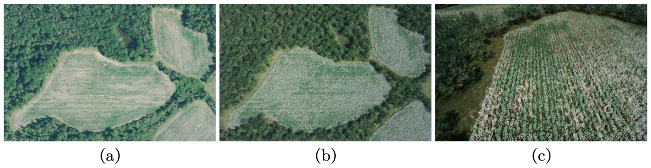


Fig. 12. Cotton field modeling results by our method. (a) An input satellite image. (b) A rendering of the reconstructed cotton scene from the bird's eye view. (c) A close-up view of the reconstructed cotton scene.

Table 1. Average computational time statistics of our approach

Stage	Seg.	Field	Tree	Road	Rendering
Small Scenes (s)	58.78	305.65	156.98	242.53	4.38
Medium Scenes (s)	86.20	1315.02	162.88	329.21	14.84
Large Scenes (s)	134.58	2201.46	167.79	447.78	96.34

The satellite images used in this work primarily come from Asia. However, since farmland in different regions can appear quite different in satellite images, several factors must be considered when modeling farm scenes in new regions. Taking Africa as an example, the primary crops grown there differ from those in Asia, so corresponding 3D crop models need to be designed. Additionally, due to differences in lighting, climate, and atmospheric conditions, the classifier will require substantial new data for training. The vegetation covers and colors in African farmland may change significantly with seasons and weather, so even for the same crops, the growth phenological stages may vary greatly between regions, necessitating a recalibration of vitality thresholds. Furthermore, the types and distribution patterns of common trees differ between regions, requiring adjustments to the distribution weight values (w_i). Lastly, road patterns may also vary, so the seed hint generation network will need some fine-tuning.

8 Results and Evaluations

We implemented our approach within the Unreal Engine 5. Our data source comprises satellite images downloaded from Mapbox (*https://www.mapbox.com/*), with a maximum resolution of 0.5m per pixel. We reported the average computational time statistics of our approach, including its major modules, as shown in Table 1. We classified test scenes into small-scale (under 1 sq km), medium-scale (between 1 to 10 sq km), and large-scale (over 10 sq km). All reported results were obtained on an off-the-shelf Desktop PC with an Intel(R) Core i7-11700 CPU, 32GB memory, and an NVIDIA GeForce GTX 3080 GPU.

8.1 Qualitative and Quantitative Evaluations

Qualitative evaluations. We conducted qualitative evaluations on 16 different farm scenes. Specifically, based on an input satellite image, we first reconstructed its corresponding 3D farm scene using our method, and then rendered a top-down view. We directly compared the rendered 2D view with the input satellite image. As shown in Figure 13, given an input satellite image in (a), our method reconstructed its 3D scene in (c), which achieved closer alignments with the structural features in the input satellite image and exhibits richer multiscale details than the result by a baseline method (b). This baseline method utilized the Unreal Engine's built-in vegetation system to scatter plants randomly within classified crop and forest areas, while enhancing vegetation layout diversity through jitter operations and model replacement functionalities. Figure 11(a)-(f) shows the impact of our defined templates in a close-up view, introducing rich variations to both the layout and characteristics of the reconstructed farmland.

Table 2. Quantitative evaluation results. Best results are highlighted in bold.

Method	**PSNR↑**	**MS-SSIM↑**	**LPIPS↓**
Baseline	17.420	0.3943	0.6851
Ours	**18.289**	0.4112	**0.5430**
Ours w/o crop layout	17.924	**0.4200**	0.6609
Ours w/o crop vitality	17.837	0.3946	0.5581
Ours w/o trees	18.087	0.4059	0.5588

We conducted another qualitative evaluation to validate the important role of our PLMs in our approach. As shown in Figure 13, the crop layout PLM (Figure 13(d)) aligns the crop orientations more accurately with actual farming practices, adding richness and diversity to the scenes. The crop vitality PLM (Figure 13(e)) imparts a more authentic self-organizing pattern to the farmland. Additionally, the tree PLM (Figure 13(f)) leads to a natural distribution of forests, thus increasing the overall realism of the rendered scenes.

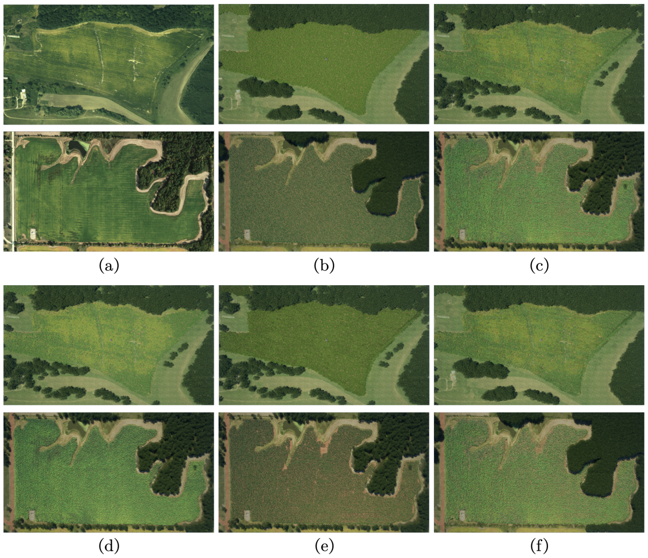

(a) (b) (c)
(d) (e) (f)

Fig. 13. (a) Input satellite images; (b) results by a baseline method; (c) results by our approach; (d) results by ours w/o crop layout; (e) results by ours w/o crop vitality; and (f) results by ours w/o trees.

Quantitative evaluations. We also conducted quantitative evaluations by directly measuring the pixel differences between the input satellite images and the rendered 2D views of the reconstructed 3D scenes. We selected the following widely-used measures: Peak Signal-to-Noise Ratio (PSNR), Multi-Scale Structural Similarity Index (MS-SSIM), and Learned Perceptual Image Patch Similarity (LPIPS). Table 2 shows the quantitative measures obtained in our evaluations. Specifically, our approach achieved the highest PSNR score of 18.289, indicating superior fidelity and accuracy in content reconstruction, particularly compared to the baseline, which scored 17.420. Although our method "w/o crop layout" strategy slightly outperformed our method in terms of the MS-SSIM measure with a score of 0.4200, our method also showed robust capabilities in maintaining image structure and fidelity with a close MS-SSIM score of 0.4112. Lastly, our method excelled in the LPIPS metric with a score of 0.5430, highlighting its significant advantage in perceptual quality, especially against the baseline.

Overall, our approach significantly outperformed both the baseline and various ablation versions of our method in image quality, particularly in content fidelity and perceptual quality, as reflected in the PSNR and LPIPS scores. Our method also demonstrated strong performance in maintaining image structure, as reflected in the near-best MS-SSIM score. These quantitative evaluation results confirmed our method's advantages in scene reconstruction and image quality preservation, producing a high similarity between top-down rendered views and original satellite images.

8.2 User Studies

To assess the effectiveness of our method, we conducted several user studies to evaluate the realism of 3D farm scenes modeled by different approaches including our approach.

Comparison with baseline. We invited 51 participants from Prolific (*https://app.prolific.com/*), consisting of 30 females and 21 males with the average age of 31.05 ± 10.10, to assess and compare the visual realism of the farm scenes by our method and by the baseline method, based on the same satellite image input. We used a total of 16 different scenes for this study. Each stimulus included an input satellite image, a rendered overhead view of the 3D scene (modeled by our approach or by the baseline), and a rendered animation of 3D navigation of the scene. All of them were placed on one screen side-by-side. Participants were asked to rate the realism of the 3D scene using a 0 to 10 point after carefully watching each stimulus. As shown in Figure 14, our method achieved a significantly higher average score (6.43) than the baseline method (4.96). The calculated p-value from t-test is less than 0.001, which indicates our method is statistically significantly better than the baseline in this comparison. In addition, as shown in Figure 14, the majority of ratings for our method fall within the high-value (7-10) range, with only a small fraction in the low-value (0-3) range. In contrast, the ratings for the baseline are approximately evenly distributed across all three ranges, with a slight majority in the mid-value (4-6) range.

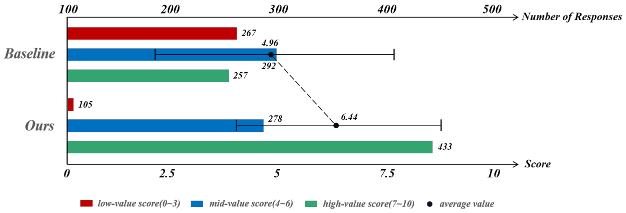


Fig. 14. The comparison study results between our method and the baseline method. Besides the average scores and the standard deviations, we also visualize the number of responses in low, medium, and high score ranges.

Ablation study. We invited 53 participants, consisting of 25 females and 28 males with the average age of 30.96 ± 9.01, to rate the realism of the overhead views of the 3D scenes generated by four different versions of our method: “ours" (full version), “ours w/o crop layout," “ours w/o crop vitality," and “ours w/o trees," given the same satellite image input. 16 different satellite images were used in this study. The average scores ± the standard deviations for the four versions are 6.72 ± 1.90 (“ours"), 5.44 ± 2.19 (“ours w/o crop layout"), 5.58 ± 2.06 (“ours w/o crop vitality"), and 5.81 ± 2.20 (“ours w/o trees"), respectively. This shows that the crop layout PLM has the most significant impact on result quality, followed by crop vitality PLM, and lastly, tree PLM. Note that the texture discrepancy between the scenes rendered by our method and real farm environments may have moderately impacted the scores.

Expert user study. We also invited 5 expert users who have over three years of modeling experience using the Unreal Engine to evaluate our approach. Each of them was asked to model a 3D farmland using both our approach and the manual modeling method (described below), based on a farm scene satellite image. In total, 5 different satellite images were used. In the manual modeling method, they initially utilized a farmland blueprint from the Unreal Engine store, which generates crops planted in multiple straight lines within a rectangular area, to model the orderly farmland areas. Subsequently, using the engine-provided brush tools, they modeled crops planted along curves and crops of different colors within the farmlands to replicate the self-organizing patterns of the crops. To ensure a fair comparison, they were allowed to use the masks extracted from our semantic segmentation step to aid modeling when using the manual approach. Finally, we asked them to respond to seven post-experiment questions using a 5-point Likert scale (1 - “strongly disagree" to 5 - “strongly agree"): (a) Is the system easy to learn? (b) Is the system easy to use? (c) Do you prefer to use this system than the alternative? (d) Does the system speed up the scene generation process? (e) Does the result meet your expectations? (f) Does it provide sufficient user control parameters? and (g) does it provide a rich variety of detail changes? We plot the obtained responses in Figure 15, which shows that, besides user friendliness, our method can significantly facilitate the generation of detailed and realistic farm scenes for experienced users.

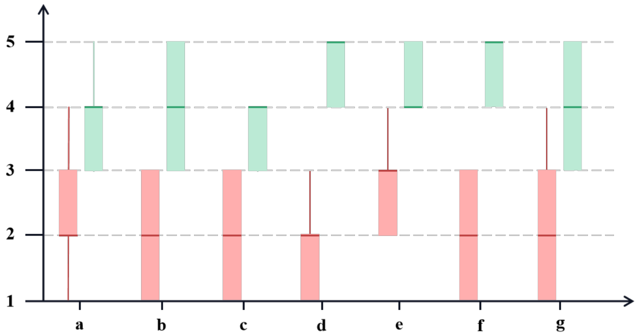


Fig. 15. Comparison of expert user ratings between traditional manual modeling method (red bars) and our method (green bars). Questions (a)-(g) are described in the Section 8.2. The horizontal lines represent the median scores, and the vertical lines indicate the ranges of the scores.

We also recorded the time they used for the modeling task with both methods, as shown in Table 3. Our method consumed less time than the manual modeling method in all 5 scenes. In addition, the time used for our method is relatively stable regardless of the scene size; on the contrary, the time used for the manual modeling method is approximately linearly correlated with the scene size. Since the most time-consuming part of our method is the extraction of farmland parameters and interactive sampling, the time required by our method is closely related to the number of farmland blocks processed rather than the size of the scene. This feature is particularly useful for large-scale farm scene modeling, as manual methods may face significantly increased time demands due to scale expansion. Figure 16 presents a visual comparison of the modeling results for five scenes using the two different methods. The comparison clearly demonstrates that our method produces farmlands with a more natural layout and appearance than the manual modeling method. A close-up comparison of Scene #1 (the first row in Figure 16) is shown in Figure 17, indicating that our method can achieve a more natural crop arrangement with richer details. For the remaining comparison results, please refer to the supplemental material.

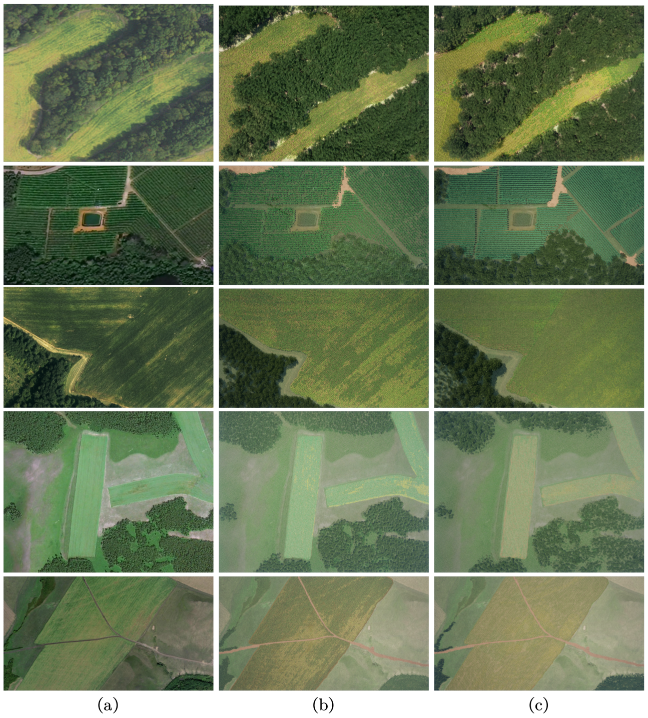


Fig. 16. Comparison of modeling results for expert users using two different methods. (a) Input satellite images; (b) results by our method; (c) manual modeling results by expert users.

9 Discussion and Conclusion

We present a practical framework to faithfully reconstruct virtual farm scenes using remote sensing imagery data. Besides generating virtual farm scenes that accurately reflect real-world farmland distributions, we design a set of PLMs to decompose and define various

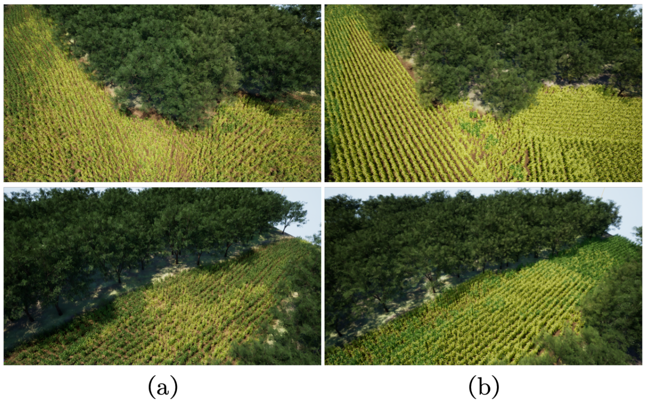


Fig. 17. Rendered close-up view comparison of two methods for scene #1 in Figure 16. (a) Rendered bird-eye and close-up views of the scene (reconstructed by our method); (b) Rendered bird-eye and close-up views of the scene (modeled by the expert manual method).

Table 3. Modeling time comparison between expert manual method and our method

Scene No.	Scene Size (m^2)	Ours(s)	Expert manual (s)
Scene 1	120 × 90	834.92	1529.20
Scene 2	210 × 170	770.19	1143.89
Scene 3	500 × 360	625.38	2071.16
Scene 4	1280 × 730	805.21	2541.91
Scene 5	1480 × 820	687.51	3312.53

farm scene patterns, and further extract the parameter values of the PLMs to faithfully recreate multi-scale farm scenes.

Despite generating encouraging results, our approach has certain limitations. First, our method is designed for specific types of farm scenes in outdoor environments. It struggles to model certain elements, such as large irrigation systems, buildings, and wind turbines. Additionally, if the farmland captured by the input satellite image is too small or lacks clear texture features, the effectiveness of our method may be compromised. Second, our current method only defines six types of field templates, which do not comprehensively cover diverse types of real-world fields. As crop categories increase, more data points will be needed to improve classification accuracy. If the number of plant species significantly expands, it may be necessary to enhance the classification module by incorporating advanced classification algorithms in the remote sensing field or deep vision models. Third, satellite images may exhibit blurred patterns due to shooting conditions and resolution issues, or inconsistent textures due to differences in crop phenological stages. Using traditional non-maximum suppression methods to automatically generate initial sampling points can result in unnatural line segments, affecting the modeling quality. Therefore, the extraction of farmland layouts cannot achieve optimal results through full automation and requires the assistance of interactive sampling. Lastly, since we use satellite images as color maps for color shifting, when houses, rocks, or other objects appear in input satellite images, the corresponding modeled grassland colors may exhibit visual abnormalities.

Future work can focus on linking extracted vitality values with the actual physical state of plants, which requires additional data sources such as near-infrared and multispectral imagery. Integrating video data captured by drones could further enhance the accuracy and detail of crop modeling. Additionally, leveraging a combination of these multi-source datasets with deep learning techniques could enable generative modeling of large-scale, unbounded, three-dimensional farmland scenes. Also, current plant models do not consider interactions between the growth process and the environment. An intriguing future direction could be to extend our framework to simulate the evolution of a complete farmland ecosystem.

Acknowledgments

This work was in part supported by the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDA0450203, the National Natural Science Foundation of China under Grant No. 62172392, and the Innovation Research Program of ICT under Grant No. E261070. Zhigang Deng was in part supported by US NSF IIS-2005430.

References

Daniel G. Aliaga, undefinedlke Demir, Bedrich Benes, and Michael Wand. 2016. Inverse procedural modeling of 3D models for virtual worlds. In *ACM SIGGRAPH 2016 Courses* (Anaheim, California) *(SIGGRAPH '16)*. Association for Computing Machinery, New York, NY, USA, Article 16, 316 pages. https://doi.org/10.1145/2897826.2927323

Daniel G. Aliaga, Carlos A. Vanegas, and Bedrich Benes. 2008. Interactive Example-Based Urban Layout Synthesis. *ACM Transactions on Graphics* 27, 5 (2008), 1–10. https://doi.org/10.1145/1409060.1409113

Bedřich Beneš, Michel Abdul Massih, Philip Jarvis, Daniel G. Aliaga, and Carlos A. Vanegas. 2011. Urban Ecosystem Design. In *I3D '11: Symposium on Interactive 3D Graphics and Games* (San Francisco California). ACM, 167–174. https://doi.org/10.1145/1944745.1944773

Martin Bokeloh, Michael Wand, and Hans-Peter Seidel. 2010. A Connection between Partial Symmetry and Inverse Procedural Modeling. *ACM Transactions on Graphics* 29, 4 (2010), 1–10. https://doi.org/10.1145/1778765.1778841

Edwin Catmull and Raphael Rom. 1974. A Class of Local Interpolating Splines. *Computer Aided Geometric Design* (1974), 317–326. https://doi.org/10.1016/B978-0-12-079050-0.50020-5

Menglei Chai, Lvdi Wang, Yanlin Weng, Yizhou Yu, Baining Guo, and Kun Zhou. 2012. Single-View Hair Modeling for Portrait Manipulation. *ACM Transactions on Graphics* 31, 4 (2012), 1–8. https://doi.org/10.1145/2185520.2185612

Guoning Chen, Gregory Esch, Peter Wonka, Pascal Müller, and Eugene Zhang. 2008. Interactive Procedural Street Modeling. *ACM Transactions on Graphics* 27, 3 (2008), 1–10. https://doi.org/10.1145/1360612.1360702

Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *Computer Vision – ECCV 2018*, Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). Vol. 11211. Springer International Publishing, 833–851. https://doi.org/10.1007/978-3-030-01234-2_49

Tao Cheng, Xusheng Ji, Gaoxiang Yang, Hengbiao Zheng, Jifeng Ma, Xia Yao, Yan Zhu, and Weixing Cao. 2020. DESTIN: A New Method for Delineating the Boundaries of Crop Fields by Fusing Spatial and Temporal Information from WorldView and Planet Satellite Imagery. *Computers and Electronics in Agriculture* 178 (2020), 105787. https://doi.org/10.1016/j.compag.2020.105787

Ilke Demir, Daniel G. Aliaga, and Bedrich Benes. 2016. Proceduralization for Editing 3D Architectural Models. In *Proc. of 2016 Fourth International Conference on 3D Vision (3DV)* (Stanford, CA). IEEE, 194–202. https://doi.org/10.1109/3DV.2016.28

Arnaud Emilien, Ulysse Vimont, Marie-Paule Cani, Pierre Poulin, and Bedrich Benes. 2015. WorldBrush: Interactive Example-Based Synthesis of Procedural Virtual Worlds. *ACM Transactions on Graphics* 34, 4 (2015), 1–11. https://doi.org/10.1145/2766975

Vivien Sainte Fare Garnot and Loic Landrieu. 2021. Panoptic Segmentation of Satellite Image Time Series with Convolutional Temporal Attention Networks. In *Proc. of 2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (Montreal, QC, Canada). IEEE, 4852–4861. https://doi.org/10.1109/ICCV48922.2021.00483

Feng Gao and Xiaoyang Zhang. 2021. Mapping Crop Phenology in Near Real-Time Using Satellite Remote Sensing: Challenges and Opportunities. *Journal of Remote Sensing* 2021 (2021). https://doi.org/10.34133/2021/8379391 arXiv:https://spj.science.org/doi/pdf/10.34133/2021/8379391

Mark Grundland and Neil A. Dodgson. 2007. Decolorize: Fast, Contrast Enhancing, Color to Grayscale Conversion. *Pattern Recognition* 40, 11 (2007), 2891–2896. https://doi.org/10.1016/j.patcog.2006.11.003

Paul Guerrero, Miloš Hašan, Kalyan Sunkavalli, Radomír Měch, Tamy Boubekeur, and Niloy J. Mitra. 2022. MatFormer: A Generative Model for Procedural Materials. *ACM Transactions on Graphics* 41, 4 (2022), 1–12. https://doi.org/10.1145/3528223.3530173

Jianwei Guo, Haiyong Jiang, Bedrich Benes, Oliver Deussen, Xiaopeng Zhang, Dani Lischinski, and Hui Huang. 2020. Inverse Procedural Modeling of Branching Structures by Inferring L-Systems. *ACM Transactions on Graphics* 39, 5 (2020), 1–13. https://doi.org/10.1145/3394105

Éric Guérin, Julie Digne, Éric Galin, Adrien Peytavie, Christian Wolf, Bedrich Benes, and Benoît Martinez. 2017. Interactive Example-Based Terrain Authoring with Conditional Generative Adversarial Networks. *ACM Transactions on Graphics* 36, 6 (2017), 1–13. https://doi.org/10.1145/3130800.3130804

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *Proc. of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Las Vegas, NV, USA). IEEE, 770–778. https://doi.org/10.1109/CVPR.2016.90

Rokgi Hong, Jinseok Park, Seongju Jang, Hyungjin Shin, Hakkwan Kim, and Inhong Song. 2021. Development of a Parcel-Level Land Boundary Extraction Algorithm for Aerial Imagery of Regularly Arranged Agricultural Areas. *Remote Sensing* 13, 6 (2021), 1167. https://doi.org/10.3390/rs13061167

Yiwei Hu, Julie Dorsey, and Holly Rushmeier. 2019. A Novel Framework for Inverse Procedural Texture Modeling. *ACM Transactions on Graphics* 38, 6 (2019), 1–14. https://doi.org/10.1145/3355089.3356516

Yiwei Hu, Paul Guerrero, Milos Hasan, Holly Rushmeier, and Valentin Deschaintre. 2022a. Node Graph Optimization Using Differentiable Proxies. In *SIGGRAPH '22: Special Interest Group on Computer Graphics and Interactive Techniques Conference* (Vancouver BC Canada). ACM, 1–9. https://doi.org/10.1145/3528233.3530733

Yiwei Hu, Chengan He, Valentin Deschaintre, Julie Dorsey, and Holly Rushmeier. 2022b. An Inverse Procedural Modeling Pipeline for SVBRDF Maps. *ACM Transactions on Graphics* 41, 2 (2022), 1–17. https://doi.org/10.1145/3502431

Maxwell Jong, Kaiyu Guan, Sibo Wang, Yizhi Huang, and Bin Peng. 2022. Improving Field Boundary Delineation in ResUNets via Adversarial Deep Learning. *International Journal of Applied Earth Observation and Geoinformation* 112 (2022), 102877. https://doi.org/10.1016/j.jag.2022.102877

Tom Kelly, John Femiani, Peter Wonka, and Niloy J. Mitra. 2017. BigSUR: Large-Scale Structured Urban Reconstruction. *ACM Transactions on Graphics* 36, 6 (2017), 1–16. https://doi.org/10.1145/3130800.3130823

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment Anything. In *Proc. of the IEEE/CVF International Conference on Computer Vision 2023*. 4015–4026.

Ares Lagae, Sylvain Lefebvre, George Drettakis, and Philip Dutré. 2009. Procedural Noise Using Sparse Gabor Convolution. *ACM Transactions on Graphics* 28, 3 (2009), 1–10. https://doi.org/10.1145/1531326.1531360

Bosheng Li, Jacek Kałużny, Jonathan Klein, Dominik L. Michels, Wojtek Pałubicki, Bedrich Benes, and Sören Pirk. 2021. Learning to Reconstruct Botanical Trees from Single Images. *ACM Transactions on Graphics* 40, 6 (2021), 1–15. https://doi.org/10.1145/3478513.3480525

Konstantinos Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis. 2018. Machine Learning in Agriculture: A Review. *Sensors* 18, 8 (2018), 2674. https://doi.org/10.3390/s18082674

Jian Liu, Qiming Qin, Jun Li, and Yunpeng Li. 2017b. Rural road extraction from high-resolution remote sensing images based on geometric feature inference. *ISPRS International Journal of Geo-Information* 6, 10 (2017), 314.

Shouyang Liu, Fred Baret, Mariem Abichou, Fred Boudon, Samuel Thomas, Kaiguang Zhao, Christian Fournier, Bruno Andrieu, Kamran Irfan, Matthieu Hemmerlé, and Benoit De Solan. 2017a. Estimating Wheat Green Area Index from Ground-Based LiDAR Measurement Using a 3D Canopy Structure Model. *Agricultural and Forest Meteorology* 247 (2017), 12–20. https://doi.org/10.1016/j.agrformet.2017.07.007

Yanchao Liu, Jianwei Guo, Bedrich Benes, Oliver Deussen, Xiaopeng Zhang, and Hui Huang. 2021. TreePartNet: Neural Decomposition of Point Clouds for 3D Tree Reconstruction. *ACM Transactions on Graphics* 40, 6 (2021), 1–16. https://dl.acm.org/doi/10.1145/3478513.3480486

Jiang Long, Mengmeng Li, Xiaoqin Wang, and Alfred Stein. 2022. Delineation of Agricultural Fields Using Multi-Task BsiNet from High-Resolution Satellite Images. *International Journal of Applied Earth Observation and Geoinformation* 112 (2022), 102871. https://doi.org/10.1016/j.jag.2022.102871

Miłosz Makowski, Torsten Hädrich, Jan Scheffczyk, Dominik L. Michels, Sören Pirk, and Wojtek Pałubicki. 2019. Synthetic Silviculture: Multi-Scale Modeling of Plant Ecosystems. *ACM Transactions on Graphics* 38, 4 (2019), 1–14.

D. Marmanis, K. Schindler, J.D. Wegner, S. Galliani, M. Datcu, and U. Stilla. 2018. Classification with an Edge: Improving Semantic Image Segmentation with Boundary Detection. *ISPRS Journal of Photogrammetry and Remote Sensing* 135 (2018), 158–172. https://doi.org/10.1016/j.isprsjprs.2017.11.009

Andelo Martinovic and Luc Van Gool. 2013. Bayesian Grammar Learning for Inverse Procedural Modeling. In *Proc. of 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (Portland, OR, USA). IEEE, 201–208. http://ieeexplore.ieee.org/document/6618877/

Khairiya Mudrik Masoud, Claudio Persello, and Valentyn A. Tolpekin. 2019. Delineation of Agricultural Field Boundaries from Sentinel-2 Images Using a Novel Super-Resolution Contour Detector Based on Fully Convolutional Networks. *Remote Sensing* 12, 1 (2019), 59. https://doi.org/10.3390/rs12010059

Till Niese, Sören Pirk, Matthias Albrecht, Bedrich Benes, and Oliver Deussen. 2022. Procedural Urban Forestry. *ACM Transactions on Graphics* 41, 2 (2022), 1–18. https://doi.org/10.1145/3502220

Gen Nishida, Ignacio Garcia-Dorado, Daniel G. Aliaga, Bedrich Benes, and Adrien Bousseau. 2016. Interactive Sketching of Urban Procedural Models. *ACM Transactions on Graphics* 35, 4 (2016), 1–11. https://doi.org/10.1145/2897824.2925951

Yoav I. H. Parish and Pascal Müller. 2001. Procedural modeling of cities. In *Proc. of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*. Association for Computing Machinery, New York, NY, USA, 301–308. https://doi.org/10.1145/383259.383292

Wojtek Pałubicki, Miłosz Makowski, Weronika Gajda, Torsten Hädrich, Dominik L. Michels, and Sören Pirk. 2022. Ecoclimates: Climate-Response Modeling of Vegetation. *ACM Transactions on Graphics* 41, 4 (2022), 1–19. https://doi.org/10.1145/3528223.3530146

C. Persello, V.A. Tolpekin, J.R. Bergado, and R.A. De By. 2019. Delineation of Agricultural Fields in Smallholder Farms from Satellite Images Using Fully Convolutional Networks and Combinatorial Grouping. *Remote Sensing of Environment* 231 (2019), 111253. https://doi.org/10.1016/j.rse.2019.111253

Michael L. V. Pitteway and Dereck J Watkinson. 1980. Bresenham's algorithm with grey scale. *Commun. ACM* 23, 11 (1980), 625–626.

Anna Rydberg and Gunilla Borgefors. 2001. Integrated Method for Boundary Delineation of Agricultural Fields in Multispectral Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing* 39, 11 (2001), 2514–2520.

Jonathan D. Salter, Cam Campbell, Murray Journeay, and Stephen R.J. Sheppard. 2009. The Digital Workshop: Exploring the Use of Interactive and Immersive Visualisation Tools in Participatory Planning. *Journal of Environmental Management* 90, 6 (2009), 2090–2101. https://doi.org/10.1016/j.jenvman.2007.08.023

Jerry Talton, Lingfeng Yang, Ranjitha Kumar, Maxine Lim, Noah Goodman, and Radomír Měch. 2012. Learning Design Patterns with Bayesian Grammar Induction. In *UIST '12: The 25th Annual ACM Symposium on User Interface Software and Technology* (Cambridge Massachusetts USA). ACM, 63–74. https://dl.acm.org/doi/10.1145/2380116.2380127

Carlos A. Vanegas, Daniel G. Aliaga, Bedrich Benes, and Paul A. Waddell. 2009. Interactive Design of Urban Spaces Using Geometrical and Behavioral Modeling. *ACM Transactions on Graphics* 28, 5 (2009), 1–10. https://doi.org/10.1145/1618452.1618457

Carlos A. Vanegas, Ignacio Garcia-Dorado, Daniel G. Aliaga, Bedrich Benes, and Paul Waddell. 2012a. Inverse Design of Urban Procedural Models. *ACM Transactions on Graphics* 31, 6 (2012), 1–11. https://doi.org/10.1145/2366145.2366187

Carlos A. Vanegas, Tom Kelly, Basil Weber, Jan Halatsch, Daniel G. Aliaga, and Pascal Müller. 2012b. Procedural Generation of Parcels in Urban Modeling. *Computer Graphics Forum* 31 (2012), 681–690. Issue 2pt3. https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2012.03047.x

Matthias P. Wagner and Natascha Oppelt. 2020. Extracting Agricultural Fields from Remote Sensing Imagery Using Graph-Based Growing Contours. *Remote Sensing* 12, 7 (2020), 1205. https://doi.org/10.3390/rs12071205

François Waldner and Foivos I. Diakogiannis. 2020. Deep Learning on Edge: Extracting Field Boundaries from Satellite Images with a Convolutional Neural Network. *Remote Sensing of Environment* 245 (2020), 111741. https://doi.org/10.1016/j.rse.2020.111741

Li-Yi Wei, Jianwei Han, Kun Zhou, Hujun Bao, Baining Guo, and Heung-Yeung Shum. 2008. Inverse Texture Synthesis. *ACM Transactions on Graphics* 27, 3 (2008), 1–9. https://doi.org/10.1145/1360612.1360651

Liegang Xia, Jiancheng Luo, Yingwei Sun, and Haiping Yang. 2018. Deep Extraction of Cropland Parcels from Very High-Resolution Remotely Sensed Imagery. In *Proc. of 2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics)* (Hangzhou). IEEE, 1–5. https://doi.org/10.1109/Agro-Geoinformatics.2018.8476002

Leilei Xu, Peng Yang, Juanjuan Yu, Fei Peng, Jia Xu, Shiran Song, and Yongxing Wu. 2023. Extraction of Cropland Field Parcels with High Resolution Remote Sensing Using Multi-Task Learning. *European Journal of Remote Sensing* 56, 1 (2023), 2181874. https://doi.org/10.1080/22797254.2023.2181874

L. Yan and D.P. Roy. 2014. Automated Crop Field Extraction from Multi-Temporal Web Enabled Landsat Data. *Remote Sensing of Environment* 144 (2014), 42–64.

Yiwei Zhao, Han Liu, Igor Borovikov, Ahmad Beirami, Maziar Sanjabi, and Kazi Zaman. 2019. Multi-Theme Generative Adversarial Terrain Amplification. *ACM Transactions on Graphics* 38, 6 (2019), 1–14. https://doi.org/10.1145/3355089.3356553