Greenscapes: Optimized Vegetation Configurations in Cities

PATRYK BARTKOWIAK, Adam Mickiewicz University, Poznań, Poland
MARCIN MARLEWSKI, Adam Mickiewicz University, Poznań, Poland
PAWEŁ WÓJCIK, Adam Mickiewicz University, Poznań, Poland
JAE JOONG LEE, Purdue University, United States of America
DOMINIK L. MICHELS, King Abdullah University of Science and Technology (KAUST), Saudi Arabia
BEDRICH BENES, Purdue University, United States of America
SÖREN PIRK, Christian-Albrechts-Universität zu Kiel, Germany
WOJCIECH PAŁUBICKI, Adam Mickiewicz University, Poznań, Poland

Fig. 1. Three complete park designs synthesized from identical Open Street Map site constraints but different style prompts. Left: *formal* design and regular spacing. Middle: *picturesque*-asymmetric composition and clustered planting. Right: *modernist*-clear geometry. All scenes are generated 3D layouts optimized under perceptual and functional objectives using a non-convex approach.

Vegetation plays a central role in cities, serving functional roles such as balancing shade and sky exposure, shaping the microclimate, carbon sequestration, and providing acoustic screening. Furthermore, aesthetic choices in landscape styles (formal, modernist, picturesque) are important for citizens' well-being and should be considered in the design of urban landscapes. In this paper, we present a novel method for placing vegetation – ranging from shrubs to large trees – in 3D urban scenes. Our method uses individual lots as input and employs a conditioned Vision-Language Model (VLM) to generate a top-down rendering of vegetation, placing large trees, small trees, and shrubs on lawn areas. This rendering is then converted into a 3D scene by detecting individual plants and selecting appropriate 3D assets. To refine the proposed configurations, we define an optimization loop that keeps the VLM-generated positions fixed after a deterministic feasibility repair and uses CMA-ES to optimize per-instance scale transformations along with discrete vegetation asset substitutions. To define the objective function for the optimization, we evaluate the generated 3D scenes with efficient surrogate models of shade and path-based acoustic exposure. Furthermore, we render in-scene views at each optimization step to compute view-dependent perceptual descriptors, enabling the generation of both aesthetic and functional vegetation layouts for urban scenes.

CCS Concepts: • **Computing methodologies → Mesh geometry models**; **Neural networks**.

Additional Key Words and Phrases: Urban Modeling, Vision-Language-Models, Scene Generation, Modeling of Parks, Tree Modeling

ACM Reference Format:
Patryk Bartkowiak, Marcin Marlewski, Paweł Wójcik, Jae Joong Lee, Dominik L. Michels, Bedrich Benes, Sören Pirk, and Wojciech Pałubicki. 2026. Greenscapes: Optimized Vegetation Configurations in Cities. In *SIGGRAPH Asia 2026 Conference Papers (SA Conference Papers '26), December 01–04, 2026, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3829340.3842288

Authors' Contact Information: Patryk Bartkowiak, Adam Mickiewicz University, Poznań, Poznań, Poland, patryk.bartkowiak@amu.edu.pl; Marcin Marlewski, Adam Mickiewicz University, Poznań, Poznań, Poland, marmar73@st.amu.edu.pl; Paweł Wójcik, Adam Mickiewicz University, Poznań, Poznań, Poland, pawwoj10@st.amu.edu.pl; Jae Joong Lee, Purdue University, West Lafayette, United States of America, lee2161@purdue.edu; Dominik L. Michels, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, dominik.michels@kaust.edu.sa; Bedrich Benes, Purdue University, West Lafayette, United States of America, bbenes@purdue.edu; Sören Pirk, Christian-Albrechts-Universität zu Kiel, Kiel, Germany, soeren.pirk@gmail.com; Wojciech Pałubicki, Adam Mickiewicz University, Poznań, Poznań, Poland, wojciech.palubicki@amu.edu.pl.

ACM ISBN 979-8-4007-2842-6/2026/12
https://doi.org/10.1145/3829340.3842288

1 Introduction

Synthesizing urban vegetation in computer graphics poses substantial challenges due to the interplay of planning constraints, aesthetic composition across multiple scales, from individual plants to city-wide ecosystems, and vegetation diversity. Modeling urban landscapes is critical for visual storytelling in games and movies [Smelik et al. 2014], urban planning [Vanegas et al. 2012], microclimate analysis [Toparlar et al. 2017], and digital twins [Ketzler et al. 2020].

Moreover, urban afforestation plays a critical role in making cities more livable [European Commission 2021; European Parliament and the Council of the European Union 2024].

While computer graphics has made significant progress in procedural city generation and building synthesis [Müller et al. 2006; Nishida et al. 2016; Parish and Müller 2001], most existing techniques either reduce vegetation to stochastic point distributions or focus solely on trees, ignoring the understory, shrubbery, and ground cover. Ecosystem simulations provide high-fidelity growth models but are designed for wild, competitive environments [Emilien et al. 2015; Gain et al. 2017; Pałubicki et al. 2022; Peytavie et al. 2024] and not for managed urban spaces [Benes et al. 2011]. Others focus on learning distributions from satellite data [Niese et al. 2022], but often lack the precise controllability required to enforce specific setbacks, clearance zones, or functional performance goals. Generative approaches have begun to explore text-to-3D city synthesis [Feng et al. 2023; Xie et al. 2024] and reconstruction [Liu et al. 2025b; Zhang et al. 2020]. However, coupling explicit landscape design rules with quantifiable 3D performance metrics, such as acoustic damping or thermal comfort, remains largely unaddressed.

We present a novel urban landscaping framework for synthesizing heterogeneous vegetation assets. For a given urban layout, our method employs a design-centric approach to generating and optimizing vegetation layout and function. The input is a 2D map of the urban layout and a text prompt describing the layout strategy. Our framework outputs aesthetically plausible vegetation configuration that fulfills high-order functional constraints. We show that modeling urban landscapes can be accomplished by coupling Vision-Language Models (VLMs) for high-level semantic structure with evolutionary optimization (CMA-ES) [Hansen and Ostermeier 2001] for geometric refinement. This enables the generation of a broad range of design typologies, such as borders, allées, and garden rooms, while adhering to parcel geometry. Unlike previous models that focus on vegetation location, our framework explicitly evaluates functional performance by estimating shade coverage, sky-view factors (SVF), and acoustic attenuation. Fig. 1 shows three park designs generated from identical geometric site constraints. Left: a formal layout with regular spacing. Center: a picturesque, asymmetric composition with clustered planting. Right: a modernist scheme characterized by clear, rigorous geometry.

Our contributions are (1) a hybrid landscaping framework combining VLM-based structural priors with geometric evolutionary optimization, (2) phenomenological functional objectives – efficient scene-design surrogates for shade, biomass, and acoustic exposure – and (3) a coupled evaluation scheme for view-dependent aesthetics. Source code, prompt templates, and configurations are publicly available at https://github.com/greenscapes-project/greenscapes.

2 Related Work

Computer graphics research on plant synthesis spans procedural modeling, reconstruction, and interactive authoring.

Early approaches model branching using rule-based systems [Honda 1971], grammars [Aono and Kunii 1984], particle systems [Reeves and Blau 1985], and fractals [Oppenheimer 1986]. L-systems [Pałubicki et al. 2019; Prusinkiewicz 1986; Prusinkiewicz and Lindenmayer 1990] provide a principled formalism for plant generation. However, specifying rules for complex trees is labor-intensive and may require learned rule inference [Guo et al. 2020; Lee et al. 2023]. To improve usability, procedural modeling methods combine botanical priors with geometric modeling and interactive controls [Lintermann and Deussen 1996; Longay et al. 2012; Weber and Penn 1995]. Subsequent work addresses self-organizing growth [Guo et al. 2020; Palubicki et al. 2009; Runions et al. 2005], inverse procedural modeling [Stava et al. 2014], urban forests [Benes et al. 2011; Niese et al. 2022], and environmental response [Pirk et al. 2012].

Tree reconstruction has been studied from single or multi-view imagery [Neubert et al. 2007; Reche-Martinez et al. 2004; Tan et al. 2008], point clouds [Livny et al. 2011; Xu et al. 2007], and videos [Li et al. 2011]. Several pipelines leverage intermediate representations, such as segmentation masks [Argudo et al. 2016] or 3D bounding volumes [Li et al. 2021], to regularize inference of branching and foliage. Bradley et al. [2013] reconstruct dense foliage by combining stereo-derived point clouds with a statistical model.

Interactive authoring aims to make plant design practical for artists and designers. Sketch-based methods use gestures and silhouettes that guide procedural branching [Chen et al. 2008; Okabe et al. 2007; Tan et al. 2008], as well as sketch-driven modeling of plants [Anastacio et al. 2009] and flowers [Ijiri et al. 2006]. TreeSketch provides a particularly complete integration of user sketches with procedural generation [Longay et al. 2012].

Recent generative pipelines have begun to address text- and structure-conditioned synthesis of large outdoor environments by combining layout planning from language and diffusion-based generation of scene representations. On the layout side, LLM and VLM-driven planners generate structured spatial specifications (e.g., bounding boxes, scene elements, or hierarchical layouts) that can be executed by downstream generators, enabling improved controllability and spatial reasoning compared to direct text-to-image generation [Feng et al. 2023; Srivastava et al. 2025; Wang et al. 2025]. For outdoor and urban environments, diffusion models have been extended from object- and indoor-centric settings toward city-scale or real-world outdoor scene synthesis, including semantic scene generation and editing in large outdoor maps [Lee et al. 2024], text-driven multi-view consistent 3D scene generation that supports long camera trajectories and outdoor content [Zhang et al. 2024], and coarse-to-fine 3D diffusion schemes designed for detailed large-scale outdoor scenes [Liu et al. 2024a]. Structure-conditioned urban generation has explored graph-based formulations, including hierarchical layout generation at the city scale from scene-graph-conditioned diffusion for controllable 3D outdoor scenes [Liu et al. 2025a]. LLM-assisted procedural modeling has also been explored for instruction-driven 3D asset construction in 3D-GPT [Sun et al. 2025], text-controlled procedural landscape generation [Liu et al. 2024b], and large-scale procedural scene generation in SceneX [Zhou et al. 2025]. While these methods substantially improve controllability, they primarily optimize for visual or structural fidelity. In contrast, our approach explicitly couples generative proposals with functional and perceptual evaluation inside an optimization loop.

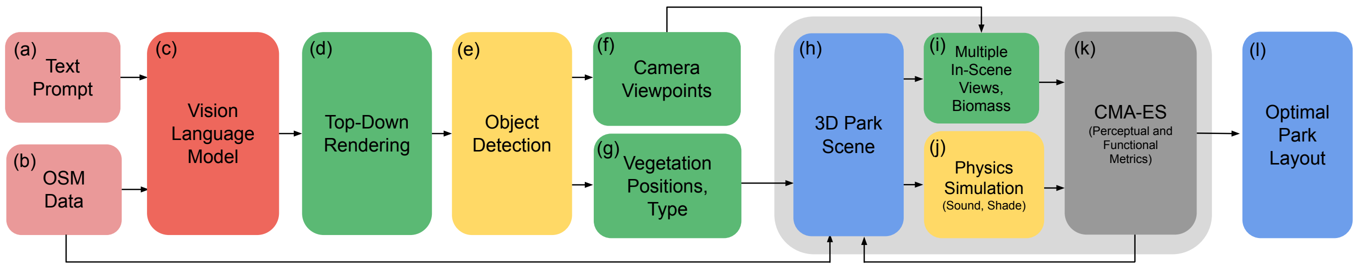


Fig. 2. Overview of our pipeline for prompt-driven park vegetation design. Inputs are a user text prompt (a) and OSM-derived site context (b). An image-to-image constrained editor generates multiple top-down layout proposals (c), which are converted to a metric top-down representation (d) and processed by instance extraction (e) to obtain candidate pedestrian viewpoints (f) and vegetation instance positions and coarse types (g). These instances are projected into world coordinates and assembled into a 3D park scene (h). We evaluate candidate scenes by rendering multiple in-scene views to compute perceptual descriptors and biomass-related quantities (i) and by running physics-inspired field models for functional measures such as shade and sound (j). CMA-ES then refines continuous instance parameters and triggers discrete species substitutions under a Gaussian-mixture asset prior using the combined perceptual and functional objectives (k), yielding the final optimized park layout (l).

3 Method

The inputs to our pipeline (Fig. 2) are a user-defined text prompt (Fig. 2a) and a lot polygon geometry from OpenStreetMap (OSM) [Haklay and Weber 2008] data expressed in a 2D coordinate frame (Fig. 2b). We rasterize the OSM polygon into a top-down conditioning image and run a *constrained image-to-image generator* (Gemini 3, [Google DeepMind 2025]) instructed to modify only a designated editable region (grass). This produces top-down vegetation-layout proposals (Fig. 2c) that are used as the sole input for class-wise instance extraction (Fig. 2d–e). All experiments use fixed prompt templates rather than interactively tuned prompts. The templates specify the editable mask, the allowed vegetation vocabulary, the required top-down orthographic style, and the requested landscape pattern or design strategy. In our experiments, we used 18 fixed templates: 17 style and layout templates for prompt-conditioned generation and one annotation template for instance extraction. In parallel, we construct a bounded set of evaluation viewpoints by prompting the VLM to annotate pedestrian standing points and view directions on the conditioning raster (Fig. 2f); the decoding procedure is detailed in the supplementary material. Detected instances are then *projected* into world coordinates, assigned assets from a library, and assembled into a 3D scene with per-instance attribute vectors and collision proxies (Fig. 2g–h). We evaluate candidate scenes by rendering in-scene buffers from the selected viewpoints and compute image- and simulation-derived metrics (Fig. 2i–j). Finally, we refine continuous per-instance parameters and discrete asset identities using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) under a low-dimensional parameterization coupled with a Gaussian Mixture Model (GMM) (Fig. 2k). We return the best feasible configuration (Fig. 2l).

3.1 OSM Preprocessing, Masks, and Assets

For the target lot and neighborhood context, we query OSM for polygonal geometry (lot boundaries, building footprints, and obstacles) and graph geometry (streets and paths).

The OSM tags are mapped into a compact set of region classes used throughout the pipeline (e.g., *plots*, *parks*, *alleys*, *urban forest*, and other vegetation-relevant categories). The layers are rasterized into aligned masks and grouped into exclusion and mandatory regions (e.g., building, road, water, pathway, open space, and grass).

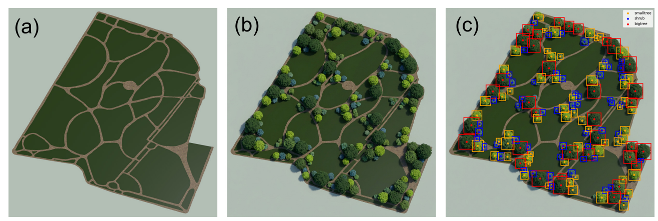


Fig. 3. Feasibility and constraint adherence of the first stages of our pipeline. In (a), we show an example conditioning raster obtained from OSM data. In (b), a constrained-edit proposal of the VLM is depicted for that input raster. In (c), we show the VLM-generated bounding box annotation overlay.

We use a library of parameterized vegetation assets categorized as large trees, small trees, and shrubs. Each vegetation asset stores an *attribute vector*

$$\mathbf{z} = [\, a,\ b,\ c,\ k_{\text{shade}},\ k_{\text{bio}},\ k_{\text{ac}} \,] \in \mathbb{R}^6, \tag{1}$$

where the ellipsoidal bounding volume($\mathcal{E} = (a, b, c)$ is defined by the two horizontal crown radii (a, b) and its height c, k_{shade} defines the shadow strength used in the shadow-propagation method [Palubicki et al. 2009], k_{bio} linearly maps canopy volume to a biomass proxy [Jenkins et al. 2003], and k_{ac} represents acoustic damping and scales a diffusion coefficient of a 3D acoustic diffusion model. The bounding box ellipsoid (a, b, c) also serves as a Gaussian prior. The values are derived from horticultural size ranges.

We rasterize the OSM geometry into a top-down orthographic image. This rasterization defines a single global axis-aligned affine transform $T : \mathbb{R}^2 \to \mathbb{R}^2$ that maps world coordinates (meters) to pixel coordinates. We use T throughout for all pixel-to-meter conversions.

3.2 Top-down Layout Generation and Instance Extraction

We generate vegetation layouts using Gemini 3 Pro [Google DeepMind 2025] on the conditioning raster, subject to *spatial constraints*: only pixels labeled as grass are editable. Obstacles (paths, buildings, water) and the outside-canvas background are regions where vegetation cannot be placed, and the pixels in these regions must remain unchanged. The prompt restricts the vocabulary of vegetation assets to (`largetree`, `smalltree`, `shrub`), each represented by a distinct canopy color; lawns form the editable ground region. Fig. 3a) illustrates the conditioning raster.

To accurately recover individual assets from the generated layout, we then perform a second pass with Gemini 3 Pro and prompt the model to return a JSON list of labeled bounding boxes (`bigtree`, `smalltree`, `shrub`) in normalized image coordinates, which we convert to pixel coordinates; for the city-scale examples, a dot-annotation variant is used instead (see supplementary material). The extraction yields an instance list $\pi_i = \{\texttt{class}_i, (u_i, v_i), s_i\}$ where (u_i, v_i) is the center of the bounding box and s_i is derived from the box dimensions.

Instance centers are reprojected to world coordinates by the inverse transform $(x_i, y_i) = T^{-1}(u_i, v_i)$. We map each detected class to a vegetation category and select a random asset within that category. Each instance stores the assigned asset identity and its attribute vector $\mathbf{z}$, as defined in Eq. (1). The size proxy s_i is converted to an initial footprint radius in meters using T and assigned a canopy ellipsoid $\mathcal{E}$ with radii (a_i, b_i, c_i). Continuous per-instance scale parameters are applied to the mesh and ellipsoid.

The VLM output is post-processed deterministically. Each detected bounding box is first assigned to one of the vegetation classes and mapped from image coordinates to lot coordinates by T^{-1}. The box center defines the initial 2D instance position, while the box extent defines an initial footprint that is converted into a 3D bounding ellipsoid. A placement is marked invalid if its center lies outside the editable scene region, if its ellipsoid extends outside the lot, intersects obstacle polygons, overlaps path geometry, or violates vegetation-to-vegetation collision constraints. The invalid placements are repaired in two steps. First, center reprojection moves the instance center to the nearest feasible location on the planting mask. Then, local obstacle correction shifts the bounding ellipsoid along the signed-distance gradient until remaining intersections are removed. The maximum number of repair iterations is 30. If the instance is still invalid, it is discarded. More technical details about the repair step of our method are given in the supplementary material of this paper.

3.3 In-scene Rendering and Metric Computation

Segmentation buffers for each scene and each selected viewpoint are rendered. We define eye-level perspective views for skyline shape assessment and a hemispherical fisheye segmentation to compute a sky view factor at all viewpoints generated in the previous stage. The buffers are inputs to the view-dependent aesthetic metrics. We also compute the scene's biomass based on the biomass of the vegetation assets. Furthermore, we simulate shading and acoustic exposure from the vegetation meshes. These functional terms are efficient scene-design surrogates rather than calibrated physical predictions.

Shade calculation. The path-level *direct-sun* shading is estimated using a deterministic *shadow propagation method* from [Palubicki et al. 2009]. Monte-Carlo or global-illumination-style estimators are substantially more expensive, while voxel shadow propagation provides a plausible approximation proxy of local light conditions [Nauber and Mäder 2024].

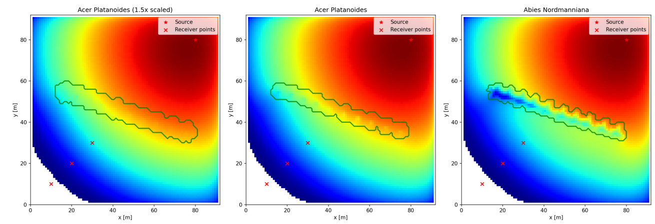
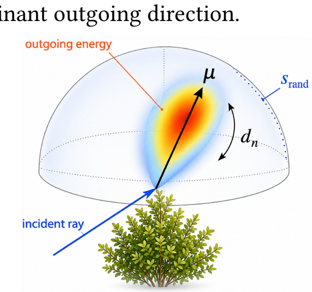


Fig. 4. Color-mapped visualization of the ear-height slice of the simulated 3D sound-energy voxel space. A sound source is placed near the upper-right corner (red marker), and a linear tree belt (green outline) separates the source from receiver samples (red markers) below. Increasing the per-vegetation attenuation parameter slows the diffusion update through the tree belt, reducing and delaying energy propagation toward the receivers; the three panels differ only in this attenuation coefficient, while keeping the geometry and discretization fixed.

Acoustic exposure is estimated with a two-stage model designed for repeated evaluation during optimization. First, we pre-compute a small set of acoustic descriptors for each vegetation asset from its 3D geometry and acoustic damping value using Monte Carlo ray tracing. Second, we use the descriptors to parameterize a diffusion-with-decay simulation of sound energy on a 3D voxel grid. The resulting exposure is measured at camera view points.

We trace 20,000 rays for each of 16 incident directions sampled over a hemisphere per vegetation asset in a pre-processing stage. Each ray performs a single interaction with the geometry. At the intersection point, the outgoing direction is sampled either as a specular or a diffuse reflection, where diffuse directions are sampled from a cosine-weighted lobe. We compute three parameters per asset from the final distribution of outgoing ray directions and energies: (1) s_{rand} is the fraction of reflected energy describing random scattering (using ISO-17497-1), (2) d_n denotes a normalized measure of how spread-out the outgoing energy is over the azimuth, where higher values imply stronger spreading (using ISO-17497-2), and (3) $\boldsymbol{\mu}$ is the dominant outgoing direction.



For reflected rays with outgoing unit directions $\boldsymbol{\omega}_i$ and remaining energies w_i, we define $\boldsymbol{\mu} = \sum_i w_i \boldsymbol{\omega}_i / \| \sum_i w_i \boldsymbol{\omega}_i \|$, which gives the normalized energy-weighted resultant direction.

Scene-level evaluations. Given a candidate scene, we discretize the park volume into a 3D voxel grid and label voxels by material class (air, vegetation, buildings, ground). We then simulate the propagation of sound-energy density $E(\mathbf{x}, t)$ with an anisotropic diffusion model with decay and a point-source term. The update is computed with an explicit finite-difference scheme (forward Euler) and a timestep chosen for stability (CFL). The top and side domain boundaries are treated as open space by enforcing E=0.

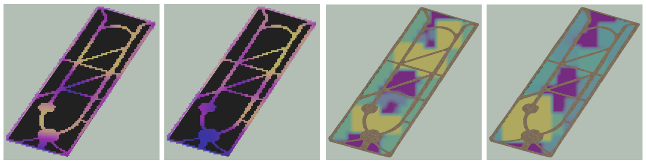


Fig. 5. Example park scenes before and after CMA-ES refinement of the VLM prior. Path shading is visualized as a scalar field over the walkable network (plasma colormap), and vegetation biomass is visualized per instance (viridis colormap). Optimization increases functional shading where needed while reallocating plant mass to meet biomass/coverage targets under the same geometric constraints; the images show the resulting redistribution of canopy placement along paths.

We propagate energy density $E(\mathbf{x}, t)$ using an anisotropic diffusion equation with decay:

$$\frac{\partial E}{\partial t} \;=\; \nabla\cdot\left(D(\mathbf{x})\,\nabla E\right) \;-\; \lambda(\mathbf{x})\,E \;+\; S(\mathbf{x}, t), \tag{2}$$

where S is a source term. Each voxel stores a diffusion tensor D and a decay rate λ determined by the class of the occupying object (air, vegetation species, building, ground), with vegetation parameters derived from the precomputed asset descriptors.

Vegetation voxels receive anisotropic diffusion and decay parameters from the precomputed descriptors using a fixed global heuristic shared across all experiments. Specifically, higher d_n increases the base diffusion magnitude as the energy spreads more in space, $\boldsymbol{\mu}$ biases diffusion along a dominant direction (anisotropy), and higher s_{rand} increases decay to represent stronger attenuation. Buildings are treated as strong decay regions, while air uses fixed background values. All numerical mapping constants are listed in the supplementary material.

We place eight receiver points at 2 m ear height, evenly spaced along the path network and record the simulated energy E over time. To isolate the effect of vegetation, three simulations are run with identical source, discretization, and boundary handling: one *free-field* reference, one *with* vegetation, and another with higher attenuating vegetation. We report excess attenuation as an energy ratio, expressed in decibels

$$\Delta L_{\text{dB}} = -10\log_{10}\left(\frac{E_{\text{with}}}{E_{\text{free}}}\right), \tag{3}$$

visualized as a heat map in Fig. 4. This provides a deterministic acoustic objective term comparable across optimization iterations.

Skyline flatness and periodicity. From the eye-level segmentation view, we compute a binary sky mask. We then extract a skyline profile $y(x)$ column-wise as the last sky pixel before the first non-sky pixel (with interpolation over missing columns). Flatness is computed from the detrended skyline (linear trend removed) using three terms: the standard deviation of $y(x)$, mean absolute first difference, and RMS second difference; these are normalized by a fixed pixel scale and combined into a score $\exp(-(\cdot))$. Periodicity is computed on the same detrended skyline as the average of the dominant-frequency power ratio in the FFT and the strongest autocorrelation peak in a window around the FFT-implied wavelength. For target-matching experiments, the skyline term measures the deviation of the extracted profile from a target profile, e.g., a sinusoid.

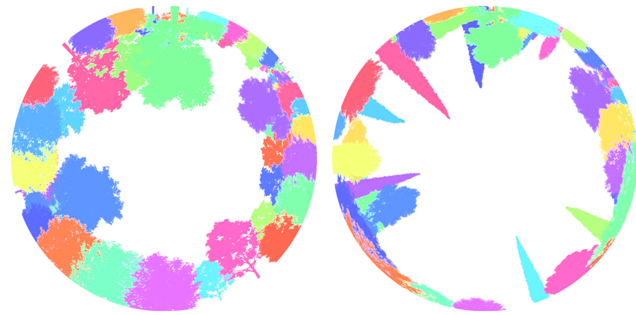


Fig. 6. Sky-view exposure before and after CMA-ES refinement from a fixed in-park viewpoint. The fisheye sky segmentation (white = sky) shows increased open-sky visibility after optimization, raising the sky view factor (SVF) to 0.70 while preserving the surrounding layout constraints.

3.4 CMA-ES Refinement

Our main goal is to generate optimal green layouts through non-convex optimization. Therefore, we process the 3D representation of our generated park output through a shade and acoustic exposure simulation pipeline, and compute all relevant skyline and biomass metrics. We then start the final optimization stage of our framework, leveraging CMA-ES. For a selected set of M CMA-ES scene operators,

$$\theta = \{(s_i^{xy}, s_i^{z}, \mathbf{z}_i')\}_{i=1}^{M}, \tag{4}$$

where (s_i^{xy}, s_i^{z}) are horizontal and vertical scale multipliers and $\mathbf{z}_i' \in \mathbb{R}^6$ is an attribute proposal in the asset space of Eq. (1). All feasible instances of the repaired configuration are optimized: the VLM proposal fixes the number, class, and 2D position of the instances; the repair operator (Sec. 3.2) runs once before optimization to establish feasibility; and CMA-ES subsequently adjusts only the per-instance scales and attribute proposals, with the GMM assignment triggering discrete species substitutions. Positions are not optimization variables; feasibility during optimization is maintained by clamping the scale parameters to their bounds.

Each species j defines one Gaussian kernel $\mathcal{N}(\boldsymbol{\mu}_j, \Sigma_j)$ in attribute space (diagonal Σ_j). Given a proposal $\mathbf{z}_i'$, we assign

$$j^* = \arg\max_j \; \pi_j\,\mathcal{N}(\mathbf{z}_i' \mid \boldsymbol{\mu}_j, \Sigma_j), \tag{5}$$

and set the discrete asset identity of instance i to j^*. We reject proposals outside an admissible Mahalanobis distance for all assets. In case no valid species is selected, we project $\mathbf{z}_i'$ back onto the closest ellipsoid defined by the Mahalanobis threshold and use it to set (a, b) and coefficients $(k_{\text{shade}}, k_{\text{bio}}, k_{\text{ac}})$. Fig. 8 illustrates the resulting *walk-and-jump* behavior. Initially, an asset starts within the Gaussian kernel that defines the species Acer campestre. As a result of multiple optimization steps, the asset's attribute proposal falls within the Mahalanobis threshold for another species (Acer platanoides) and switches its species tag. As a consequence, we replace the asset with another (belonging to Acer platanoides) and continue the CMA-ES optimization loop.

Each candidate θ is evaluated as $E(\theta)$, where E is the weighted sum of five normalized terms – SVF, a skyline-fit term (flatness, periodicity, or deviation from a target profile, depending on the experiment), path shading, acoustic exposure, and biomass – plus a small regularizer penalizing deviations from the initial scales.

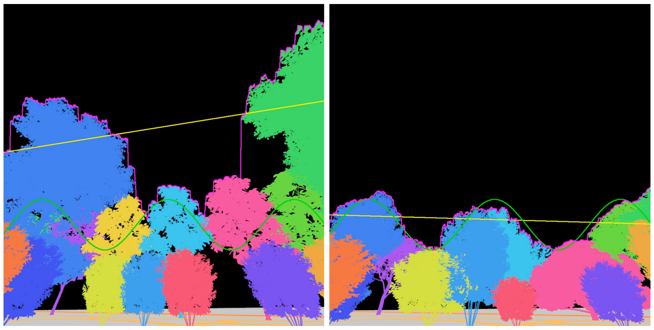


Fig. 7. Skyline periodicity matching before (left) and after (right) CMA-ES refinement from a fixed viewpoint. The green curve is the target sinusoidal skyline profile; CMA-ES improves agreement by jointly adjusting per-instance scales and performing discrete species substitutions to introduce canopies with more suitable sizes and crown shapes, yielding a more periodic silhouette while maintaining feasibility constraints.

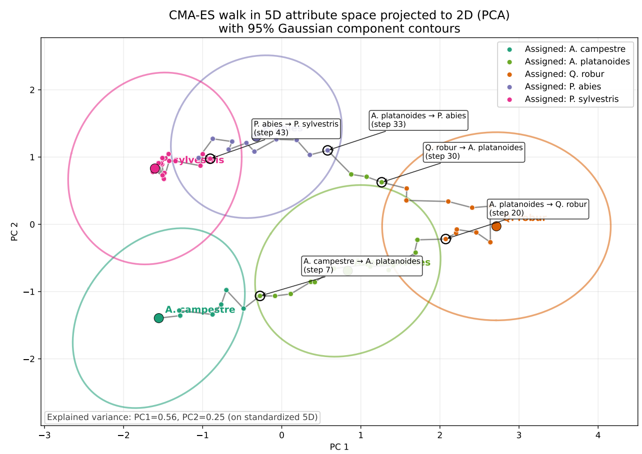


Fig. 8. Visualization of the CMA-ES trajectory in the 6D attribute space after PCA projection to 2D. Elliptical contours indicate projected Gaussian components (species priors) used by the GMM-based substitution rule; the polyline shows successive CMA-ES proposals and the induced discrete component assignments.

CMA-ES is run with a fixed seed, iteration budget G, initial step size σ_0, and population size $\lambda_{\text{pop}} = \max(4, \lfloor 4 + 3 \ln n \rfloor)$ for search dimensionality n; the bounds on $\left(s_i^{xy}, s_i^z\right)$ are enforced by clamping. The best feasible configuration over all generations is returned.

4 Results and Validation

In this section, we present the results of our modeling framework.

Park-Design Strategies. Figure 14 shows three design concepts: formal (geometric or axial), picturesque (English landscape), and modernist. They are commonly treated as distinct and iconic historical styles in landscape design and landscape architecture [Rogers 2001]. To quantify the design principles, we propose orientation entropy and linearity: formal layouts tend to have lower entropy and higher linearity, while picturesque layouts are more irregular.

We embed each scene from its `GeoJSON` vegetation layout (tree and shrub point locations) by computing two deterministic layout descriptors in a common metric frame (EPSG:3857): *orientation entropy* is the Shannon entropy of a 36-bin histogram of nearest-neighbor bearing angles modulo 180° (low for axis-aligned grids; high for irregular and circulinear layouts), while *linearity* is the fraction of points whose two nearest neighbors are approximately collinear (within 5°), reported as a percentage. We plot 33 reconstructed reference parks with known design strategies and 30 generated parks from three OSM sites and style-specific prompts (three templates per style, with LLM-generated textual variants); the generated parks are rendered as circles. The clusterings show that the two descriptors separate the three strategies and that prompt-conditioned generations align with the corresponding reference distributions, supporting both metric validity and reproducible style control from the `GeoJSON` layouts.

Prompt-conditioned Controllability. Figure 10 shows results with our 3D generative pipeline from a single OSM input, but different text prompts. We use the parcel boundary and OSM-derived path network of St. Mary's Park and reconstruct a reference vegetation layout from the NYC Tree Map inventory (geolocated trees with species and size attributes) [NYC Parks 2024]. Starting from the single OSM masks, we vary only the text prompt (e.g., *agroecological*, *East Asian*, *infrastructural*, *post-modernist*, *formal*, *picturesque*, *modernist*, *memorial*), yielding diverse 3D layouts that appear as plausible design alternatives to the reconstructed St. Mary's Park baseline (top left panel in Figure 10). The results show generated complex vegetation layouts; e.g., the *agroecological* layout features haphazardly distributed trees and bushes, whereas the *modernist* and *memorial* layouts are sparser and more organized.

CMA-ES Refinement. Figure 12 shows an objective-ablation study on Joyce Kilmer Park. Panel (a) shows the reconstructed reference layout from the NYC tree inventory (3D trees placed at their real locations), used as the input condition. Panels (b–f) show five optimization configurations in which one objective term is assigned a dominant weight while the remaining terms are assigned low or zero weights. This protocol isolates the qualitative effect of each criterion while keeping the site boundary, path network, feasibility constraints, asset library, and optimizer fixed. A key observation is that CMA-ES can substantially change the perceived park design without changing the underlying planting positions. Most of the visible differences arise from continuous rescaling and GMM-driven discrete substitutions (crown width, height propensity, and species-level density proxies). The skyline-targeted run (b) exhibits a distinct, non-local signature: alternating canopy heights emerge along the view direction to match a periodic skyline profile. *Increasing SVF* (c) results in species with tall or wide crowns being replaced by slimmer forms so that the same spatial arrangement gives the impression of a more open, plaza-like room from interior viewpoints. *Minimizing shade* (d) selects narrower, more open crowns and lower effective canopy density. *The noise-reduction target* in (e) yields a belt-like planting appearance by selecting species with higher attenuation for regions near the viewpoints. Finally, *maximizing biomass* (f) increases visual mass by selecting broader-canopy assets and higher scale factors, yielding a denser canopy and reduced long-range sightlines. The exact weights and targets for each experiment are listed in the supplementary material and released with the code.

Planting Layouts of Lots. Figure 11 provides a qualitative SOTA comparison to Figure 6 in the procedural urban-forestry method of Niese et al. [2022]. Their method defines representative procedural planting strategies, including semi-random, boundary-biased, and clustered placement. Our prompt-conditioned pipeline qualitatively reproduces these baseline planting patterns.

Figure 13 demonstrates this capability on multiple reconstructed site lots derived from OSM. For each example, we use the OSM parcel boundary and building footprint polygons, extrude the footprints into 3D building masses (blue), and treat these as exclusion constraints for vegetation placement. Despite substantial variation in parcel shape, footprint size, and building placement, the generated layouts remain constraint-adherent and adapt the planting structure to the remaining ground area: vegetation concentrates along parcel edges and setbacks, fills residual wedges and corners, and forms strip-like bands where the geometry induces narrow corridors.

Runtime Analysis. The system was implemented in Rust using Bevy for rendering and CUDA for acoustic pre-processing. All results were generated on a laptop running Windows 11 and equipped with an Intel Core i9-14900HX CPU (24 cores), 32 GB of RAM, and an NVIDIA GeForce RTX 4080 mobile GPU. The average wall-clock time for one optimization generation (processing the full population of candidate scenes) is approximately 23 seconds. Consequently, a typical full experimental run of 100 generations requires approximately 38 minutes to converge. The initialization phase (VLM generation and detection) incurs a fixed one-time latency of $\sim$ 15–20 seconds depending on API response times.

Quantitative Validation and Robustness. Table 1 summarizes $N = 30$ independent runs with independently sampled VLM layouts (the VLM API is unseeded; downstream stages use fixed seeds) and reports absolute deviations from the optimization targets. The reconstructed park layout yields a high total cost and fails to meet several prompt-specific objectives, such as SVF and Acoustics. The VLM prior exhibits substantial sample sensitivity, with large run-to-run variance in several terms (e.g. skyline-related error), whereas the CMA-ES refinement consistently reduces both the mean target error and the variance across runs, indicating that the optimization stage significantly improves objective alignment over the VLM prior.

To isolate the role of the VLM prior, we additionally perform a component ablation on Joyce Kilmer Park (Table 2). We compare random-position initialization, random initialization followed by CMA-ES, the VLM-derived prior, and the final VLM-initialized CMA-ES result. The VLM prior already improves shade scores over random placement while keeping biomass in a comparable range, indicating that the prompt-conditioned layout provides a useful semantic initialization rather than only random feasible samples. Starting from random placement, CMA-ES improves the functional scores, but remains less effective than VLM-initialized CMA-ES because random layouts often place vegetation away from paths. The final VLM+CMA-ES result achieves the highest shade and acoustic scores while minimizing biomass.

Table 1. Quantitative error analysis over $N = 30$ experimental runs reports the *mean absolute deviation from design targets* ($\mu \pm \sigma$) to demonstrate robustness. While the *VLM Prior* exhibits high variance, the *Optimized* result shows significantly lower standard deviation, indicating that our method reliably converges to the design goals regardless of initialization. Real Park is a fixed reference geometry.

Metric (Target)	Real Park (Ref) $\lvert\Delta\rvert$	VLM Prior $\lvert\Delta\rvert$	Optimized $\lvert\Delta\rvert$
SVF (0.75)	0.12	0.06 ± 0.04	**0.03 ± 0.01**
Shade (2.5)	0.32	0.80 ± 0.20	**0.28 ± 0.07**
Acoustics (3.5)	5.76	0.50 ± 0.14	**0.21 ± 0.06**
Biomass (75k)	9,736	18,683 ± 6,174	**896 ± 190**
Skyline Fit (10k)	10,213	9,433 ± 1,815	**2,986 ± 456**
Total Cost	1.30	0.40 ± 0.10	**0.05 ± 0.01**

Table 2. Ablation of the two components of our method: initialization (random vs. VLM-derived prior) and CMA-ES refinement (on vs. off)-reported on Shade, Biomass, and the optimization *Cost*, the weighted squared distance to the Shade and Biomass targets (lower is better). CMA-ES on the VLM prior achieves the lowest cost (0.006), reaching the shade target (2.04 vs. 2.00) while keeping biomass within ~2.3% of baseline; CMA-ES on random initialization falls short of the shade target (1.91) and overshoots biomass by $\sim$ 5.5%, ending $\sim$ 20× further from the joint target.

Run	Shade	Biomass	Cost
Target	2.00	56,317	0.0000
Random Init.	1.65	54,473	0.2012
Random+CMA-ES	1.91 ± 0.04	59, 447 ± 1, 176	0.1205 ± 0.0075
VLM Prior	1.70	56,317	0.0227
VLM+CMA-ES	2.04 ± 0.06	57, 581 ± 780	0.0059 ± 0.0030

5 Discussion and Limitations

Generating plausible vegetation placements in urban scenes is challenging because solutions must simultaneously satisfy geometric constraints and align with human expectations for planting structure and landscape style. Prior work tends to restrict the design space; for example, Niese et al. [2022] focuses primarily on trees and does not model understory vegetation, nor does it explicitly couple functional performance with perceptual architectural strategies; in contrast, our method configures heterogeneous multi-class plantings under such coupled objectives.

A core bottleneck is that many landscape-architectural principles are qualitative and therefore difficult to formalize in a procedural manner. Our pipeline separates responsibilities accordingly: the constrained image-edit prior provides stylistically coherent, human-plausible structure, while CMA-ES operates on a low-dimensional 3D parameterization to optimize scene-derived objectives via statistical evaluation of in-scene renderings and physics-inspired surrogate metrics. However, this gain in expressive power of scene modeling comes at the cost of algorithm runtime. Another limitation is that our functional metrics are designed as optimization surrogates, approximative in nature, rather than first-principles-based physical simulations. Finally, our method has a slower runtime than procedural urban vegetation configuration approaches, which report placing 50K trees per minute ([Niese et al. 2022]) but do not optimize for functional or aesthetic objectives.

6 Conclusion

We introduced a pipeline for urban vegetation design that combines a constrained image-edit generative prior with a low-dimensional, simulation-driven CMA-ES refinement loop. The generative stage produces stylistically coherent top-down layouts that capture perceptual landscape strategies, while the optimization stage improves functional and view-dependent objectives by performing continuous scale adjustments and discrete asset substitutions within an attribute-space GMM. Our results show prompt-conditioned controllability, metric-consistent clustering with reconstructed exemplars of known styles, and substantial refinement gains beyond the generative prior, suggesting a practical route to integrate high-level aesthetic intent with surrogate-based performance criteria in automated park design.

Acknowledgments

We thank the reviewers for their valuable comments and suggestions. This work is supported by ERC grant #101170158 - WildfireTwins.

References

F. Anastacio, P. Prusinkiewicz, and M. C. Sousa. 2009. Sketch-based parameterization of L-systems using illustration-inspired construction lines and depth modulation. *Computers & Graphics* 33, 4, 440–451.

M. Aono and T.L. Kunii. 1984. Botanical Tree Image Generation. *IEEE Comput. Graph. Appl.* 4(5) (1984), 10–34.

O. Argudo, A. Chica, and C. Andujar. 2016. Single-picture reconstruction and rendering of trees for plausible vegetation synthesis. *Computers & Graphics* 57 (2016), 55–67.

B. Benes, M. A. Massih, P. Jarvis, D. G. Aliaga, and C. A. Vanegas. 2011. Urban Ecosystem Design. In *I3D*. 167–174.

D. Bradley, T. Nowrouzezahrai, and P. Beardsley. 2013. Image-based reconstruction and synthesis of dense foliage. *ACM Trans. Graph.* 32, 4 (2013), 74:1–74:10.

X. Chen, B. Neubert, Y.-Q. Xu, O. Deussen, and S. B. Kang. 2008. Sketch-based tree modeling using Markov random field. *ACM Trans. Graph.* 27, 5 (2008), 109:1–109:9.

A. Emilien, U. Vimont, M.-P. Cani, P. Poulin, and B. Benes. 2015. WorldBrush: Interactive Example-Based Synthesis of Procedural Virtual Worlds. *ACM Trans. on Graphics* 34, 4, Article 106 (2015), 11 pages.

European Commission. 2021. The 3 Billion Tree Planting Pledge for 2030. Commission Staff Working Document SWD(2021) 651; factsheet accompanying the EU Forest Strategy for 2030. https://environment.ec.europa.eu/topics/forests/3-billion-trees_en Accessed 2026-08-13.

European Parliament and the Council of the European Union. 2024. Regulation (EU) 2024/1991 of the European Parliament and of the Council of 24 June 2024 on nature restoration and amending Regulation (EU) 2022/869. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1991/oj/eng See Article 8 regarding restoration of urban ecosystems. Accessed: 20 Dec 2025.

W. Feng, W. Zhu, T.-J. Fu, V. Jampani, A. Akula, X. He, S. Basu, X. E. Wang, and W. Y. Wang. 2023. LayoutGPT: Compositional Visual Planning and Generation with Large Language Models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

J. Gain, H. Long, G. Cordonnier, and M.-P. Cani. 2017. EcoBrush: Interactive Control of Visually Consistent Large-Scale Ecosystems. *Comp. Graph. Forum* 36, 2 (2017), 63–73. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13107 doi:10.1111/cgf.13107

Google DeepMind. 2025. *Gemini 3*. https://deepmind.google/technologies/gemini/ Large multimodal language model.

J. Guo, H. Jiang, B. Benes, O. Deussen, D. Lischinski, and H. Huang. 2020. Inverse Procedural Modeling of Branching Structures by Inferring L-Systems. *ACM Trans. Graph.* 39, 5 (2020).

M. Haklay and P. Weber. 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing* 7, 4 (2008), 12–18. doi:10.1109/MPRV.2008.80

N. Hansen and A. Ostermeier. 2001. Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation* 9, 2 (2001), 159–195. doi:10.1162/106365601750190398

H. Honda. 1971. Description of the form of trees by the parameters of the tree-like body: Effects of branching angle and the branch length on the sample of the tree-like body. *Journal of Theoretical Biology* 31, 2 (1971), 331–338.

T. Ijiri, S. Owada, and T. Igarashi. 2006. Seamless Integration of Initial Sketching and Subsequent Detail Editing in Flower Modeling. *CGF* 25, 3 (2006), 617–624.

Jennifer C. Jenkins, David C. Chojnacky, Linda S. Heath, and Richard A. Birdsey. 2003. National-Scale Biomass Estimators for United States Tree Species. *Forest Science* 49, 1 (2003), 12–35. doi:10.1093/forestscience/49.1.12

B. Ketzler, V. Naserentin, F. Latino, C. Zangelidis, L. Thuvander, and A. Logg. 2020. Digital twins for cities: A state of the art review. *Built environment* 46, 4 (2020), 547–573.

J. Lee, S. Lee, C. Jo, W. Im, J. Seon, and S.-E. Yoon. 2024. SemCity: Semantic Scene Generation with Triplane Diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 28337–28347.

J.J. Lee, B. Li, and B. Benes. 2023. Latent l-systems: Transformer-based tree generator. *ACM Transactions on Graphics* 43, 1 (2023), 1–16.

B. Li, J. Kałużny, J. Klein, D. L. Michels, W. Pałubicki, B. Benes, and S. Pirk. 2021. Learning to Reconstruct Botanical Trees from Single Images. *ACM Trans. Graph.* 40, 6, Article 231 (2021).

C. Li, O. Deussen, Y.-Z. Song, P. Willis, and P. Hall. 2011. Modeling and Generating Moving Trees from Video. *ACM Trans. on Graphics* 30, 6, Article 127 (2011), 127:1–127:12 pages.

B. Lintermann and O. Deussen. 1996. Interactive modelling of branching structures. In *ACM SIGGRAPH 96 Visual Proceedings: The Art and Interdisciplinary Programs of SIGGRAPH '96 (SIGGRAPH '96)*. Association for Computing Machinery, New York, NY, USA, 148. doi:10.1145/253607.253880

J.-H. Liu, S.-K. Zhang, C. Zhang, and S.-H. Zhang. 2024b. Controllable Procedural Generation of Landscapes. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*. Association for Computing Machinery, 6394–6403. doi:10.1145/3664647.3681129

W. Liu, Y. Zhong, Y. Li, X. Chen, J. Cui, H. Zhang, L. Xu, X. Lou, Y. Shi, J. Yu, et al. 2025b. CityGo: Lightweight Urban Modeling and Rendering with Proxy Buildings and Residual Gaussians. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*. 1–10.

Y. Liu, X. Li, X. Li, L. Qi, C. Li, and M.-H. Yang. 2024a. Pyramid diffusion for fine 3d large scene generation. In *European Conference on Computer Vision*. Springer, 71–87.

Y. Liu, X. Li, Y. Zhang, L. Qi, X. Li, W. Wang, C. Li, X. Li, and M.-H. Yang. 2025a. Controllable 3D Outdoor Scene Generation via Scene Graphs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 28052–28062. https://openaccess.thecvf.com/content/ICCV2025/html/Liu_Controllable_3D_Outdoor_Scene_Generation_via_Scene_Graphs_ICCV_2025_paper.html

Y. Livny, S. Pirk, Z. Cheng, F. Yan, O. Deussen, D. Cohen-Or, and B. Chen. 2011. Texture-lobes for Tree Modelling. *ACM Trans. on Graphics* 30, 4, Article 53 (2011), 10 pages.

S. Longay, A. Runions, F. Boudon, and P. Prusinkiewicz. 2012. TreeSketch: Interactive Procedural Modeling of Trees on a Tablet. In *Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling (SBIM)*. 107–120.

P. Müller, P. Wonka, S. Haegler, A. Ulmer, and L. Van Gool. 2006. Procedural modeling of buildings. *ACM Transactions on Graphics (TOG)* 25, 3 (2006), 614–623.

T. Nauber and P. Mäder. 2024. Light Distribution Models for Tree Growth Simulation. *Comp. Graph. Forum* (2024). doi:10.1111/cgf.15268

B. Neubert, T. Franken, and O. Deussen. 2007. Approximate image-based tree-modeling using particle flows. *ACM Trans. Graph.* 26, 3 (2007), 88–es.

T. Niese, S. Pirk, M. Albrecht, B. Benes, and O. Deussen. 2022. Procedural Urban Forestry. *ACM Trans. Graph.* 41, 2, Article 20 (March 2022), 18 pages. doi: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, Article 130 (2016), 130:1–130:11 pages. doi:10.1145/2897824.2925951

NYC Parks. 2024. *New York City Tree Map*. https://tree-map.nycgovparks.org/tree-map Accessed: 2026-01-22. The Tree Map states it includes both street trees and park trees in landscaped areas..

M. Okabe, S. Owada, and T. Igarashi. 2007. Interactive Design of Botanical Trees Using Freehand Sketches and Example-based Editing. In *ACM SIGGRAPH Courses* (San Diego, California). ACM, Article 26.

P. E. Oppenheimer. 1986. Real time design and animation of fractal plants and trees. *Proc. of SIGGRAPH* 20, 4 (1986), 55–64.

W. Palubicki, K. Horel, S. Longay, A. Runions, B. Lane, R. Měch, and P. Prusinkiewicz. 2009. Self-organizing tree models for image synthesis. *ACM Trans. Graph.* 28, 3 (2009), 58:1–58:10.

Wojtek Pałubicki, Andrzej Kokosza, and Agata Burian. 2019. Formal Description of Plant Morphogenesis. *Journal of Experimental Botany* 70, 14 (2019), 3601–3613. doi:10.1093/jxb/erz210

W. Pałubicki, M. Makowski, W. Gajda, T. Hädrich, D. L. Michels, and S. Pirk. 2022. Ecoclimates: Climate-response Modeling of Vegetation. *ACM Trans. Graph.* 41, 4, Article 155 (July 2022), 19 pages. doi:10.1145/3528223.3530146

Y. I. H. Parish and P. Müller. 2001. Procedural modeling of cities. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 301–308.

A. Peytavie, J. Gain, E. Guérin, O. Argudo, and E. Galin. 2024. DeadWood: Including Disturbance and Decay in the Depiction of Digital Nature. *ACM Trans. Graph.* 43, 2, Article 21 (Feb. 2024), 19 pages. doi:10.1145/3641816

Fig. 9. City-scale vegetation placement examples generated from OSM-derived urban context. Panels show the area around (a) Sydney Park, Australia, (b) the Mausoleum Charlottenburg area in Berlin, Germany, and (c) the Sołacz district in Poznań, Poland, with over 4000 placed tree instances per scene. Panel (d) shows an in-scene rendering of a generated layout.

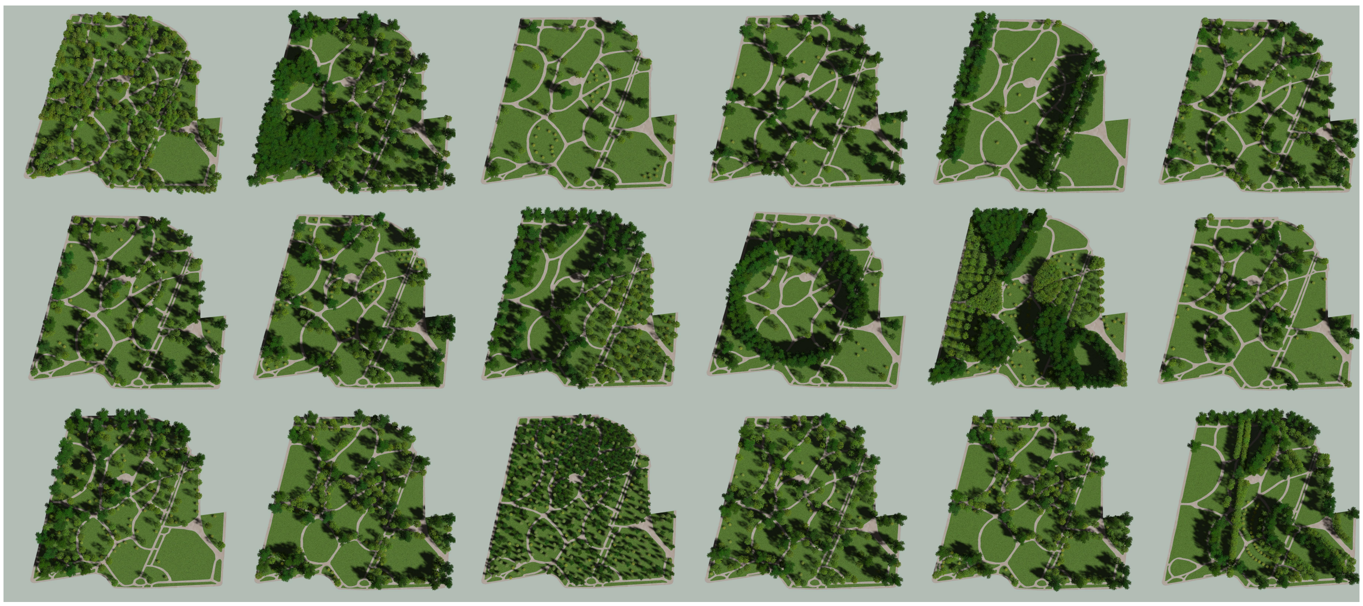


Fig. 10. The top left panel is a reconstruction of vegetation positions of St. Mary's park in New York. All other panels show the final 3D layout produced by the reconstructed 3D scene from VLM prior. Varying only the text prompt yields diverse, style-consistent compositions including *agroecological*, *East Asian*, *infrastructural*, *post-modernist*, *formal*, *picturesque*, *modernist*, and *memorial*.

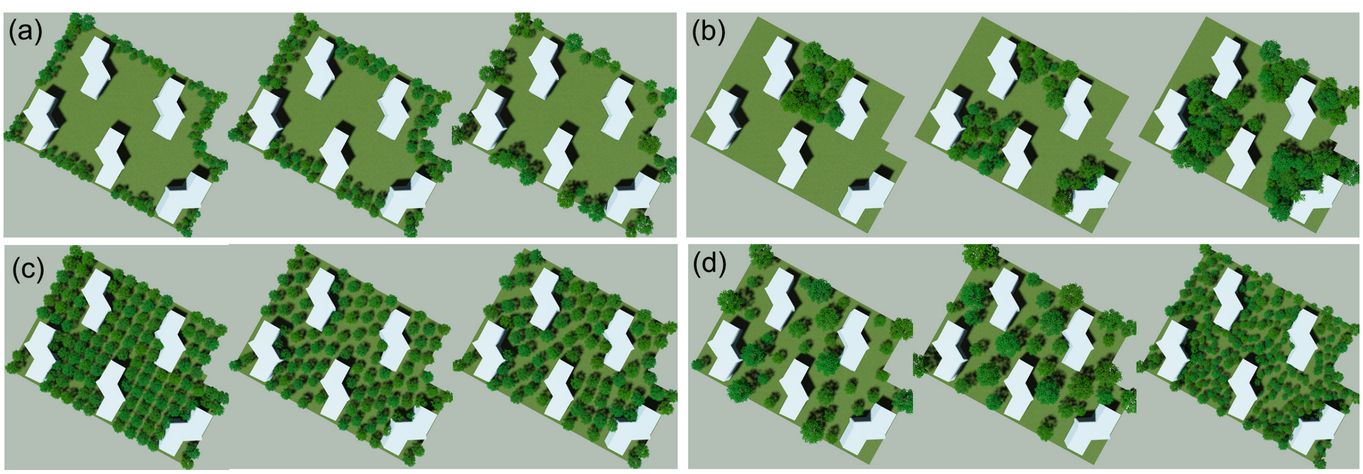


Fig. 11. All panels show outputs generated by our method using text prompts that encode the vegetation-placement patterns illustrated in Fig. 6 [Niese et al. 2022] showing: semi-random placement, where trees are sampled throughout the active lot region; boundary placement, where samples are concentrated in a band near the lot boundary; and cluster placement, where samples are concentrated around local cluster regions. In the reference method, these strategies are implemented with Variable Radii Poisson-Disk Sampling.

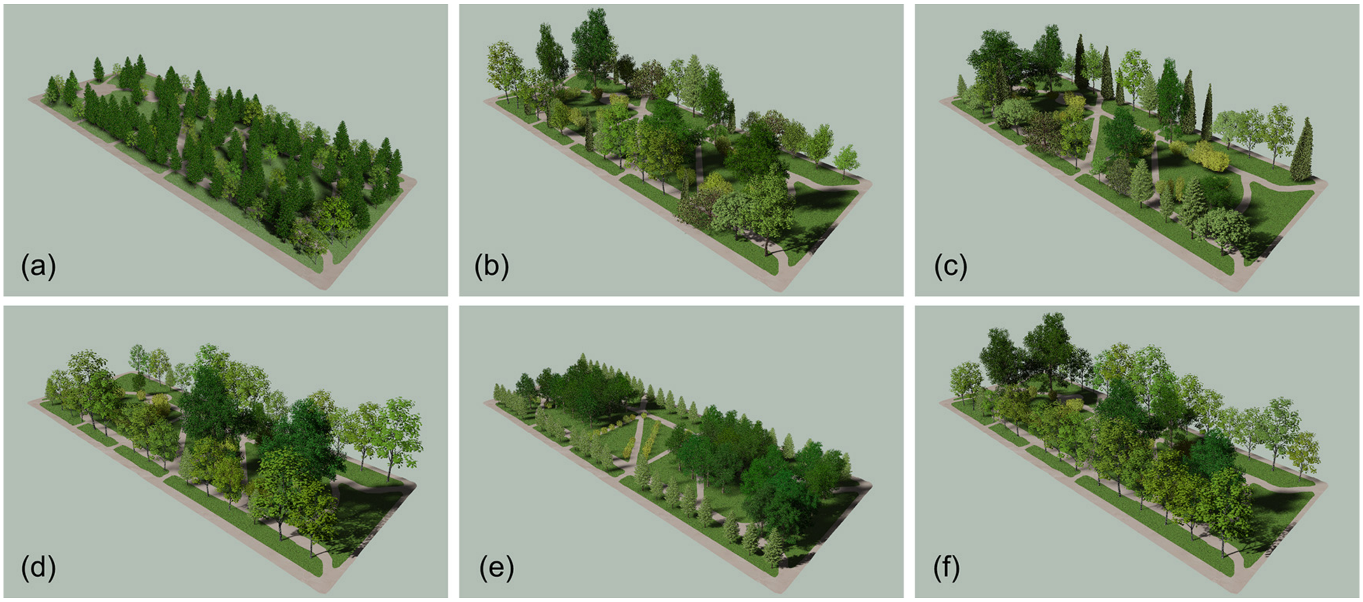


Fig. 12. Objective-ablation results for Joyce Kilmer Park. Panel (a) shows the reconstructed reference layout used as the input condition. Panels (b–f) show CMA-ES refinements obtained by assigning one objective term a dominant weight while keeping the remaining terms at low or zero weights: (b) skyline sine-function fit, (c) SVF ratio, (d) minimal path shading, (e) acoustic/noise-reduction objective, and (f) maximum biomass. All runs use the same site boundary, path network, feasibility constraints, asset library, and optimizer settings; the visual differences therefore illustrate the qualitative effect of each dominant objective.

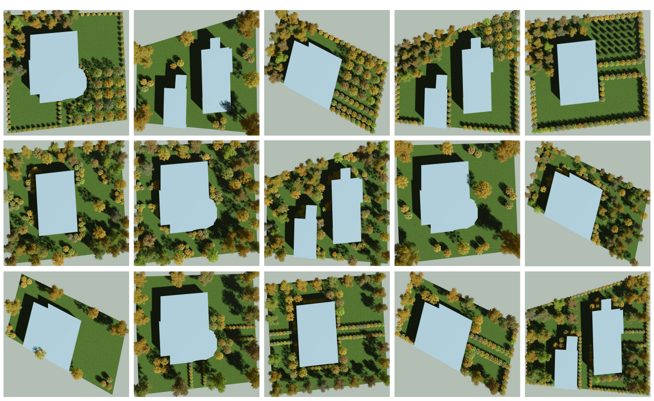


Fig. 13. Examples of generated planting layouts for different site lots. For each lot geometry, our method produces distinct plant configurations, demonstrating its ability to adapt planting structure and spatial organization to varying parcel shapes and constraints.

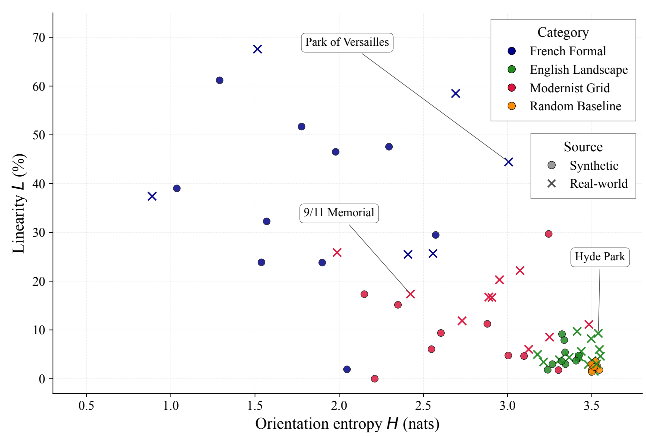


Fig. 14. We plot 33 reconstructed reference parks with known architectural styles and 30 generated parks produced from OSM sites and style prompts. Colors indicate the intended style (*formal* blue, *picturesque* green, *modernist* red). Disks versus crosses distinguish generated from reconstructed references. The resulting clustering shows that the proposed entropy and linearity metrics separate the three strategies and that the generated parks align with the corresponding reference distributions.

S. Pirk, O. Stava, J. Kratt, M. A. M. Said, B. Neubert, R. Měch, B. Benes, and O. Deussen. 2012. Plastic trees: interactive self-adapting botanical tree models. *ACM Trans. on Graphics* 31, 4, Article 50 (2012), 10 pages.

P. Prusinkiewicz. 1986. Graphical Applications of L-systems. In *Proceedings on Graphics Interface '86/Vision Interface '86*. Canadian Information Processing Society, 247–253.

P. Prusinkiewicz and A. Lindenmayer. 1990. *The Algorithmic Beauty of Plants*. Springer-Verlag New York, Inc., New York, USA.

A. Reche-Martinez, I. Martin, and G. Drettakis. 2004. Volumetric reconstruction and interactive rendering of trees from photographs. *ACM Trans. Graph.* 23, 3 (2004), 720–727.

W. T. Reeves and R. Blau. 1985. Approximate and probabilistic algorithms for shading and rendering structured particle systems. In *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '85)*. Association for Computing Machinery, 313–322. doi:10.1145/325334.325250

E. B. Rogers. 2001. *Landscape Design: A Cultural and Architectural History*. Harry N. Abrams, New York.

A. Runions, M. Fuhrer, B. Lane, P. Federl, A.-G. Rolland-Lagan, and P. Prusinkiewicz. 2005. Modeling and visualization of leaf venation patterns. *ACM Trans. Graph.* 24, 3 (2005), 702–711.

R. M. Smelik, T. Tutenel, R. Bidarra, and B. Benes. 2014. A survey on procedural modelling of virtual worlds. *Comp. Graph. Forum* 33, 6 (2014), 31–50.

D. Srivastava, X. Zhang, H. Wen, C. Wen, and Z. Tu. 2025. Lay-Your-Scene: Natural Scene Layout Generation with Diffusion Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 17909–17919.

O. Stava, S. Pirk, J. Kratt, B. Chen, R. Mech, O. Deussen, and B. Benes. 2014. Inverse Procedural Modelling of Trees. *Comp. Graph. Forum* 33, 6 (2014), 118–131.

C. Sun, J. Han, W. Deng, X. Wang, Z. Qin, and S. Gould. 2025. 3D-GPT: Procedural 3D Modeling with Large Language Models. In *2025 International Conference on 3D Vision (3DV)*. IEEE, 1253–1263. doi:10.1109/3DV66043.2025.00119

P. Tan, T. Fang, J. Xiao, P. Zhao, and L. Quan. 2008. Single Image Tree Modeling. *ACM Trans. on Graphics* 27, 5, Article 108 (2008), 7 pages.

Y Toparlar, Bert Blocken, B Maiheu, and B Heijmans. 2017. A review on the CFD analysis of urban microclimate. *Renewable and Sustainable Energy Reviews* 80 (2017), 1613–1640.

C. A. Vanegas, I. Garcia-Dorado, D. G. Aliaga, B. Benes, and P. Waddell. 2012. Inverse Design of Urban Procedural Models. *ACM Trans. on Graphics* 31, 6, Article 168 (2012), 11 pages.

J. Wang, N. Cao, Y. Ding, M. Xie, F. Gu, and C. Chen. 2025. SKE-Layout: Spatial Knowledge Enhanced Layout Generation with LLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 19414–19423.

J. Weber and J. Penn. 1995. Creation and rendering of realistic trees. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '95)*. Association for Computing Machinery, 119–128. doi:10.1145/218380.218427

H. Xie, Z. Hong, G. Yao, and D. Xu. 2024. CityDreamer: Compositional Generative Model of Unbounded 3D Cities. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

H. Xu, N. Gossett, and B. Chen. 2007. Knowledge and heuristic-based modeling of laser-scanned trees. In *ACM Trans. Graph.*, Vol. 26. 19.

S. Zhang, Y. Zhang, Q. Zheng, R. Ma, W. Hua, H. Bao, W. Xu, and C. Zou. 2024. 3D-SceneDreamer: Text-Driven 3D-Consistent Scene Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10170–10180.

X. Zhang, A. Shehata, B. Benes, and D. Aliaga. 2020. Automatic Deep Inference of Procedural Cities from Global-Scale Spatial Data. *ACM Trans. Spatial Algorithms Syst.* 7, 2, Article 6 (Oct. 2020), 28 pages. doi:10.1145/3423422

M. Zhou, Y. Wang, J. Hou, S. Zhang, Y. Li, C. Luo, J. Peng, and Z. Zhang. 2025. SceneX: Procedural Controllable Large-Scale Scene Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 39. 10806–10814. doi:10.1609/aaai.v39i10.33174