



“An autumm park”

scene
decomposition

“A DSLR photo of a wooden park bench.”

“A DSLR photo of an ornate stone fountain.”

⋮

“A DSLR photo of a colorful playground slide.”

freeze parameters

Point-E
(optional)

Point cloud

3D Gaussian
Filtering

3D Gaussians

rendering

add noise

Pretrained 2D Diffusion Model

U-Net

object initialize

Reconstructive Generation

Multi-timesteps Sampling

$$L_{MTS} = E_{t, \epsilon, c} [\omega(t) || \sum_{i=1}^m \epsilon_{\phi}(x_{t_i}, t_i, y) - \epsilon_{\phi}(x_{t_i}, t_i, \emptyset) ||]$$

$$\begin{aligned} &\epsilon_{\phi}(x_{t_i}, t_i, \emptyset) \\ &\epsilon_{\phi}(x_{t_i}, t_i, y) \end{aligned}$$

Formation Pattern Sampling

composition

Point cloud

Indoor
environments
initialize

3D Gaussians

layout

Stage1: surroundings

Stage2: ground

Stage3: all

Camera Sampling