



“An autumn 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

object initialize

Reconstructive Generation



U-Net

Formation Pattern Sampling

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

$\epsilon_{\phi}(x_{t_i}, t_i, \emptyset)$
 $\epsilon_{\phi}(x_{t_i}, t_i, y)$



Stage3: all

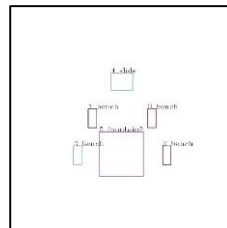


Stage2: ground



Stage1: surroundings

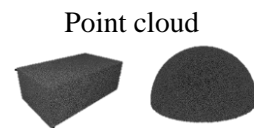
composition



layout



3D Gaussians



Indoor environments initialize
Outdoor environments initialize

Camera Sampling