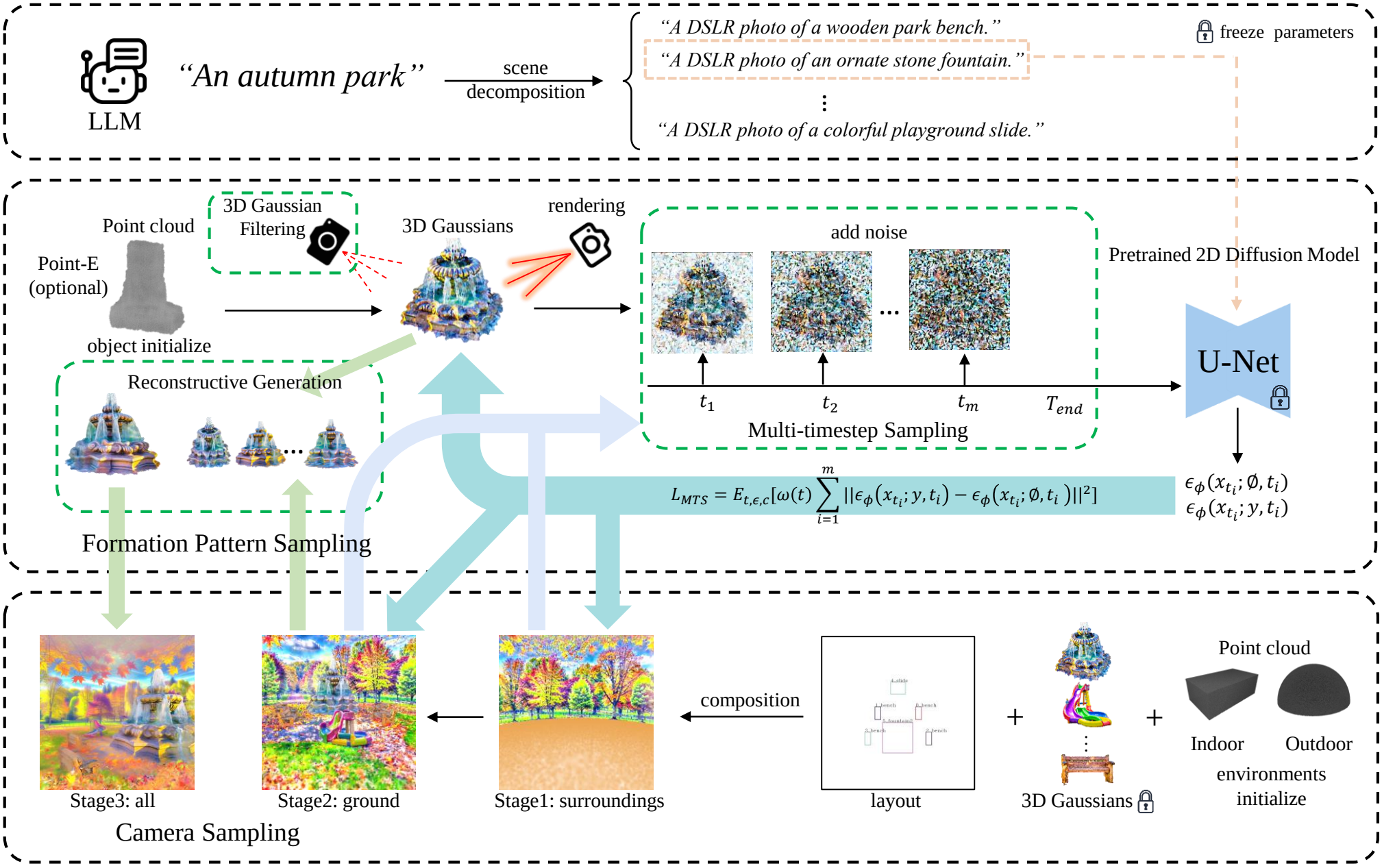
LLM
"An autumn park"
scene decomposition
"A DSLR photo of a wooden park bench."
"A DSLR photo of an ornate stone fountain."
"A DSLR photo of a colorful playground slide."
freeze parameters
Point cloud
Point-E (optional)
object initialize
3D Gaussian Filtering
3D Gaussians
rendering
add noise
Pretrained 2D Diffusion Model
U-Net
Reconstructive Generation
t_1
t_2
t_m
T_{end}
Multi-timestep Sampling
$L_{MTS} = E_{t,\epsilon,c}[\omega(t) \sum_{i=1}^{m} ||\epsilon_\phi(x_{t_i}; y, t_i) - \epsilon_\phi(x_{t_i}; \emptyset, t_i)||^2]$
$\epsilon_\phi(x_{t_i}; \emptyset, t_i)$
$\epsilon_\phi(x_{t_i}; y, t_i)$
Formation Pattern Sampling
composition
layout
3D Gaussians
Point cloud
Indoor
Outdoor
environments initialize
Stage3: all
Stage2: ground
Stage1: surroundings
Camera Sampling