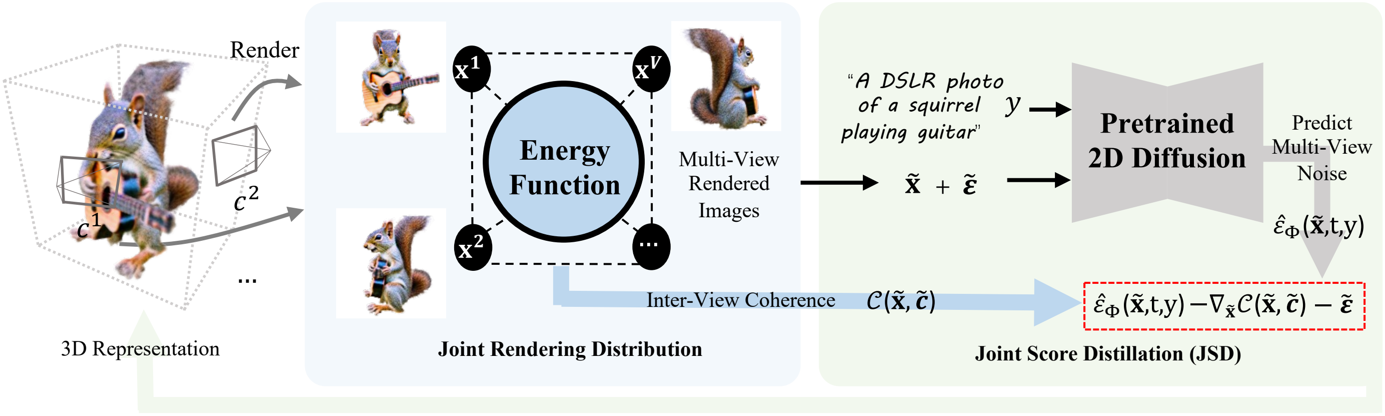
Render
c^1
c^2
...
3D Representation
$\mathbf{x}^1$
$\mathbf{x}^V$
$\mathbf{x}^2$
...
Energy Function
Multi-View Rendered Images
Joint Rendering Distribution
Inter-View Coherence
$\mathcal{C}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{c}})$
"A DSLR photo of a squirrel playing guitar"
y
$\tilde{\mathbf{x}} + \tilde{\boldsymbol{\varepsilon}}$
Pretrained 2D Diffusion
Predict Multi-View Noise
$\hat{\varepsilon}_\Phi(\tilde{\mathbf{x}}, t, y)$
$\hat{\varepsilon}_\Phi(\tilde{\mathbf{x}}, t, y) - \nabla_{\tilde{\mathbf{x}}} \mathcal{C}(\tilde{\mathbf{x}}, \tilde{\boldsymbol{c}}) - \tilde{\boldsymbol{\varepsilon}}$
Joint Score Distillation (JSD)