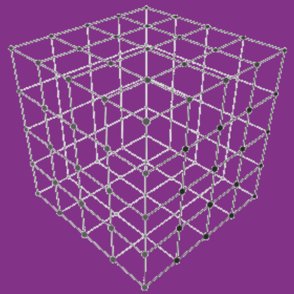
清華大學 交叉信息研究院
Institute for Interdisciplinary Information Sciences, Tsinghua University

Generative 3D Mesh modeling with text-to-texture generator



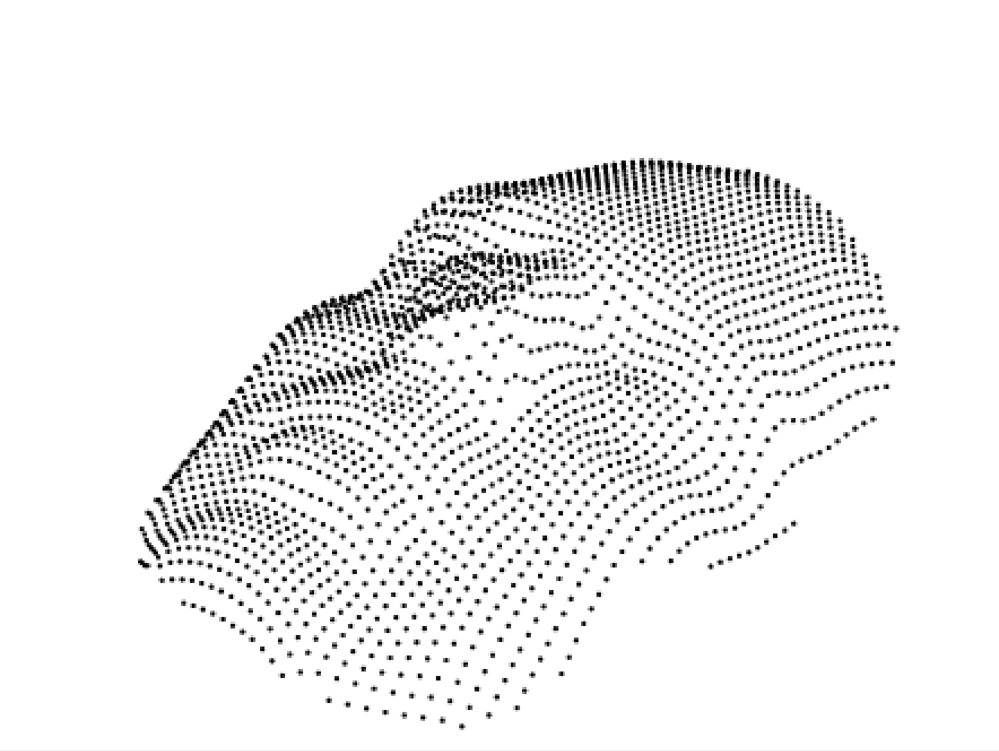
Presented By:
Ziru Huang* and *Jialuo Li

Introduction

Different 3D representation

Voxel-based

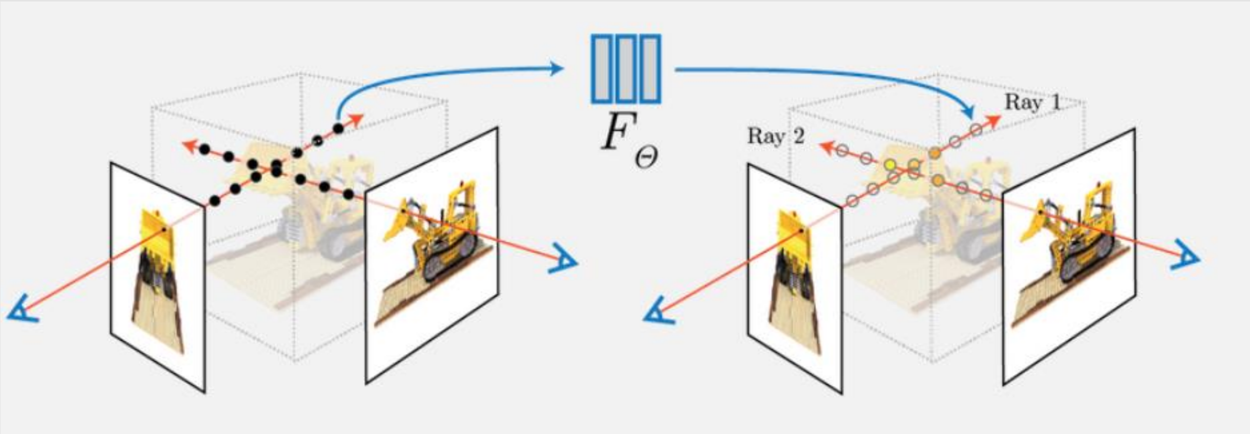
Point cloud



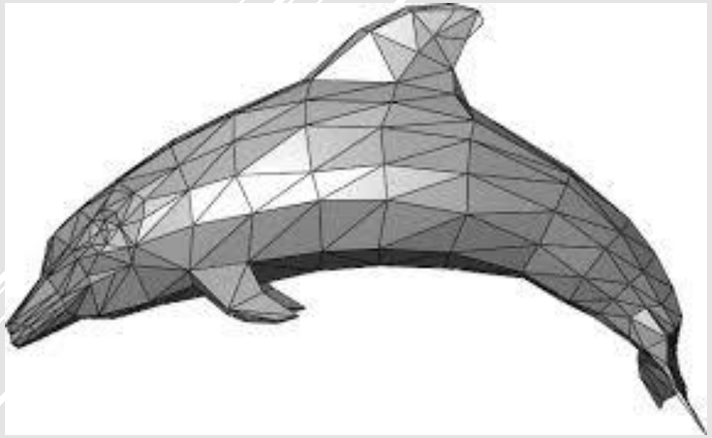
Introduction

Different 3D representation

Nerf



Mesh



What we need?

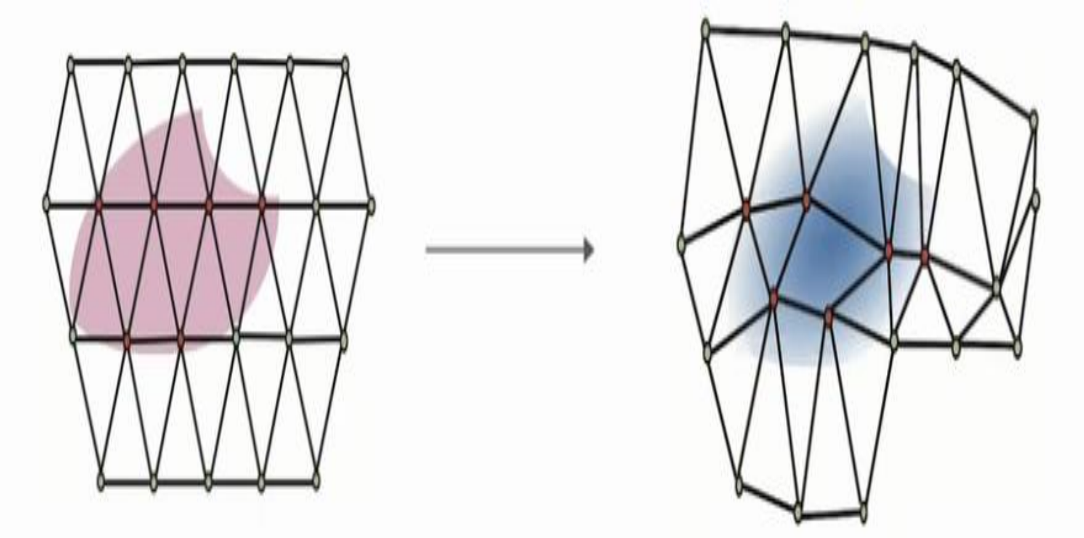
- Proper representation to generate mesh
- Proper representation for training neural network

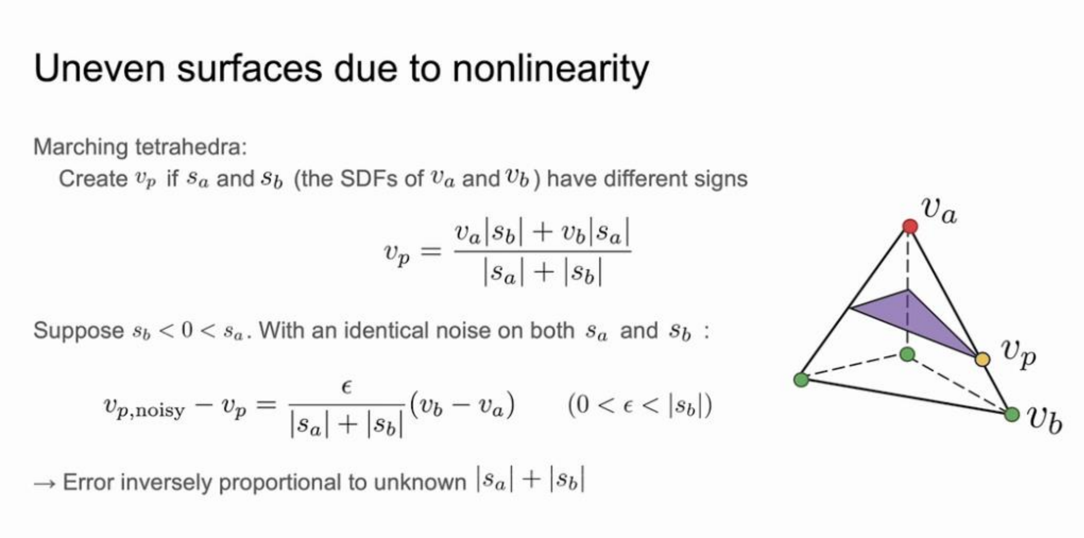
Methodology

—— Parametrizing meshes

Deep Marching Tetrahedra(DMTet): parametrize meshes with deformable tetrahedral grids

- Each vertex stores an offset (3-dim) and a SDF-value
- Use interpolation to infer the mesh



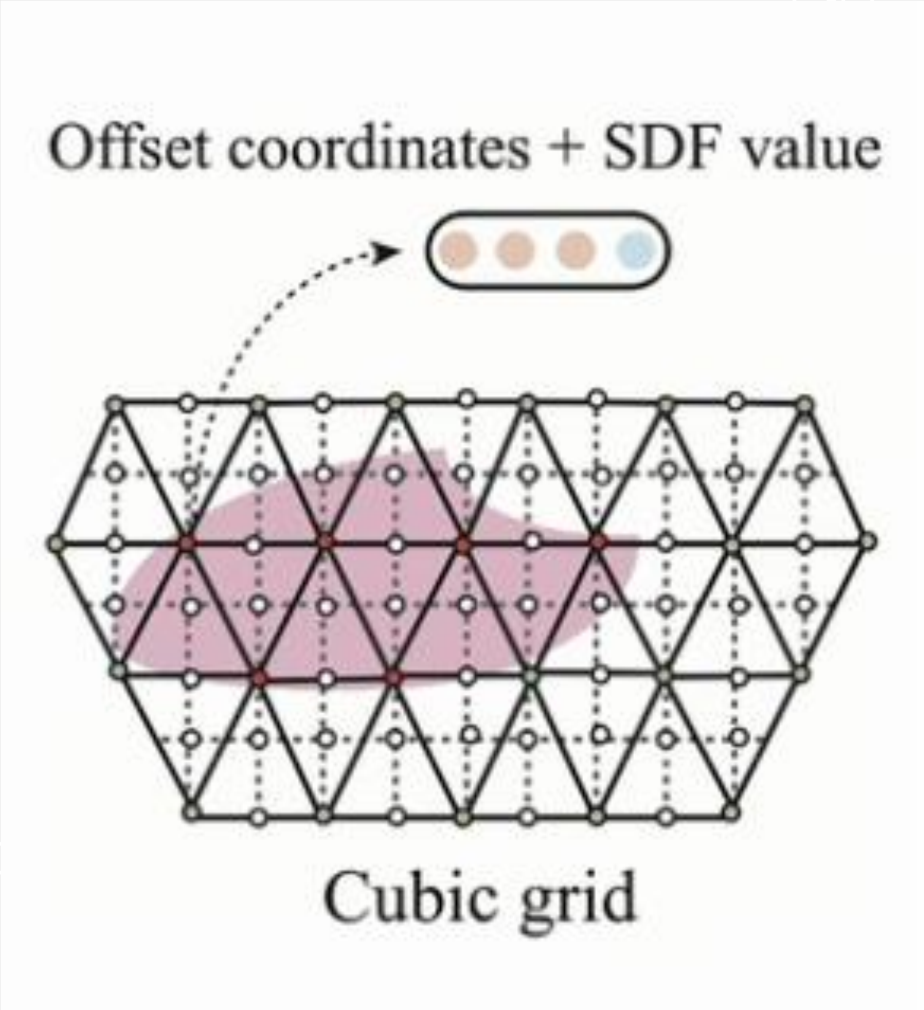


Deep Marching Tetrahedra: a Hybrid Representation for High-Resolution 3D Shape Synthesis

Methodology

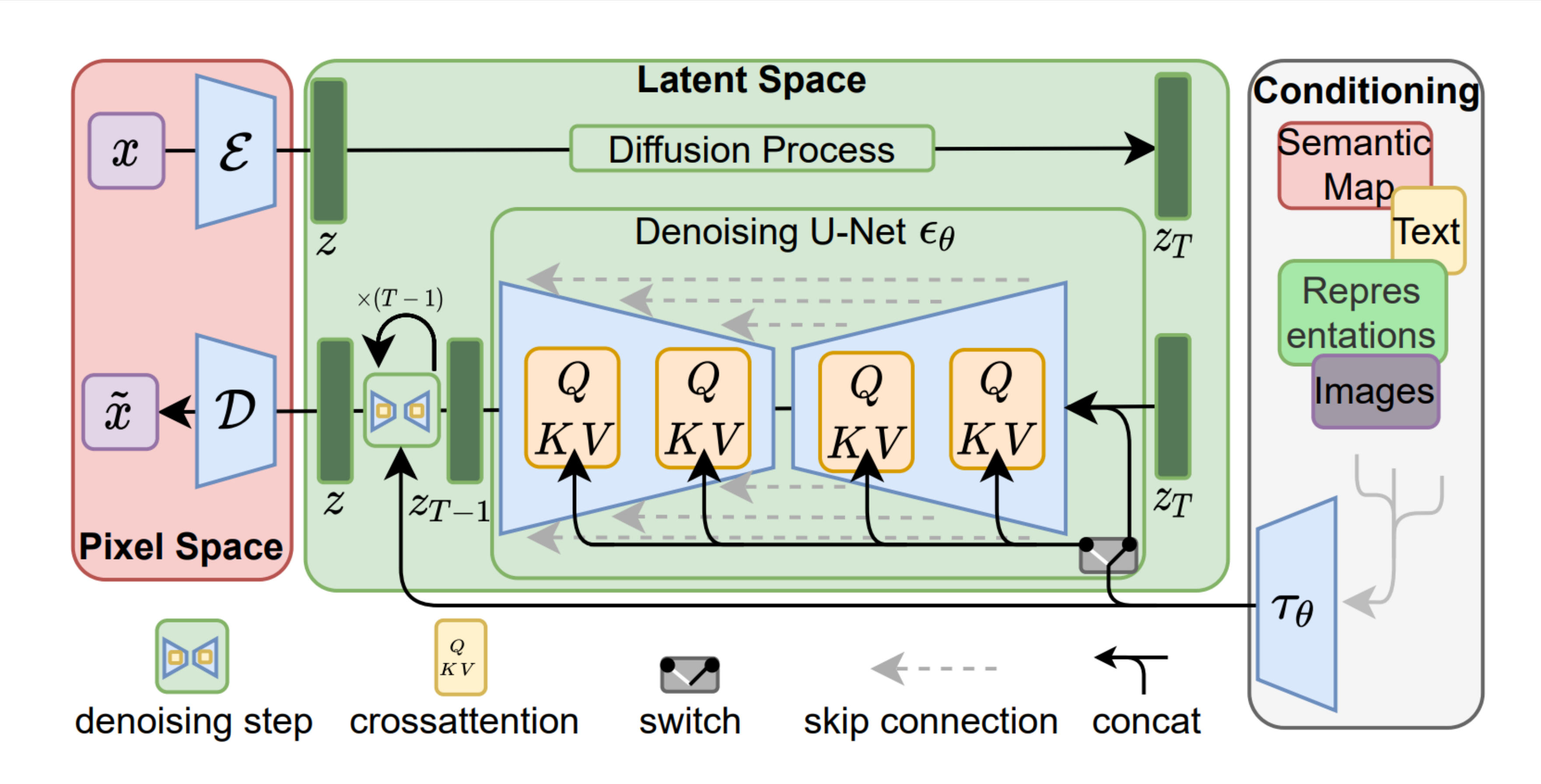
—— Parametrizing meshes

Augment tetrahedral grids to cubic grids -> 3D CNN



Methodology

—— Class-conditioned generative model



High-Resolution Image Synthesis with Latent Diffusion Models

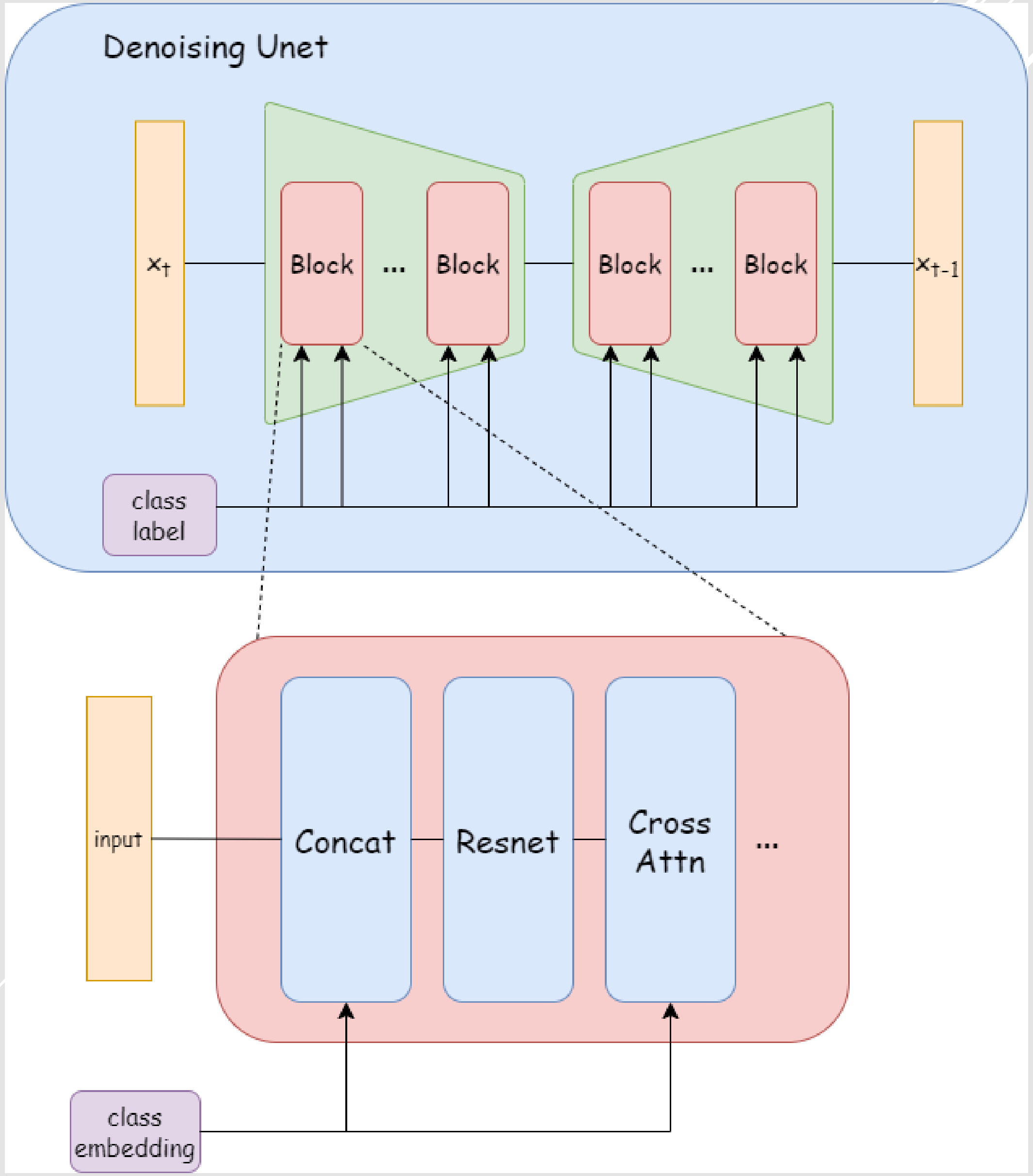
Methodology

—— Class-conditioned generative model

- Concatenate Resnet Block input and class embedding
- Use cross attention to embed class labels
- Use classifier guidance to control the generation. Train an additional classifier with only the encoder part of Unet as backbone

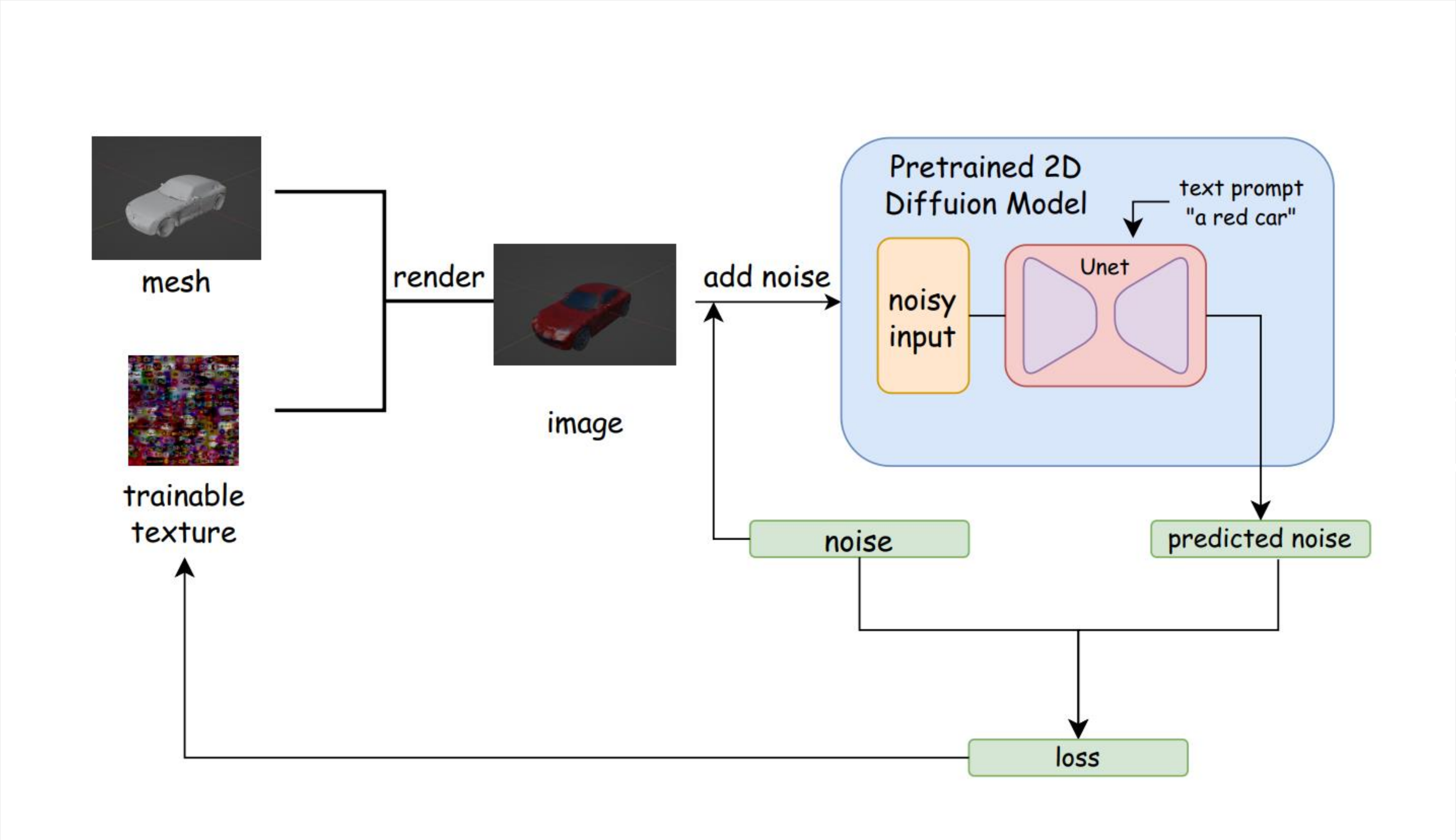
Algorithm 1 Classifier guided diffusion sampling, given a diffusion model $(\mu_\theta(x_t), \Sigma_\theta(x_t))$, classifier $p_\phi(y|x_t)$, and gradient scale s.

Input: class label y, gradient scale s
$x_T \leftarrow$ sample from $\mathcal{N}(0, \mathbf{I})$
for all t from T to 1 **do**
  $\mu, \Sigma \leftarrow \mu_\theta(x_t), \Sigma_\theta(x_t)$
  $x_{t-1} \leftarrow$ sample from $\mathcal{N}(\boxed{\mu + s\Sigma\,\nabla_{x_t} \log p_\phi(y|x_t)}, \Sigma)$
end for
return x_0



Methodology

—— Text-to-texture



- Trainable texture
- Knowledge distilled from a pretrained 2D T2I diffusion model

THANK YOU