

## Distribution-aligned MVG

Camera Pose  $\pi$



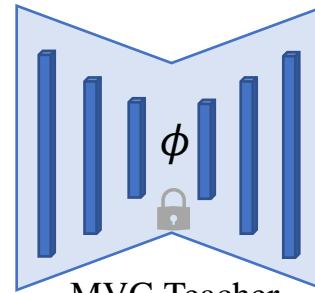
Timestep  $s$



Input Image  $y$



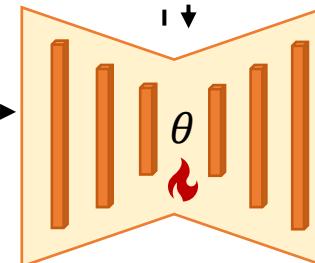
CVC Sampling



MVG Teacher

Align  
GT  
Distribution

Predict  $\hat{x}_0$



MVG Student



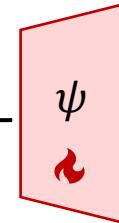
$\hat{x}_0$

Sample  $\hat{x}_{s-1}$  by 3D Renderings  $\hat{x}_0$

Recursively in 4 Steps

## Adversarially Enhanced Reconstruction

Discriminator



Real/  
Fake

GAN Loss

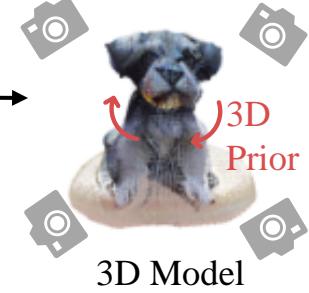


Recon  
Loss

Align  
GT  
Distribution

$\tilde{x}_0$  Render

Reconstruction  
Model



Add Noise to Timestep  $s - 1$

$$\hat{x}_{s-1} = \alpha_t \tilde{x}_0 + \beta_t \epsilon'$$