46

DO NOT PUT ANYTHING HERE

Adapting 3D Object Representations for Pre-grasp Generation

DO NOT PUT ANYTHING HERE

Al Hassan, Chaitanya Chawla, Eyob Dagnachew

DO NOT PUT ANYTHING HERE

Motivation

• Why are pre-grasps important?

- Enables robust and reliable grasps under uncertainty
- Improves motion planning and execution efficiency
- Critical for generalization across diverse objects and tasks
- Why Latent-Diffusion Models?
 - Reduces computational cost seen in pixel-space diffusion.
 - Increased ability to capture more high-level semantic details
 - Enhanced manipulation of outputs of diffusion workflow
 - Facilitates multi model diffusion models



Results: Qualitative Analysis

X-CUBE (Text input)

Our model (Image input)

Image is generated using stable diffusion from text prompt



A chair made from polished oak



An old-fashioned rotary phone







GT | Frame 0

Results: VAE Ablations

Left: Ground truth, Right: Reconstruction





V1: X-Cube-based approach







V2: Sliding Window Attention approach



Results - Inference time/Model Size



Result - 3D model output quality



Results: Pre-grasp Generation

Pre-grasps for our generated assets



Results: Grasp Quality



Future Work

- Scale up the resolution and dataset size of our model to better compare with X-CUBE
- Use native 3d sparse data structures to improve efficiency
- Confirm pre-grasp quality in simulation by producing force perturbations
- Test framework's capability for sim2real on real-robot