Geometric Algebra Grasp Diffusion for Dexterous Manipulators

Tao Zhong and Christine Allen-Blanchette

Princeton University

Overview

Dexterous grasping in unstructured environments, where objects are encountered in diverse poses, remains a key challenge in robotics. Existing methods often predict grasps in canonical object poses, limiting their effectiveness in real-world scenarios. Our novel approach introduces a **symmetry-aware diffusion-based grasp generation framework**, which uses **geometric algebra** to enforce $SE(3)$ **equivariance**, enabling robust and scalable grasp generation.

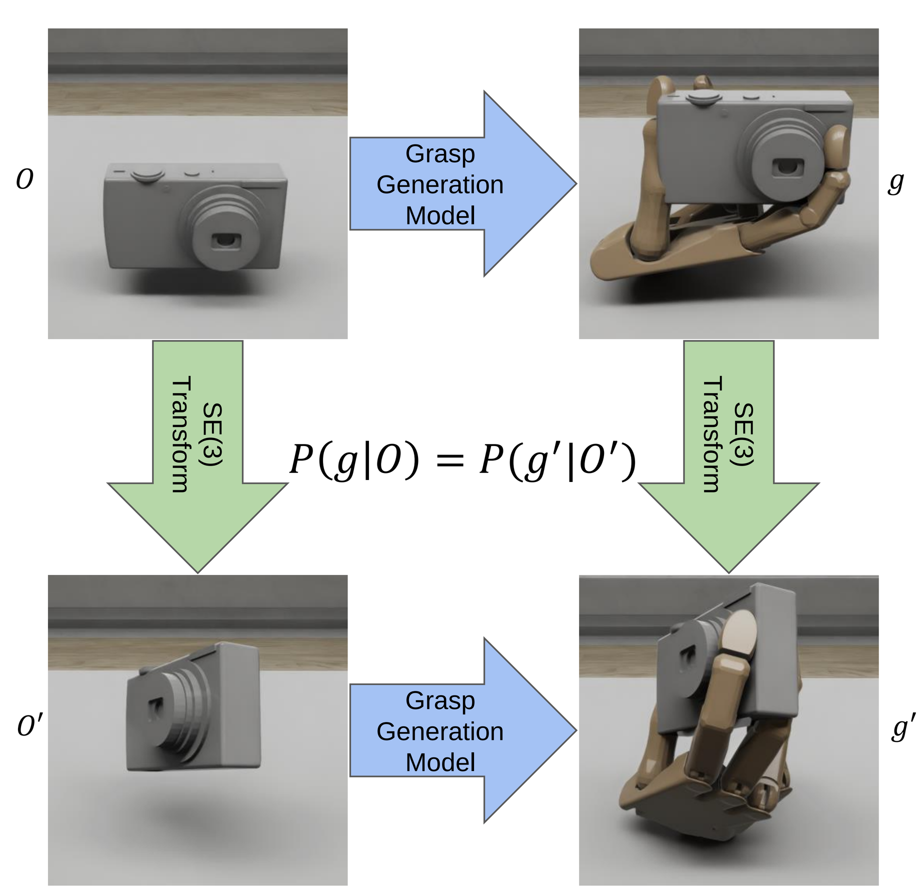

Figure 1: Our grasp generation model maintains the probability $P(\mathbf{g}|O) = P(\mathbf{g}'|O')$ under $SE(3)$ transformations.

Methodology

Grasp Generation Problem

- Given an object point cloud $O \in \mathbb{R}^{N\times 3}$, we aim to generate a dexterous grasp $\mathbf{g} = [\mathbf{r}, \mathbf{p}, \mathbf{q}]$, where $\mathbf{r}$ represents rotation, $\mathbf{p}$ represents position, and $\mathbf{q}$ represents hand joint configurations. The generation process is modeled as $p(\mathbf{g} \mid O)$.
- We ensure **equivariance** to $SE(3)$ transformations, meaning the grasp generation remains invariant to changes in object pose:

$$p([\mathbf{r}, \mathbf{p}, \mathbf{q}] \mid O) = p([\rho \cdot (\mathbf{r}, \mathbf{p}), \mathbf{q}] \mid \rho_{\mathbb{R}^3}(O)). \quad (1)$$

Geometric Algebra Framework

- We leverage 16-dimensional multivectors in the **projective geometric algebra** $\mathbb{G}_{3,0,1}$ for representing grasps and object point clouds.
- This allows us to directly encode symmetry properties into the neural network architecture shown in Fig 2 and 3, ensuring **equivariance**.

Diffusion-Based Grasp Generation

- The grasp generation process is modeled using a **conditional diffusion model**:

$$q(\mathbf{g}_{1:T} \mid \mathbf{g}_0) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{g}_t; \sqrt{1-\beta_t}\mathbf{g}_{t-1}, \beta_t I) \quad (2)$$

- Training involves minimizing the noise prediction loss:

$$L_\epsilon = \|\epsilon_\theta(\mathbf{g}_t, O, t) - \epsilon_t\|_2^2 \quad (3)$$

Physics-Informed Refinement

- A physics-informed refinement layer ensures physical plausibility with stability loss:

$$L_{\text{stability}} = \frac{1}{M} \sum_{m=1}^{M} (\|\dot{\mathbf{p}}_{\text{obj}}(T_{\text{sim}})_m\|_2^2 + \|\dot{\mathbf{r}}_{\text{obj}}(T_{\text{sim}})_m\|_2^2). \quad (4)$$

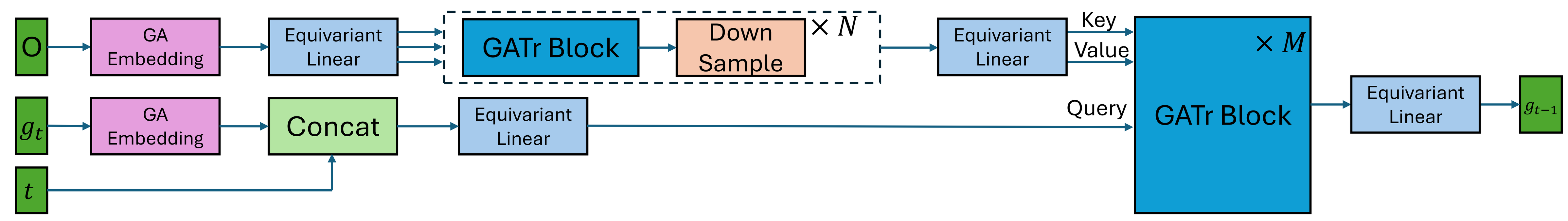

Figure 2: **Model architecture for equivariant dexterous grasp generation.** Point cloud O and grasp configuration g_t are embedded using $\mathbb{G}_{3,0,1}$ embeddings, processed through GATr [1] blocks for $SE(3)$ equivariance, with down-sampling to reduce computational load.

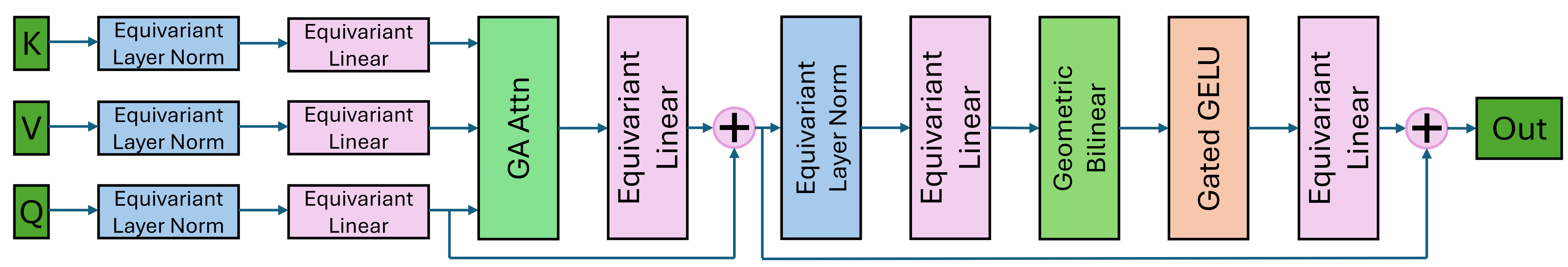

Figure 3: **The GATr Block** processes key, value, and query inputs with equivariant layers, using Geometric Algebra-based attention

Results

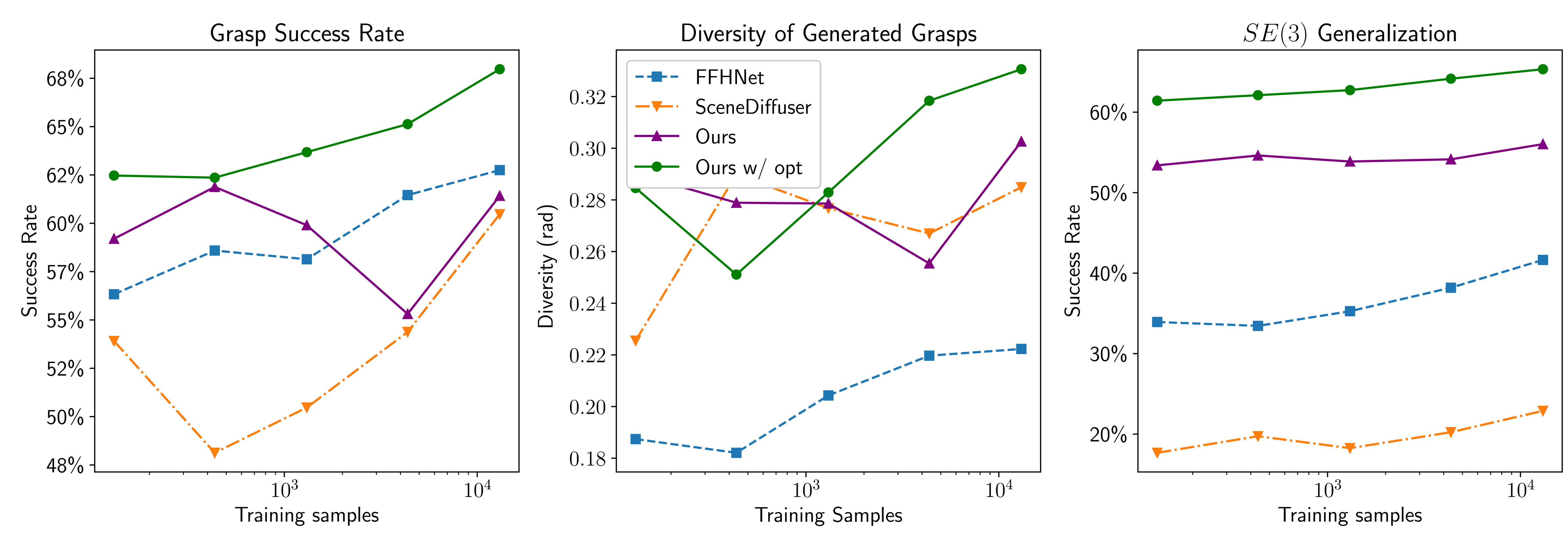

Figure 4: **Experimental Results.** Grasp success rate and diversity metrics across different data amounts. "Ours w/ opt" refers to our model with the physics-informed refinement layer.

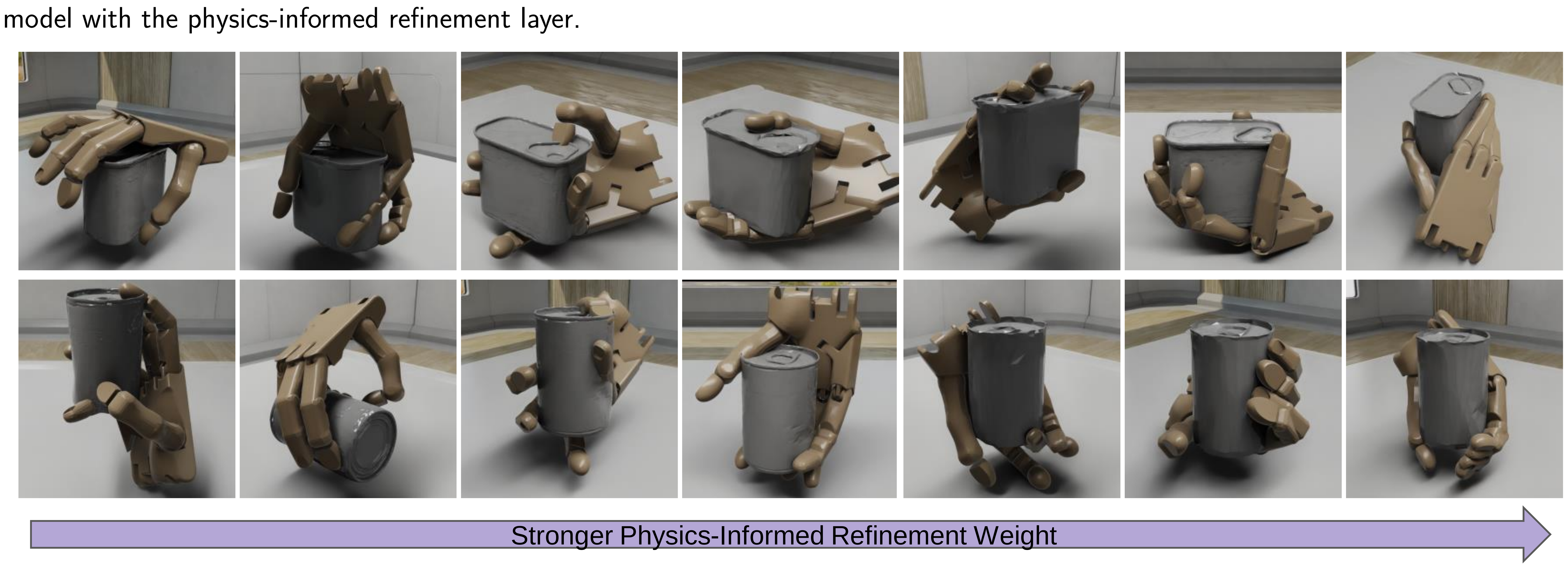

Figure 5: **Example Grasps Generated by our Method.** Our model generates stable grasps for unseen objects.

- Our method outperforms baseline methods [2, 3] in **grasp success rate** and **grasp diversity**, especially in **low-data regimes**, indicating improved data efficiency.
- Our approach remains robust when tested on objects with unseen $SE(3)$ transformations, outperforming baseline methods [2, 3]. The model's ability to generalize to out-of-distribution data ensures applicability in real-world manipulation tasks.
- We visualize grasps generated by our model, showcasing its ability to generate stable and diverse grasps for unseen objects. The refinement weight λ controls the grasp type, favoring **power grasps** for higher values and **precision fingertip grasps** for smaller values.

Ablation: Impact of Refinement

	Ours	SceneDiffuser
without Refinement	61.42	60.47
with Refinement	67.89	65.31

Table 1: Comparison of the grasp success rate with and without physics-informed refinement layer.

- The physics-informed refinement layer improves grasp stability by **5%-10%**, guiding the diffusion process toward physically feasible grasps.

References & Links

[1] Brehmer, et al. NeurIPS 2023
[2] Huang et al. CVPR 2023
[3] Meyer et al. ICRA 2022

- More details at: `https://gagrasp.github.io/`

SCAN ME