Geometric Algebra Grasp Diffusion for Dexterous Manipulators

Tao Zhong[1] and Christine Allen-Blanchette[1]

***Abstract*— We propose a novel framework for dexterous grasp generation that leverages geometric algebra representations to enforce equivariance to $SE(3)$ transformations. By encoding the $SE(3)$ symmetry constraint directly into the architecture, our method improves data and parameter efficiency, while enabling robust grasp generation across diverse object poses. Additionally, we incorporate a differentiable physics-informed refinement layer, which ensures generated grasps are physically plausible and stable. Extensive experiments demonstrate the model's superior performance in generalization, stability, and adaptability compared to existing methods. Additional details at gagrasp.github.io**

***Index Terms*— Multifingered Hands, Equivariant Neural Networks, Deep Learning in Grasping and Manipulation, Dexterous Manipulation.**

I. Introduction

Dexterous grasping remains a fundamental challenge in robotics, especially in unstructured environments with diverse object poses. Most existing datasets [1–3] contain grasps in canonical poses, limiting the generalizability of learning-based methods. These methods often require data augmentation [4, 5] or assume canonical poses [6, 7], making them prone to out-of-distribution issues.

To address this, we propose a framework that integrates geometric algebra (GA) to achieve equivariance and invariance under $SE(3)$ transformations. GA provides a unified approach to handle geometric data, directly embedding symmetry properties into the network [8, 9]. This allows us to develop a symmetry-aware diffusion model for grasp generation that can generalize more effectively across diverse poses encountered in real-world tasks.

We further incorporate a differentiable physics-informed loss to ensure that generated grasps are both geometrically feasible and physically plausible. Inspired by classifier guidance in diffusion models [10], our approach allows for direct optimization of grasps, enhancing stability and physical plausibility.

Our contributions are threefold: (1) We introduce a symmetry-aware diffusion-based grasp generation framework leveraging GA for equivariance under the broader Euclidean $E(3)$ group, (2) incorporate a differentiable physics loss for physical plausibility, and (3) demonstrate improved generalization and data efficiency through extensive experiments.

[1]Tao Zhong and Christine Allen-Blanchette are with the Department of Mechanical and Aerospace Engineering, School of Engineering and Applied Science, Princeton University, Princeton, New Jersey 08544, USA Correspondence: {tzhong, ca15}@princeton.edu

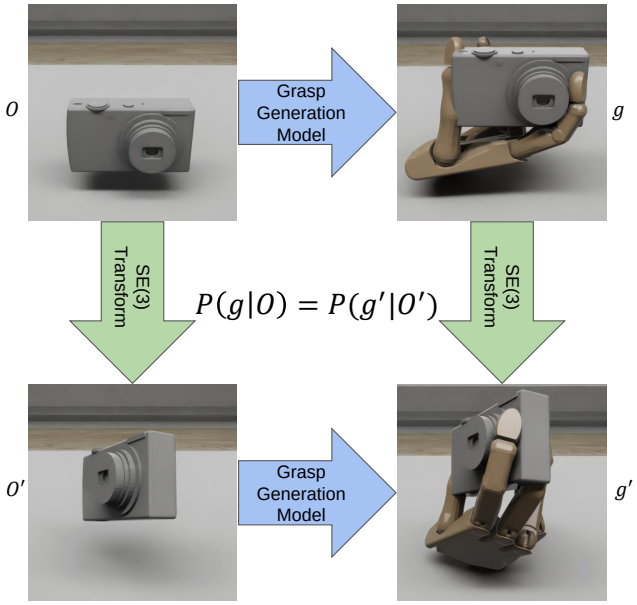

Fig. 1: Illustration of the symmetries in robotic grasping problems. Our grasp generation model leverages $SE(3)$ symmetries, maintaining the probability $P(\mathbf{g}|O) = P(\mathbf{g}'|O')$ under transformations.

II. Related Work

Vision-based Grasp Prediction methods aim to map visual inputs to grasp configurations. Traditional approaches [3, 11–14] often use a two-stage process, generating a contact heatmap followed by optimization [15]. More recent methods explore end-to-end learning via GANs [4, 5, 16, 17], VAEs [18], and diffusion models [6, 7, 19]. However, many of these assume canonical object poses, limiting generalizability. Data augmentation techniques [4, 5, 16] address this, but they increase complexity.

Diffusion Models have shown promise across various robotics tasks, including motion planning [20–23], navigation [24], and manipulation [25–27]. For grasping, Urain et al. [19] applied diffusion models to generate parallel-jaw grasps, while others extended this to dexterous grasping [6, 7]. Despite their advantages, such as stable training dynamics and high-quality sample generation, challenges like physical plausibility and handling OOD poses persist.

Equivariant Neural Networks learn representations that transform predictably with input transformations, enhancing model generalization [28–31]. Group equivariant CNNs [32, 33] are popular for implementing equivariance to rotations [34], similarity transformations [29], and Euclidean transformations [35, 36], though they have limitations in input types and architectures [37]. Recent approaches leverage

Transformation	Geometric Objects	$\mathbb{G}_{3,0,1}$ Element
Identity	Scalar	Scalar
Reflection	Plane	Vector
Rotation / Translation	Line	Bivector
Roto / Transflection	Point	Trivector
Screw	Pseudoscalar	Quadvector

TABLE I: Embeddings of group elements in $E(3)$ and common geometric objects in $\mathbb{G}_{3,0,1}$. Each transformation or geometric object can be represented by components of the multivector.

geometric algebra [8, 38] to enforce $SE(3)$ equivariance, improving generalization for unseen transformations in grasp configurations.

Differentiable Physics in Grasping has emerged as a key approach to ensuring the physical plausibility of grasps. While heuristic-based refinement steps [4, 16] are common, they can be computationally expensive. More formal approaches using differentiable physics [1, 3, 39–41] provide guarantees of physical plausibility but are often slower. In this work, we adapt the differentiable metric from Turpin et al. [1, 41] to guide the generation process, directly optimizing grasp configurations for stability and contact richness.

III. Preliminaries

A. Geometric Algebra

Geometric algebra (GA) provides a unified framework for representing geometric objects and transformations, extending vector algebra with multivectors. A multivector in 3D GA $\mathbb{G}_{3,0,0}$ can be expressed as:

$$x = x_s + \sum_{i=1}^{3} x_i e_i + \sum_{i<j} x_{ij} e_i e_j + x_{123} e_1 e_2 e_3, \quad (1)$$

where x_s, x_i, x_{ij}, and x_{123} are real coefficients. In our work, we use the projective geometric algebra $\mathbb{G}_{3,0,1}$ [8, 38], which extends the 3D GA $\mathbb{G}_{3,0,0}$ with a fourth homogeneous coordinate $x_0 e_0$, resulting in a 16-dimensional multivector in GA representation. The geometric product in GA satisfies closure, associativity, distributivity, and other key properties, making it ideal for representing transformations such as rotations and translations. In $\mathbb{G}_{3,0,1}$, transformations can be represented uniformly as multivectors, with the action of a transformation u on a geometric object v given by the sandwich product $u[v] = uvu^{-1}$. For example, when applied to dexterous grasp generation, these operations allow us to describe object orientations and hand configurations in a way that is consistent across different reference frames, enhancing model robustness to changes in pose. Table I summarizes common geometric objects and their corresponding GA elements.

For a more comprehensive study of geometric algebra and its applications, we direct readers to [8, 9, 38, 42].

B. Problem Formulation

We aim to generate dexterous grasps for objects represented by point clouds. Given a point cloud $O \in \mathbb{R}^{N\times 3}$, the goal is to sample grasps G parameterizing a dexterous hand's pose. Each grasp $\mathbf{g} \in G$ is represented by $[\mathbf{r}, \mathbf{p}, \mathbf{q}]$, where $\mathbf{r} \in \mathbb{R}^6$ and $\mathbf{p} \in \mathbb{R}^3$ represent the rotation and translation of the hand base, respectively, and $\mathbf{q} \in \mathbb{R}^k$ denotes the joint configurations. We model the grasp generation process as learning a distribution $p(\mathbf{g} \mid O)$ conditioned on the observed point cloud.

C. Equivariance and Invariance in Grasp Generation

In grasp generation, we require that the mapping from object representation to hand base pose be equivariant under $SE(3)$ transformations, while the mapping to joint configurations should be invariant. A function $f : X \to Y$ is said to be equivariant with respect to the transformation group $\mathcal{G}$ if $f(\rho_X(x)) = \rho_Y(f(x))$ for a group element ρ and an input x, where ρ_Z denotes the action of the group element ρ on the space Z. Invariance is defined as $f(\rho_X(x)) = \rho_Y(f(x)) = f(x)$. This ensures that the generated grasps remain consistent under transformations, mathematically expressed as:

$$p([\mathbf{r}, \mathbf{p}, \mathbf{q}] \mid O) = p([\rho \cdot (\mathbf{r}, \mathbf{p}), \mathbf{q}] \mid \rho_{\mathbb{R}^3}(O)), \quad (2)$$

where $\rho \in SE(3)$, and $g \cdot h$ denotes the composition of group elements g, and $h \in \mathcal{G}$.

IV. Methodology

We introduce a framework for generating dexterous grasps using a conditional diffusion model that handles the high-dimensional grasp space while ensuring robustness to variations in object poses.

A. Diffusion-based Grasp Generation

Our pipeline generates high-quality grasps by iteratively refining sampled grasps through a diffusion process [43]. Starting with a grasp $\mathbf{g}_0 = [\mathbf{r}_0, \mathbf{p}_0, \mathbf{q}_0]$, it is diffused into Gaussian noise over T timesteps:

$$q(\mathbf{g}_{1:T} \mid \mathbf{g}_0) = \prod_{t=1}^{T} \mathcal{N}(\mathbf{g}_t; \sqrt{1-\beta_t}\mathbf{g}_{t-1}, \beta_t I) \quad (3)$$

where β_t is the scheduled noise variance at timestep t.

To recover the original grasp, the model iteratively removes the added noise using a noise predictor ϵ_θ:

$$L_\epsilon = \|\epsilon_\theta(\mathbf{g}_t, O, t) - \epsilon_t\|_2^2 \quad (4)$$

where ϵ_t is the ground-truth noise at timestep t.

The denoising process is defined as:

$$p_\theta(\mathbf{g}_{0:T} \mid O) = p(\mathbf{g}_T) \prod_{t=1}^{T} \mathcal{N}(\mathbf{g}_{t-1}; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t) \quad (5)$$

where $\boldsymbol{\mu}_t = \mu_\theta(\mathbf{g}_t, O, t)$ and $\boldsymbol{\Sigma}_t = \Sigma_\theta(\mathbf{g}_t, O, t)$ can be inferred from $\epsilon_\theta(\mathbf{g}_t, O, t)$ and the parameters of the diffusion process.

B. Equivariant Model Architecture

Building on the principle that an equivariant denoising model ensures an invariant marginal distribution [38, 44], we incorporate symmetries into our model architecture using

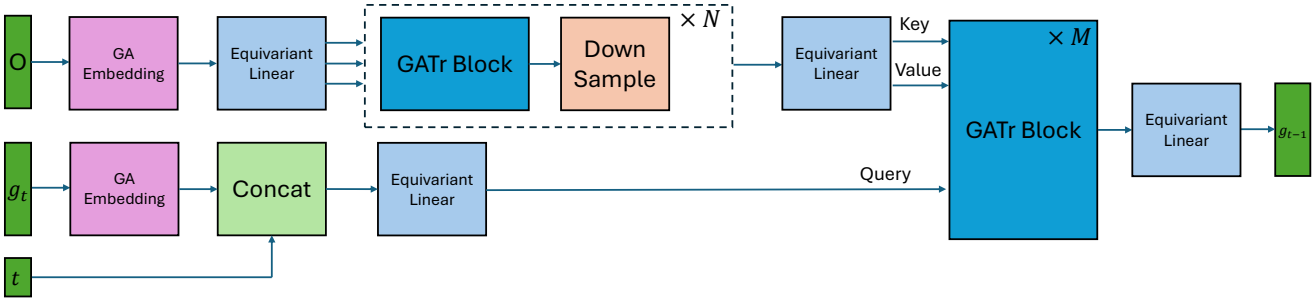

Fig. 2: Model architecture for equivariant dexterous grasp generation. Point cloud O and grasp configuration g_t are embedded using $\mathbb{G}_{3,0,1}$ embeddings, processed through GATr blocks for $SE(3)$ equivariance, with down-sampling to reduce computational load.

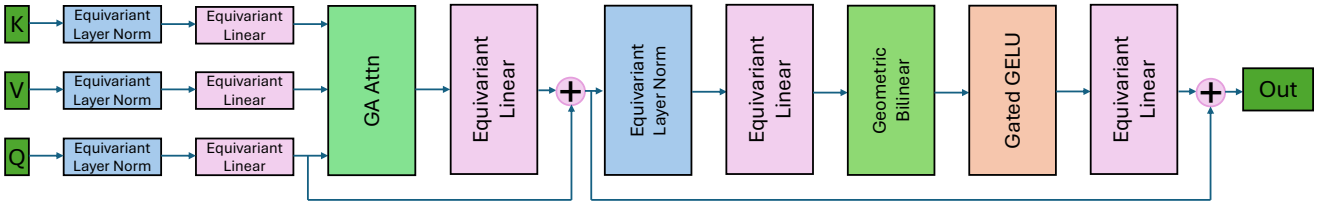

Fig. 3: The GATr Block processes key, value, and query inputs with equivariant layers, using Geometric Algebra-based attention and nonlinearities to maintain $SE(3)$ equivariance for the grasp configuration.

multivector representations in $\mathbb{G}_{3,0,1}$ leveraging the primitives defined by Brehmer et al. [38], as shown in Figure 2 and 3.

Equivariant Linear Layers ensure geometric properties are preserved:

$$\phi(x) = \sum_{k=0}^{4} w_k \langle x \rangle_k + \sum_{k=0}^{3} v_k e_0 \langle x \rangle_k \tag{6}$$

where $\langle x \rangle_k$ is the grade projection that isolates the k-grade components of the multivector x, and w_k and v_k are learnable parameters.

Geometric Bilinears combine geometric products and join operations. The geometric product xy allows for the mixing of different grades, and the join operation $\text{Join}(x, y) = (x^* \wedge y^*)^*$, where x^* is the dual of x and $x \wedge y$ denotes the exterior (wedge) product between x and y, is necessary for constructing meaningful geometric features. The geometric bilinear layer is then defined as: $\text{Geometric}(x, y; z) = \text{Concat}(xy, z_{0123}(x^* \wedge y^*)^*)$, where z_{0123} is the pseudoscalar component of a reference multivector z calculated from the network inputs.

Equivariant Attention adapts dot-product attention to multivectors. Given multivector-valued query, key, and value tensors (Q, K, V) with n_i tokens and n_c channels, attention scores are computed using the inner product in $\mathbb{G}_{3,0,1}$:

$$\text{Attention}(Q, K, V)_{i'c'} = \sum_i \text{Softmax}\left(\frac{\sum_c \langle Q_{i'c'}, K_{ic'} \rangle}{\sqrt{8n_c}}\right) V_{ic'} \tag{7}$$

where i, i' index tokens, c, c' index channels, and $\langle \cdot, \cdot \rangle$ is the invariant inner product, a regular dot product on 8 of 16 dimensions without e_0 terms.

Equivariant Nonlinearities and Normalization are crucial for preserving the geometric properties during transformations. We use scalar-gated GELU (Gaussian Error Linear Unit) nonlinearities [45]: $\text{GatedGELU}(x) = \text{GELU}(x_1)x$, where x_1 is the scalar component of the multivector x. Normalization is achieved through an E(3)-equivariant LayerNorm:

$$\text{LayerNorm}(x) = \frac{x}{\sqrt{\mathbb{E}_c[\langle x, x \rangle]}}$$

where the expectation $\mathbb{E}_c$ of the inner product is taken over the channels.

Down-Sampling Layers reduce the point set size for computational efficiency, inspired by PointTransformer [46]. The process includes farthest point sampling (FPS) in xyz space, followed by k-nearest neighbors (kNN) pooling, with max pooling on the scalar component of multivectors to retain critical geometric information.

Symmetry Breaking adjusts the model to handle cases where full $E(3)$ symmetry is unnecessary, particularly in grasping tasks requiring $SE(3)$ symmetry. This is achieved by introducing pseudoscalar features to encode handedness, enabling the model to better adapt to specific tasks.

C. Physics-Informed Differentiable Refinement Layer

Inspired by Turpin et al. [1, 41], our refinement layer uses gradients from a differentiable physics simulator [47, 48] to iteratively adjust grasp configurations. The object is initialized with velocity $\dot{\mathbf{p}}_{\text{obj}}(0)$ in pose $(\mathbf{r}_{\text{obj}}(0), \mathbf{p}_{\text{obj}}(0))$ with a grasp $\mathbf{g}$ being executed. After T_{sim} timesteps, final velocities $(\dot{\mathbf{r}}_{\text{obj}}(T_{\text{sim}}), \dot{\mathbf{p}}_{\text{obj}}(T_{\text{sim}}))$ are computed across M simulations with varying initial velocities. The stability loss we are trying to optimize is then given by:

$$L_{\text{stability}} = \frac{1}{M} \sum_{m=1}^{M} \left(\left\| \dot{\mathbf{p}}_{\text{obj}}(T_{\text{sim}})_m \right\|_2^2 + \left\| \dot{\mathbf{r}}_{\text{obj}}(T_{\text{sim}})_m \right\|_2^2 \right). \tag{8}$$

Additionally, joint configuration constraints are enforced via:

$$L_{\text{range}} = \left\| \mathbf{q} - \frac{\mathbf{q}^{\text{up}} + \mathbf{q}^{\text{low}}}{2} \right\|_2^2, \quad L_{\text{limit}} = \max(\mathbf{q} - \mathbf{q}^{\text{up}}, 0) + \max(\mathbf{q} - \mathbf{q}^{\text{low}}, 0) \tag{9}$$

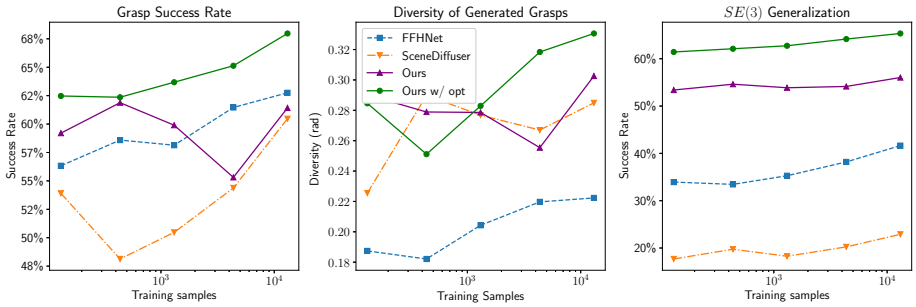

Fig. 4: Experimental Results. Grasp success rate and diversity metrics across different data amounts. "Ours w/ opt" refers to our model with the physics-informed refinement layer. **Left:** Our model outperforms others in grasp success rate, especially with fewer training samples. **Middle:** Our model generates more diverse grasps. **Right:** $SE(3)$ generalization performance, highlighting robustness in handling out-of-distribution data with random $SE(3)$ transformations.

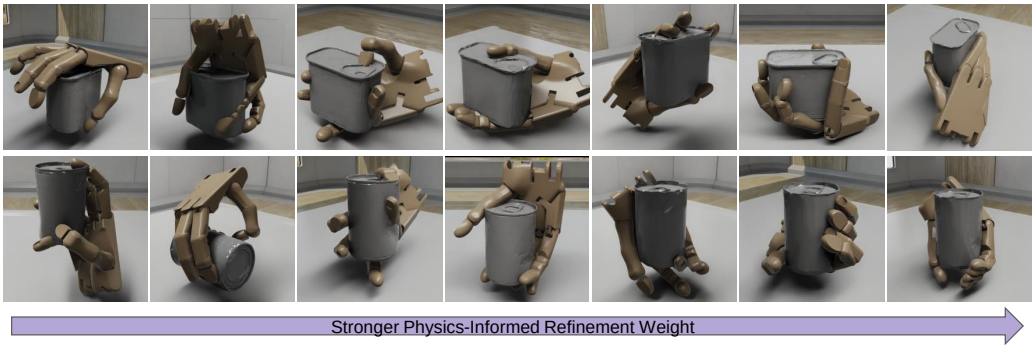

Fig. 5: Example Grasps Generated by our Method. Our model generates stable grasps for unseen objects. As the Physics-Informed Refinement weight λ increases (left to right), the model produces more stable power grasps, while a smaller λ results in precision fingertip grasps, showing adaptability in grasp types.

	Ours	**SceneDiffuser** [7]
without Refinement	61.42	60.47
with Refinement	67.89	65.31

TABLE II: Comparison of the grasp success rate with and without physics-informed refinement layer for our model and SceneDiffuser.

where $\mathbf{q}^{\text{up}}$ and $\mathbf{q}^{\text{low}}$ denote the upper and lower limit of the joints, respectively.

The overall objective function is: $L_{\text{phys}} = L_{\text{stability}} + \alpha_1 L_{\text{range}} + \alpha_2 L_{\text{limit}}$, where α_1 and α_2 are hyperparameters.

The gradients of this objective function adjust the mean during denoising: $\hat{\boldsymbol{\mu}}_t = \boldsymbol{\mu}_t + \lambda \nabla_g L_{\text{phys}}$, where $\boldsymbol{\mu}_t$ is defined as in Equation 5, and λ is a scaling factor that controls the influence of the physics-based optimization.

V. Experiments

In this section, we explore the effectiveness of our framework through experiments designed to answer the following questions: (1) Does the use of geometric algebra (GA) representation enhance data and parameter efficiency? (2) How does the method perform on out-of-distribution (OOD) data with random $SE(3)$ transformations? (3) How much does the physics-informed refinement layer contribute to grasp quality and stability?

We evaluate our method using the Shadowhand subset of the MultiDex [3] dataset, containing 16,069 dexterous grasps for 58 objects, split into 48 training and 10 test objects. Grasp configurations are represented by $\mathbf{g} = [\mathbf{r}, \mathbf{p}, \mathbf{q}] \in \mathbb{R}^{33}$, with objects represented by point clouds $O \in \mathbb{R}^{2048\times 3}$. We follow Li et al. [3] and assess grasp success rates in IssacGym [49], testing grasps under external forces along $\pm xyz$ axes, with success defined by stability across all tests. We also measure diversity as the mean standard deviation among all revolute joints. Baseline models include SceneDiffuser [7], FFHNet [5], and our model without the refinement layer.

Our model outperforms baselines in both grasp success rates and diversity, particularly in low-data regimes, as shown in Figure 4. Despite having about 20% of the parameters of FFHNet [5], it performs on par with or better than the baseline models. The physics-informed refinement layer further boosts success rates by 5%-10% (Table II) and enhances robustness to $SE(3)$ transformations, making our approach more adaptable to OOD scenarios. Qualitative results (Figure 5) demonstrate the flexibility of our model in generating stable grasps, where stronger λ values produce more stable power grasps, and smaller λ values favor precision fingertip grasps. This adaptability highlights the model's effectiveness in varying grasping tasks.

VI. Conclusion

We introduce a symmetry-aware, diffusion-based framework for dexterous grasp generation leveraging geometric algebra. Our approach enhances generalization to out-of-distribution data, improves data and parameter efficiency, and produces physically plausible grasps via a physics-informed refinement layer. The model's robustness and flexibility make it well-suited for real-world robotic tasks, showing significant improvements over existing methods. Future work will explore further optimizations and broader applications of our framework in manipulation scenarios.

References

[1] D. Turpin, T. Zhong, S. Zhang, G. Zhu, E. Heiden, M. Macklin, S. Tsogkas, S. Dickinson, and A. Garg, "Fast-grasp'd: Dexterous multi-finger grasp generation through differentiable simulation," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8082–8089.

[2] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen, "The columbia grasp database," in *2009 IEEE international conference on robotics and automation*. IEEE, 2009, pp. 1710–1716.

[3] P. Li, T. Liu, Y. Li, Y. Geng, Y. Zhu, Y. Yang, and S. Huang, "Gendexgrasp: Generalizable dexterous grasping," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 8068–8074.

[4] J. Lundell, F. Verdoja, and V. Kyrki, "Ddgc: Generative deep dexterous grasping in clutter," *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6899–6906, 2021.

[5] V. Mayer, Q. Feng, J. Deng, Y. Shi, Z. Chen, and A. Knoll, "Ffhnet: Generating multi-fingered robotic grasps for unknown objects in real-time," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 762–769.

[6] Z. Weng, H. Lu, D. Kragic, and J. Lundell, "Dexdiffuser: Generating dexterous grasps with diffusion models," *arXiv preprint arXiv:2402.02989*, 2024.

[7] S. Huang, Z. Wang, P. Li, B. Jia, T. Liu, Y. Zhu, W. Liang, and S.-C. Zhu, "Diffusion-based generation, optimization, and planning in 3d scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16 750–16 761.

[8] D. Ruhe, J. K. Gupta, S. De Keninck, M. Welling, and J. Brandstetter, "Geometric clifford algebra networks," in *International Conference on Machine Learning*. PMLR, 2023, pp. 29 306–29 337.

[9] D. Ruhe, J. Brandstetter, and P. Forré, "Clifford group equivariant neural networks," *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[10] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[11] J. Varley, J. Weisz, J. Weiss, and P. Allen, "Generating multi-fingered robotic grasps via deep learning," in *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2015, pp. 4415–4420.

[12] S. Brahmbhatt, A. Handa, J. Hays, and D. Fox, "Contactgrasp: Functional multi-finger grasp synthesis from contact," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2386–2393.

[13] J. Lu, H. Kang, H. Li, B. Liu, Y. Yang, Q. Huang, and G. Hua, "Ugg: Unified generative grasping," *arXiv preprint arXiv:2311.16917*, 2023.

[14] A. Wu, M. Guo, and K. Liu, "Learning diverse and physically feasible dexterous grasps with generative model and bilevel optimization," in *Conference on Robot Learning*. PMLR, 2023, pp. 1938–1948.

[15] A. T. Miller and P. K. Allen, "Graspit! a versatile simulator for robotic grasping," *IEEE Robotics & Automation Magazine*, vol. 11, no. 4, pp. 110–122, 2004.

[16] J. Lundell, E. Corona, T. N. Le, F. Verdoja, P. Weinzaepfel, G. Rogez, F. Moreno-Noguer, and V. Kyrki, "Multi-fingan: Generative coarse-to-fine sampling of multi-finger grasps," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 4495–4501.

[17] E. Corona, A. Pumarola, G. Alenya, F. Moreno-Noguer, and G. Rogez, "Ganhand: Predicting human grasp affordances in multi-object scenes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5031–5041.

[18] A. Mousavian, C. Eppner, and D. Fox, "6-dof graspnet: Variational grasp generation for object manipulation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 2901–2910.

[19] J. Urain, N. Funk, J. Peters, and G. Chalvatzaki, "Se (3)-diffusionfields: Learning smooth cost functions for joint grasp and motion optimization through diffusion," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5923–5930.

[20] J. Carvalho, M. Baierl, J. Urain, and J. Peters, "Conditioned score-based models for learning collision-free trajectory generation," in *NeurIPS 2022 Workshop on Score-Based Methods*, 2022.

[21] X. Fang, C. Garrett, C. Eppner, T. Lozano-Pérez, L. Kaelbling, and D. Fox, "Dimsam: Diffusion models as samplers for task and motion planning under partial observability," in *CoRL 2023 Workshop on Learning Effective Abstractions for Planning (LEAP)*, 2023.

[22] J. Carvalho, A. T. Le, M. Baierl, D. Koert, and J. Peters, "Motion planning diffusion: Learning and planning of robot motions with diffusion models," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2023, pp. 1916–1923.

[23] M. Janner, Y. Du, J. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," in *International Conference on Machine Learning*. PMLR, 2022, pp. 9902–9915.

[24] A. Sridhar, D. Shah, C. Glossop, and S. Levine, "Nomad: Goal masked diffusion policies for navigation and exploration," in *First Workshop on Out-of-Distribution Generalization in Robotics at CoRL 2023*, 2023.

[25] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[26] Z. Xian, N. Gkanatsios, T. Gervet, and K. Fragkiadaki, "Unifying diffusion models with action detection transformers for multi-task robotic manipulation," in *7th Annual Conference on Robot Learning*, 2023.

[27] U. Mishra and Y. Chen, "Reorientdiff: Diffusion model based reorientation for object manipulation," in *CoRL 2023 Workshop on Learning Effective Abstractions for Planning (LEAP)*, 2023.

[28] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, "Learning so (3) equivariant represen-

tations with spherical cnns," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 52–68.

[29] J. Lei, C. Deng, K. Schmeckpeper, L. Guibas, and K. Daniilidis, "Efem: Equivariant neural field expectation maximization for 3d object segmentation without scene supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 4902–4912.

[30] B. Elesedy, "Group symmetry in pac learning," in *ICLR 2022 workshop on geometrical and topological representation learning*, 2022.

[31] S. Zhu, B. An, and F. Huang, "Understanding the generalization benefit of model invariance from a data perspective," *Advances in Neural Information Processing Systems*, vol. 34, pp. 4328–4341, 2021.

[32] T. S. Cohen, M. Geiger, and M. Weiler, "A general theory of equivariant cnns on homogeneous spaces," *Advances in neural information processing systems*, vol. 32, 2019.

[33] R. Kondor and S. Trivedi, "On the generalization of equivariance and convolution in neural networks to the action of compact groups," in *International conference on machine learning*. PMLR, 2018, pp. 2747–2755.

[34] T. S. Cohen, M. Geiger, J. Köhler, and M. Welling, "Spherical cnns," *arXiv preprint arXiv:1801.10130*, 2018.

[35] H. Ryu, J. Kim, J. Chang, H. S. Ahn, J. Seo, T. Kim, J. Choi, and R. Horowitz, "Diffusion-edfs: Bi-equivariant denoising generative modeling on se (3) for visual robotic manipulation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[36] J. Brehmer, J. Bose, P. De Haan, and T. S. Cohen, "Edgi: Equivariant diffusion for planning with embodied agents," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[37] C. Deng, O. Litany, Y. Duan, A. Poulenard, A. Tagliasacchi, and L. J. Guibas, "Vector neurons: A general framework for so (3)-equivariant networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 200–12 209.

[38] J. Brehmer, P. De Haan, S. Behrends, and T. S. Cohen, "Geometric algebra transformer," *Advances in Neural Information Processing Systems*, vol. 36, 2023.

[39] M. Liu, Z. Pan, K. Xu, K. Ganguly, and D. Manocha, "Deep differentiable grasp planner for high-dof grippers," in *Robotics: Science and Systems*, 2020.

[40] T. Liu, Z. Liu, Z. Jiao, Y. Zhu, and S.-C. Zhu, "Synthesizing diverse and physically stable grasps with arbitrary hand structures using differentiable force closure estimator," *IEEE Robotics and Automation Letters*, vol. 7, no. 1, pp. 470–477, 2021.

[41] D. Turpin, L. Wang, E. Heiden, Y.-C. Chen, M. Macklin, S. Tsogkas, S. Dickinson, and A. Garg, "Grasp'd: Differentiable contact-rich grasp synthesis for multi-fingered hands," in *European Conference on Computer Vision*. Springer, 2022, pp. 201–221.

[42] L. Dorst, D. Fontijne, and S. Mann, *Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2007.

[43] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[44] J. Köhler, L. Klein, and F. Noé, "Equivariant flows: exact likelihood generative learning for symmetric densities," in *International conference on machine learning*. PMLR, 2020, pp. 5361–5370.

[45] D. Hendrycks and K. Gimpel, "Gaussian error linear units (gelus)," *arXiv preprint arXiv:1606.08415*, 2016.

[46] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun, "Point transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 16 259–16 268.

[47] M. Macklin, "Warp: A high-performance python framework for gpu simulation and graphics," https://github.com/nvidia/warp, March 2022, nVIDIA GPU Technology Conference (GTC).

[48] E. Heiden, M. Macklin, Y. S. Narang, D. Fox, A. Garg, and F. Ramos, "DiSECt: A Differentiable Simulation Engine for Autonomous Robotic Cutting," in *Proceedings of Robotics: Science and Systems*, Virtual, July 2021.

[49] J. Liang, V. Makoviychuk, A. Handa, N. Chentanez, M. Macklin, and D. Fox, "Gpu-accelerated robotic simulation for distributed reinforcement learning," in *Conference on Robot Learning*. PMLR, 2018, pp. 270–282.