Gaussian Splatting as $SE(3)$ Equivariant Features for Imitation Learning

Sanghyun Hahn [1 2] **Taekyun Ha** [1 3] **Inhee Lee** [1 4]

Abstract

We present a novel approach to one-shot imitation learning for robotic manipulation that generalizes from a single expert demonstration to unseen object instances and poses. Given multi-view RGB observations and the full trajectory of single pick-and-place task, our method enables a robot to perform similar tasks involving novel objects and novel SE(3) configurations. While prior works employ SE(3)-equivariant descriptor fields to enhance sample efficiency, these approaches often demand task-specific neural network training and are computationally intensive. Our method instead leverages Gaussian Splatting, a compact 3D scene representation that naturally supports SE(3) equivariance while encoding rich geometric and appearance details. By avoiding task-specific model training, our lightweight 3D reconstruction enables rapid descriptor-based optimization for correspondence and planning. Our framework significantly reduces both computational cost and supervision overhead, facilitating scalable and adaptable imitation learning while showing comparable performance to baseline.

1. Introduction

Robots that can generalize manipulation behaviors from a single demonstration have long been a central goal in Embodied AI. As the complexity of real-world environments increases, manually designing policies or shaping reward functions becomes increasingly infeasible. Environments with novel object instances, varied poses, and visual clutter present significant challenges that are difficult to anticipate in closed-form rule systems. In such settings, *Imitation Learning* (IL) offers a more scalable alternative: rather than engineering reward signals or heuristics, the robot observes an expert demonstration and directly replicates the task (Duan et al., 2017; Oh et al., 2018; Ross & Bagnell, 2010; Jung et al., 2024). However, one of the fundamental limitations of IL is its reliance on a large number of high-quality demonstrations, which are expensive and time-consuming to collect.

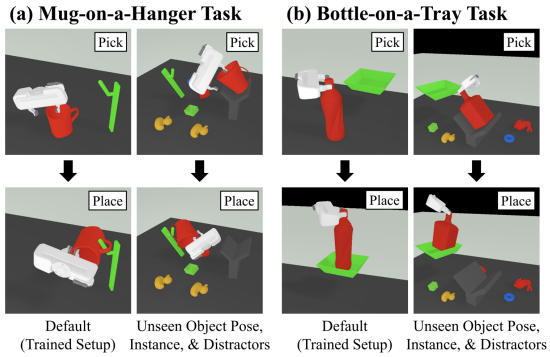

Figure 1. **Objective of the project**. We aim to achieve one-shot imitation learning for unseen object poses, unseeen similar objects, and with distractors. Figure borrowed from (Ryu et al., 2024).

To address this challenge, recent work has explored *SE(3)-equivariant descriptor fields* (Simeonov et al., 2022; Ryu et al., 2022; 2024), which learn features that are consistent across object rotations and translations. These methods have shown strong generalization to novel poses and configurations, allowing robots to extrapolate from fewer demonstrations. However, their practical utility remains limited: they typically require task-specific neural network training, offline optimization (e.g., diffusion-based sampling), and significant computational resources to achieve high accuracy. Moreover, despite their improved sample efficiency, they often fall short of true *one-shot* generalization and remain sensitive to training data distributions.

In this work, we propose an alternative approach that sidesteps the need for large-scale task-specific training while still achieving robust generalization across unseen objects and SE(3) configurations. Our key insight is that recent advances in *3D Gaussian Splatting* (3D-GS) (Kerbl et al., 2023) provide a powerful, compact representation of a scene that inherently supports SE(3)-equivariant reasoning. By reconstructing the scene from a single demonstration of multi-view images and representing it as a set of Gaussian primitives, we form a rich, pose-aware descriptor field that

[1]Seoul National University [2]2020-17482 steve0221@gmail.com [3]2025-28058 johnha97@gmail.com [4]2023-24599 ininin0516@gmail.com. Correspondence to: Sanghyun Hahn <steve0221@snu.ac.kr>.

encodes geometry, appearance, and object identity—without the need for explicit feature learning or neural descriptors.

We use this representation to enable one-shot imitation of pick-and-place tasks. Given a single expert demonstration consisting of multi-view RGB images, full robot trajectories, and gripper states, our method reconstructs the scene as a set of Gaussians, establishes dense correspondences in SE(3) space, and adapts the trajectory to novel configurations of pick and place objects. Crucially, this pipeline avoids any network pretraining or task-specific fine-tuning. Instead, our method relies on fast optimization over the Gaussian field to infer new action parameters.

Despite its lightweight design, our method—termed *Gaussian-EFs*—achieves competitive performance compared to prior work that uses diffusion models and pretrained neural fields. On a challenging mug placement benchmark, our method reaches a 92% overall success rate using only a single demonstration, closely matching the 95% success rate of a diffusion-based baseline that requires much heavier training and supervision. These results highlight the effectiveness of 3D-GS as a plug-in replacement for neural descriptor modules in robotic learning systems.

Our contributions can be summarized as follows:

- We propose a lightweight one-shot imitation learning framework that substitutes the task-specific SE(3)-equivariant neural descriptor with a 3D Gaussian Splatting-based representation.
- Our method enables generalization to novel object instances and SE(3) poses from a single multi-view demonstration, without requiring pretraining or neural network inference at test time.
- We empirically show that our approach achieves comparable or superior performance to existing descriptor-field-based methods, while being significantly more efficient and easier to deploy.

By reducing reliance on large datasets, pretrained models, and computational overhead, our method brings imitation learning closer to real-world applicability, enabling low-effort deployment in dynamic or user-specific environments.

2. Preliminaries

2.1. Gaussian Splatting

Gaussian Splatting (Kerbl et al., 2023) is a well-developed 3D representation for scene reconstruction, showing advantages on quality and rendering speed compared to previous volume-rendering approaches (Mildenhall et al., 2021; Müller et al., 2022). 3D-GS is similar to colorized point clouds, with a volume shaped like a 3D Gaussian.

Specifically, 3D Gaussians $\{G_i\}_{i=1}^{M}$ where the i-th Gaussian is represented by $G_i = \{\boldsymbol{\mu}_i, \boldsymbol{q}_i, \boldsymbol{s}_i, \boldsymbol{c}_i, \alpha_i\}$, where $\boldsymbol{\mu} \in \mathbb{R}^3$ is the Gaussian center, $\boldsymbol{s} \in \mathbb{R}^3$ and $\boldsymbol{q} \in SO(3)$ are respectively the scaling factor and the rotation represented in quaternion to define the covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$. $\boldsymbol{c_i} \in \mathbb{R}^3$ is the color, and $o_i \in \mathbb{R}$ is opacity. For a 3D query location $\boldsymbol{x} \in \mathbb{R}^3$, its Gaussian weight $\mathbf{g}(\boldsymbol{x})$ is represented as:

$$\mathbf{g}(\boldsymbol{x}) = e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{1}$$

where the symmetric 3D covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{3\times3}$ is represented by

$$\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{R}^T. \tag{2}$$

$\boldsymbol{R} = \text{quat2rot}(\boldsymbol{q})$ is a rotation matrix converted from $\boldsymbol{q}$, and $\boldsymbol{S} = \text{diag}(\boldsymbol{s})$ is a diagonal matrix from scaling factor $\boldsymbol{s}$.

Similar to rendering point clouds using a point-rasterizer, 3D-GS rasterizes these 3D Gaussians $\{G_i\}_{i=1}^{M}$ by sorting them in depth order in camera space and projecting them to the image plane. If N number of Gaussians are projected on 2D location $\boldsymbol{p} \in \mathbb{R}^2$, the pixel color $C(\boldsymbol{p})$ is given by α-blended rendering as follows:

$$C(\boldsymbol{p}) = \sum_{i\in N} c_i\alpha_i \prod_{j=1}^{i-1}(1-\alpha_j), \tag{3}$$

$$\alpha_i = \mathbf{g}_i^{2D}(\boldsymbol{p}) \cdot o_i, \tag{4}$$

where $\mathbf{g}_i^{2D}$ is the weight after the 2D projection of 3D Gaussian $\mathbf{g}_i$ to the image plane, and we use the Jacobian of the affine approximation of the projective transformation, following previous approaches (Zwicker et al., 2001; Kerbl et al., 2023). As the output of 3D scene reconstruction, we obtain the parameters of 3D Gaussians $\mathcal{G} = \{G_i\}_{i=1}^{M}$ by optimizing them with reconstruction loss calculated from rendering Eq. (3).

2.2. $SE(3)$ Equivariance in Imitation Learning

$SE(3)$, the Special Euclidean group in 3D, is a Lie Group that represents all possible rigid transformations in the 3D space. An element of $SE(3)$ can be represented as a 4×4 matrix:

$$T = \begin{bmatrix} \mathbf{R} & t \\ 0 & 1 \end{bmatrix} \quad \begin{array}{rcl} \mathbf{R} \in \text{SO(3)} & : & 3\times3 \text{ rotation matrix} \\ t \in \mathbb{R}^3 & : & \text{translation vector} \end{array} \tag{5}$$

Equivariance denotes the property of a function whose output changes identically to the action applied on the input. The $SE(3)$ equivariance of a vector field $f : \mathbb{R}^3 \times \mathcal{X} \to \mathbb{R}^{2l+1}$ can be defined as:

$$D_l(\mathbf{R})f(\mathbf{x}|X) = f(T\mathbf{x}|T \circ X) \quad \forall T = (\mathbf{R}, \mathbf{v}) \in \text{SE(3)}, \; X \in \mathcal{X}, \; \mathbf{x} \in \mathbb{R}^3$$

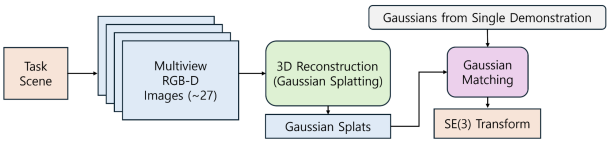

Figure 2. **Overview of the proposed pipeline.** The task scene is reconstructed using 3D Gaussian Splatting from multi-view RGB-D images. Gaussian Matching aligns the reconstructed scene to the demonstration Gaussians and estimates SE(3) transforms, which are used to adapt and execute the pick-and-place trajectory.

where $D_l(\mathbf{R})$ is the degree l *Wigner D-matrix* of the representation $D(\mathbf{R})$, and $\mathcal{X}$ is a set with action $T \circ \mathbf{x} = R\mathbf{x} + \mathbf{v}$. In the scope of robotic manipulation, x is an arbitrary 3D point with the vector field f conditioned on the object point cloud X. This property enables the robot to expand its prior knowledge to unseen poses, reducing the number of demonstrations across multiple object poses.

3. Methods

Our goal is to enable a robotic manipulator to generalize a pick-and-place task to novel object instances and poses using only a single demonstration. To achieve this, we propose a lightweight and modular pipeline that combines 3D Gaussian Splatting with Gaussian-based SE(3) alignment and trajectory adaptation. The pipeline consists of three main stages: (1) *Segmented Gaussian Splatting*, (2) *Gaussian Matching*, and (3) *Trajectory Generation*. An overview of the full pipeline is illustrated in Fig. 2.

The system takes as input a set of multi-view RGB-D images captured from the task scene, along with precomputed Gaussian representations from a single expert demonstration. It reconstructs the task scene using Gaussian Splatting, performs feature-based Gaussian matching to align objects between the demonstration and the current task, and finally uses the estimated SE(3) transforms to adapt and execute the manipulation trajectory.

3.1. Segmented Gaussian Splatting

In the first stage, we construct compact 3D representations of the pick and place objects from the demonstration using 3D Gaussian Splatting (Kerbl et al., 2023). This process begins with multi-view RGB images captured from a wrist-mounted camera during the expert demonstration.

To isolate the relevant objects, we use Grounded Segment Anything Model (Grounded SAM) (Kirillov et al., 2023; Li et al., 2023) to segment the pick and place objects across all views. These object masks are then used to filter the image content, allowing us to reconstruct only the segmented

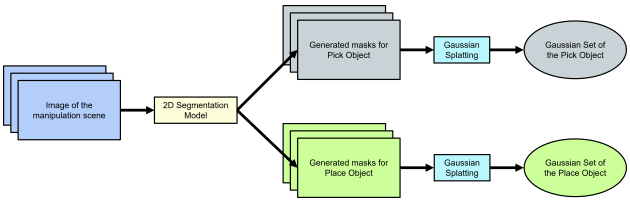

Figure 3. **Pipline of Segmented Gaussian Splatting**. We plan to generate the Gaussian Set of the Pick and Place Object by exploiting a 2D Segmentation Model and Gaussian Splatting.

regions through vanilla Gaussian Splatting. This results in two sets of Gaussians, one for each segmented object:

$$G_{\text{pick}} = \{g_i\}_{i=1}^{N_{\text{pick}}}, \quad G_{\text{place}} = \{g_j\}_{j=1}^{N_{\text{place}}}$$

Each Gaussian g is parameterized by its centroid position $\boldsymbol{x} \in \mathbb{R}^3$, opacity $\alpha \in \mathbb{R}$, rotation (expressed as axis-angle vector) $\boldsymbol{r} \in \mathbb{R}^3$, and scale $\boldsymbol{s} \in \mathbb{R}^3$. These features implicitly encode object shape and structure, acting analogously to SE(3)-equivariant descriptor fields from prior work (Ryu et al., 2022; 2024).

3.2. Gaussian Matching via SE(3) Alignment

The next step is to compute the SE(3) transformation that aligns the demonstration scene to the current task scene. Specifically, we compute the transformation that maps the demonstration pick object to the task pick object, and likewise for the place object. This is accomplished by matching their corresponding Gaussian sets.

We concatenate each Gaussian's parameters into a single feature vector:

$$F(g) = \text{concat}(\boldsymbol{\mu}, \alpha, \boldsymbol{q}, \boldsymbol{s}) \in \mathbb{R}^{10}$$

Matching is performed via a feature-augmented version of the Iterative Closest Point (ICP) algorithm (Besl & McKay, 1992), in which correspondences are determined by nearest neighbors in feature space and the SE(3) transformation is optimized to minimize bidirectional alignment error.

The optimization objective is as follows:

$$\min_{T \in SE(3)} \Big[\sum_{g \in G_{\text{demo}}} \| F(T \cdot g) - F\left(\mathcal{N}(T \cdot g, G_{\text{task}})\right) \|_2^2 + \sum_{g' \in G_{\text{task}}} \| F(g') - F\left(\mathcal{N}(g', T \cdot G_{\text{demo}})\right) \|_2^2 \Big] \quad (6)$$

Here, $\mathcal{N}(g, G)$ denotes the nearest Gaussian in set G to Gaussian g, under L_2 distance in feature space, and $T \cdot g$ applies the SE(3) transformation to the Gaussian's centroid and orientation.

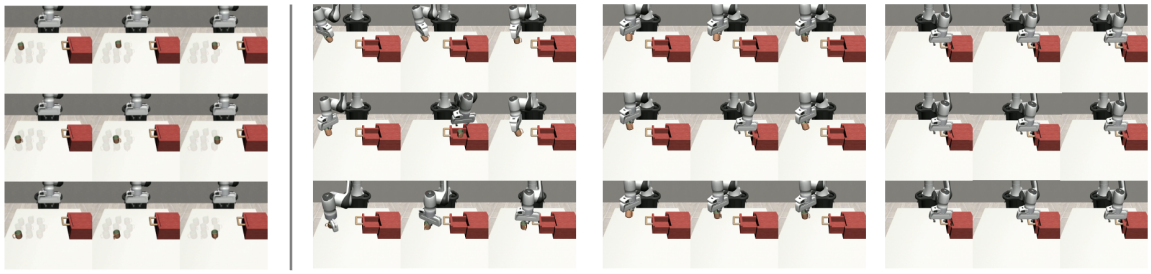

Figure 4. **Generalization to unseen SE(3) poses.** Our method successfully adapts the demonstrated pick-and-place trajectory to novel target object poses, including variations in position and orientation. Even under large SE(3) transformations, the robot accurately grasps and places the object using only a single demonstration.

Table 1. Mug Pick–Place Success Rate Results

Method	**Without**			**Mug**		
	Pretrain	**More than 1 Demo.**	**Neural Net.**	Pick	Place	Total
R-NDFs [1]	✗	✗	✗	0.83	**0.97**	0.81
SE(3)-DiffusionFields [2]	✓	✗	✗	0.75	*(n/a)*	*(n/a)*
Diffusion-EDFs [3]	✓	✗	✗	**0.99**	0.96	**0.95**
Gaussian-EFs (Ours)	✓	✓	✓	0.94	0.90	0.92

This alignment step is applied separately to both pick and place objects, resulting in two transforms:

$$T_{\text{pick}},\ T_{\text{place}} \in SE(3)$$

These define how to warp the demonstration frame to match the task scene configuration. We find that the inclusion of full Gaussian features—beyond position alone—significantly improves alignment accuracy, especially in cluttered environments or for objects with ambiguous geometry.

3.3. Trajectory Generation and Execution

Having estimated the object-level SE(3) transforms, we now generate the robot motion plan by adapting the demonstration trajectory to the task scene. The demonstration provides the full end-effector trajectory and gripper state sequence, including the grasp pose $T_{\text{grasp, demo}}$ for the demo pick object.

We apply the computed transforms to derive the grasp pose for the target object as:

$$T_{\text{grasp, task}} = T_{\text{pick}} \cdot T_{\text{grasp, demo}}$$

This yields the pose at which the robot should grasp the novel object.

After grasping, we compute a placement goal that aligns the object with the placement pose from the demonstration. Since the object has already been transformed by T_{pick}, we map it back into the demonstration frame via:

$$T_{\text{place\_goal}} = T_{\text{place}} \cdot T_{\text{pick}}^{-1} \cdot T_{\text{grasp, task}}$$

Finally, we use an off-the-shelf motion planner to generate a collision-free trajectory from the grasp pose to the placement goal. The manipulator then executes the same relative motion as shown in the demonstration, now grounded in the geometry of the task scene.

3.4. Summary of Advantages

This three-stage pipeline allows the robot to perform a pick-and-place task on previously unseen objects with arbitrary SE(3) configurations, using only a single demonstration. In contrast to prior work based on neural descriptor fields or diffusion models, our method does not require task-specific pretraining or multiple demonstrations. It instead leverages the rich, differentiable structure of 3D Gaussians to establish dense object correspondences and adapt actions accordingly.

In our experiments (Table 1), this lightweight pipeline achieves success rates comparable to fully trained baselines while being significantly faster to set up and generalize. The modular design also permits easy integration with various motion planners and gripper controllers.

4. Results

4.1. Implementation Details

The Gaussian Matching algorithm first constructs 100 candidate rigid transforms in $SE(3)$ by setting each translation to the vector connecting the center of mass of the two objects, and selecting the rotation from a set of 100 uniform samples over $SO(3)$. Starting from every candidate, it minimizes the objective in Equation (6) with the Adam optimizer (Kingma & Ba, 2014). The transformation that attains the lowest final loss is reported as the algorithm's estimate of the true transformation matrix.

The trajectory generation process is performed by an Operational Space Controller implemented by PyBullet (Coumans & Bai, 2016–2021), allowing the robot to follow the demonstration trajectory transformed corresponding to the estimated $SE(3)$ transformations. All experiments were conducted on a custom simulation environment on robosuite ().

4.2. Quantitative Results

We evaluate our method, Gaussian-EFs, on a mug pick-and-place task and compare it with three existing approaches: R-NDFs, SE(3)-DiffusionFields, and Diffusion-EDFs. The results are summarized in Table 1.

Our method achieves a 92% overall success rate, outperforming R-NDFs and SE(3)-DiffusionFields, and approaching the performance of the fully trained Diffusion-EDFs (95%), while maintaining a significantly lighter pipeline. Notably, Gaussian-EFs is the only method that succeeds without requiring pretrained weights, multiple demonstrations, or heavy neural network components, making it highly efficient and deployable in practical settings. R-NDFs achieves moderate performance (81% total), but relies on hand-crafted descriptors and struggles with SE(3) generalization. SE(3)-DiffusionFields performs poorly in the mug setting (75% pick success, no placement results reported), and requires pretraining on shape-specific tasks. Diffusion-EDFs achieves the highest total success rate (95%) but at the cost of requiring pretrained networks and offline diffusion-based optimization. In contrast, Gaussian-EFs achieves comparable performance (94% pick, 90% place, 92% total) while maintaining one-shot generalization with minimal overhead.

These results demonstrate that Gaussian-EFs provides a strong balance between accuracy and efficiency, making it a promising direction for one-shot imitation learning with SE(3) equivariance.

4.3. Qualitative Results

Our method demonstrates strong generalization across both pose variation and object diversity. As shown in Fig. 4, the robot successfully performs the pick-and-place task even when the target mug appears in significantly different SE(3) configurations, including varying orientations and translations from the demonstration scene. The estimated SE(3) transformations reliably align the manipulated objects, allowing trajectory adaptation without the need for retraining.

Furthermore, our approach generalizes to novel mug instances that differ in geometry, scale, and color from those seen in the demonstration. Despite these changes, the Gaussian-based matching preserves semantic correspondences, such as handle alignment and graspable surfaces. These results highlight the flexibility of our descriptor-free pipeline, showing that the combination of segmented Gaussian splatting and lightweight optimization provides a robust solution for one-shot manipulation under real-world variability.

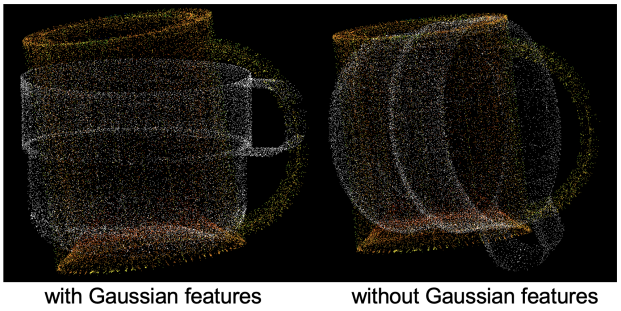

Figure 5. **Gaussian Matching**. Gaussian features allows the matching of cups with different pose and different shape in same category.

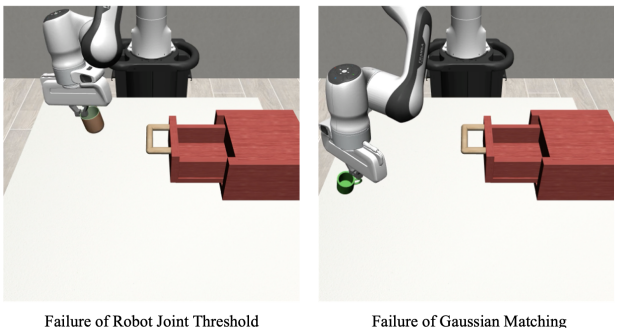

Figure 6. **Failure cases.** Our method occasionally fails due to two main reasons. *Left:* The robot is commanded to a pose that exceeds its joint limits (e.g., shoulder or elbow configurations beyond its physical range), resulting in unreachable targets or trajectory deviations. *Right:* Inaccurate alignment between demonstration and task objects—particularly with tall, narrow mugs—can cause the robot to miss the grasp point despite being close, leading to grasp failure.

5. Conclusion and Discussion

We presented a lightweight and training-free framework for one-shot imitation learning using 3D Gaussian Splatting. By leveraging SE(3)-equivariant Gaussian features extracted from a single demonstration, our method generalizes pick-and-place tasks to novel objects and poses without requiring additional training or multiple examples. Through extensive experiments, we demonstrate that our pipeline achieves strong performance across a variety of SE(3) configurations and unseen object instances.

Limitations and Future Work

The system occasionally fails due to physical constraints of the robotic hardware. For example, as shown in Fig. 6, the robot may be commanded to reach a configuration that violates joint limits, such as extreme shoulder or elbow positions. In such cases, the motion planner cannot execute the desired trajectory, resulting in task failure.

Also, failure can occur in cases where object alignment is not sufficiently accurate. For instance, with tall and narrow mugs, the Gaussian-based matching may lead to slight pose misalignments that cause the end-effector to miss the grasping region by a small margin—enough to compromise execution.

Another limitation lies in the reliance on high-quality 3D reconstruction. Our pipeline assumes access to a sufficient number of calibrated RGB-D views (e.g., $\sim$27 views) to produce clean Gaussian splats of the objects. In scenarios with limited camera access, occlusion, or poor depth quality, the resulting Gaussians may be too sparse or noisy for robust matching.

To address these issues, we plan to explore the use of single-view Gaussian reconstruction methods and learnable priors to reduce the dependence on multiple cameras. We also aim to investigate hybrid approaches that combine our descriptor-free pipeline with fast pose estimation networks to support faster execution and closed-loop control. Additionally, integrating robustness against robot kinematic constraints into the motion planning module is a promising direction for enabling safer and more reliable deployments in real-world settings.

References

Besl, P. J. and McKay, N. D. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pp. 586–606. Spie, 1992.

Coumans, E. and Bai, Y. Pybullet, a python module for physics simulation for games, robotics and machine learning. http://pybullet.org, 2016–2021.

Duan, Y., Andrychowicz, M., Stadie, B., Jonathan Ho, O., Schneider, J., Sutskever, I., Abbeel, P., and Zaremba, W. One-shot imitation learning. *Advances in neural information processing systems*, 30, 2017.

Jung, D., Lee, H., and Yoon, S. Sample-efficient adversarial imitation learning. *Journal of Machine Learning Research*, 25(31):1–32, 2024.

Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.

Li, Z., Zhou, Q., Zhang, X., Zhang, Y., Wang, Y., and Xie, W. Open-vocabulary object segmentation with diffusion models. 2023.

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Müller, T., Evans, A., Schied, C., and Keller, A. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4): 1–15, 2022.

Oh, J., Guo, Y., Singh, S., and Lee, H. Self-imitation learning. In *International conference on machine learning*, pp. 3878–3887. PMLR, 2018.

Ross, S. and Bagnell, D. Efficient reductions for imitation learning. In Teh, Y. W. and Titterington, M. (eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, volume 9 of *Proceedings of Machine Learning Research*, pp. 661–668, Chia Laguna Resort, Sardinia, Italy, 13–15 May 2010. PMLR.

Ryu, H., Lee, H.-i., Lee, J.-H., and Choi, J. Equivariant descriptor fields: Se (3)-equivariant energy-based models for end-to-end visual robotic manipulation learning. *arXiv preprint arXiv:2206.08321*, 2022.

Ryu, H., Kim, J., An, H., Chang, J., Seo, J., Kim, T., Kim, Y., Hwang, C., Choi, J., and Horowitz, R. 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*, pp. 18007–18018, 2024.

Simeonov, A., Du, Y., Tagliasacchi, A., Tenenbaum, J. B., Rodriguez, A., Agrawal, P., and Sitzmann, V. Neural descriptor fields: Se (3)-equivariant object representations for manipulation. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 6394–6400. IEEE, 2022.

Zwicker, M., Pfister, H., van Baar, J., and Gross, M. Ewa volume splatting. In *Proceedings Visualization, 2001. VIS '01.*, 2001.