IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

DOSE3: Diffusion-based Unified Out-of-distribution Detection on $\mathbb{SE}(3)$ Trajectories

Hongzhe Cheng[1], Tianyou Zheng[1], Ziyong Ma[1], Tianyi Zhang[2], Matthew Johnson-Roberson[3,1], Weiming Zhi[4,3]

***Abstract*—Out-of-Distribution (OOD) detection, the task of identifying when an input falls outside the distribution seen at training time, is critical for deploying safe and reliable systems. Traditional OOD methods require retraining models whenever the in-distribution has changed. Recent work introduces *unified* models for OOD detection, where metrics can be constructed from an unconditional diffusion model trained on an arbitrary dataset, and the inlier distribution can be changed without retraining the diffusion model. However, these unified approaches have been largely confined to Euclidean or latent space domains. In contrast, real-world robotics systems often perceive and act through sequences of 6 degrees-of-freedom poses in the Special Euclidean Group $\mathbb{SE}(3)$, taking into account both translations and orientation changes over time. In this work, we extend OOD detection to trajectories in *Special Euclidean Group in 3D* ($\mathbb{SE}(3)$) by presenting a *Diffusion-based Out-of-distribution detection on* $\mathbb{SE}(3)$ (DOSE3). DOSE3 constructs an OOD metric from the noise estimator model of a diffusion model over $\mathbb{SE}(3)$ to separate outlier samples from inlier distributions. We demonstrate DOSE3's strong performance on OOD detection frameworks through extensive validation on multiple real-world robotics and autonomous systems datasets, covering vehicle and robot manipulator motion trajectories.**

I. Introduction

OOD detection represents a fundamental machine learning challenge focused on identifying data samples that deviate from expected inlier distributions. This capability is particularly crucial in safety-critical applications within robotics and autonomous driving, where accurate identification of motion trajectory samples which differ from observed training samples can prevent system failures. Recent advances in OOD detection have explored various unsupervised approaches to learn inlier data representations. These include likelihood-based methods that employ different likelihood measures for OOD determination [1], [2], [3], [4], and reconstruction-based approaches that utilize pretrained generative models to assess sample similarity [5], [6], [7]. However, these methods typically require dataset-specific training, necessitating retraining for different in-distribution (ID) and OOD datasets [8].

Manuscript received: July 4, 2025; Revised October 7, 2025; Accepted November 15, 2025.

This paper was recommended for publication by Editor Aleksandra Faust upon evaluation of the Associate Editor and Reviewers' comments.

[1]Hongzhe Cheng, Tianyou Zheng, Ziyong Ma and Matthew Johnson-Roberson are with the Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA

[2]Tianyi Zhang is with Aurora, USA.

[3]Weiming Zhi and Matthew Johnson-Roberson are with the College of Connected Computing, Vanderbilt, TN, USA

[4]Weiming Zhi is with the School of Computer Science, The University of Sydney, Australia

Digital Object Identifier (DOI): see top of this page.

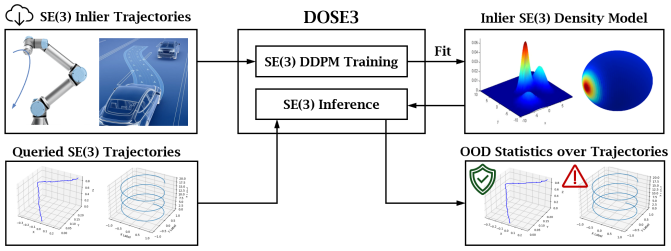


Fig. 1: Sequences of rigid poses are ubiquitous in robotics to describe objects moving in the real world. We propose **DOSE3**, a *unified* diffusion model over the $\mathbb{SE}(3)$ manifold to detect out-of-distribution pose sequences accurately that can adapt to new inlier datasets without retraining. This means the queried inlier trajectories can differ from the original trajectories used to train the diffusion model.

Recent research [9] seeks to address this limitation by training a unified model, where OOD detection can be conducted on a different inlier dataset without re-training. This means a model trained entirely *dataset A*, can also detect whether *dataset C* is OOD relative to dataset B. In this work, we present **DOSE3**, a unified framework that handles full pose sequences.

Existing trajectory OOD detection research primarily focuses on data spaces of latent spaces that are Euclidean, often overlooking explicit manifold space structures. Our work targets OOD detection for rigid body pose data, encompassing both position and orientation information. This type of data is fundamental to numerous applications in physics, engineering, and robotics that analyze object pose evolution over time [10]. We present theoretical insights and practical algorithms for detecting OOD data in rigid body pose sequences. Our framework, ***D**iffusion-based **O**ut-of-distribution detection on* $\mathbb{SE}(3)$ (**DOSE3**), introduces a novel unified generative approach for trajectory space OOD detection. We define a manifold-specific diffusion process for rigid transformations on $\mathbb{SE}(3)$ and develop a high-dimensional OOD statistic for out-of-distribution sample identification.

We validate our approach using established robotics and automation datasets, from the Oxford RobotCar [11], KITTI [12], IROS20 [13], as well as a newly-collected robot manipulator motion trajectory dataset. These datasets enable comprehensive evaluation across varying OOD similarity levels. Our key contributions include: (1) The **DOSE3** framework that *diffuses* over $\mathbb{SE}(3)$ sequences, incorporating manifold structures into OOD detection; (2) A novel OOD statistic derived from our $\mathbb{SE}(3)$ manifold diffusion estimator for sample degree measurement; (3) Comprehensive empirical validation demonstrating **DOSE3**'s effectiveness in distinguishing between in-distribution and OOD samples across diverse real-

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

world trajectory datasets. By connecting diffusion models with trajectory OOD detection, **DOSE3** advances the development of robust and scalable methods for autonomous systems and 3D trajectory analysis applications.

II. Related Work

OOD detection: OOD detection plays a crucial role in safety-critical applications such as autonomous driving. Existing methods can generally be categorized into likelihood-based and reconstruction-based approaches. Likelihood-based OOD detection methods involve training a model on ID data and deriving a likelihood statistic from test samples to serve as an OOD metric. Early work focused on learning discriminative representations to detect OOD samples and identify distributional shifts [5], [1] or used the reconstruction probability of VAEs [14], [15] for anomaly detection. More recent research has explored generative models due to their ability to model high-dimensional data and facilitate likelihood estimation [16]. However, studies have shown that generative models may assign higher likelihoods to OOD samples than to ID ones [17], [18]. To address this issue, various refinements have been proposed, including likelihood ratios [2], Watanabe-Akaike Information Criterion (WAIC) [3], improved noise contrastive priors [4], and Energy-based Model (EBM)s [19]. However, these enhancements remain ineffective in high-dimensional scenarios [7]. Another approach considers measuring how *typical* a test input is [20], but this method suffers from poor performance at the sample level. Normalizing flows [21] have also been investigated for OOD detection as they provide direct likelihood estimation, yet they still suffer from overconfidence issues [22]. Reconstruction-based OOD detection methods, on the other hand, aim to reconstruct input samples and compare them to their reconstructions to measure similarity. Additionally, curvature-based methods such as [23] estimate local decision-boundary curvature via low-rank Hessian sketches, enabling efficient OOD detection by capturing rapid output variations in the model's feature space.

Diffusion-based OOD Detection: Diffusion models (DMs) have achieved remarkable performance in generative tasks across various modalities, including images [24], [25], videos [26], and audio [27]. More recently, research has emphasized the robustness of DMs in sampling and their potential use in OOD detection. Utilizing the reconstruction mean squared error (MSE) of DDPMs as an OOD score has been shown to enhance image-space OOD detection [6], [7]. However, these models require retraining for different in-distribution datasets. A growing trend in machine learning research is the development of unified learning frameworks that generalize across various tasks [8]. In an effort to construct a unified DM for OOD detection, DiffPath [9] demonstrated that the rate of change and curvature of the forward diffusion trajectory can serve as effective OOD metrics, eliminating the need for retraining on different datasets.

The aforementioned OOD detection methods, whether likelihood-based or reconstruction-based, are primarily focused on images or Euclidean space data. In contrast, some research from the robotics community implicitly incorporates OOD detection under the framework of trajectory planning or optimization in 2D spaces [28], [29], [30]. However, these models struggle to generalize to complex real-world scenarios that involve three-dimensional interactions [31], [32]. Additionally, neural networks have been shown to suffer from performance reductions when directly using Euler angles or quaternions [33], as they cannot imbue continuity. While there is a growing interest in extending diffusion models to non-Euclidean spaces [34], [35], these methods are limited to generating individual samples on manifolds rather than modeling entire trajectories. To the best of our knowledge, **DOSE3** is the first approach to leverage manifold-based diffusion over trajectories, enabling a unified OOD detection framework.

III. Preliminaries

In this section, we first provide background on the architecture of diffusion models. We then discuss the recent advancements in constructing *Unified* OOD detection models using diffusion models. Finally, we introduce the *Special Euclidean Group in 3D*, $\mathbb{SE}(3)$, and elaborate on its geometric structure and related statistical foundations.

Denoising Diffusion Probabilistic Model (DDPM): Diffusion models have gained widespread attention in generative modeling due to their strong ability to synthesize high-fidelity data. These models employ a forward diffusion process, where data $\mathbf{x_0}$ is gradually corrupted by adding Gaussian noise over T timesteps, ultimately producing a noisy distribution $\mathbf{x_T}$ that approximates a standard normal distribution. The goal is to learn the reverse diffusion process, which systematically denoises $\mathbf{x_T}$ to recover the original data distribution. At the core of this reverse process is the ϵ-model, typically implemented as a neural network trained to predict the noise $\boldsymbol{\epsilon}$ added at each timestep t. The forward diffusion process, expressed in equation 1, illustrates how standard Gaussian noise is introduced to perturb the original sample $\mathbf{x}_0$. The backward process, given in equation 2, employs the estimator model $\boldsymbol{\epsilon}_\theta$, which estimates the true Gaussian noise $\boldsymbol{\epsilon}$ and enables data recovery by removing the noise.

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{1}$$

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z} \tag{2}$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and α_t, β_t, $\bar{\alpha}_t$ are predefined noise schedule parameters, and $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$. The theoretical foundation of diffusion models is grounded in variational inference, where the evidence lower bound (ELBO) in equation 3 is maximized to ensure that the learned reverse process closely approximates the true data distribution.

$$\begin{aligned}\mathcal{L}_{\text{ELBO}} = &\mathbb{E}_q[D_{\text{KL}}(q(\mathbf{x}_T|\mathbf{x}_0) \parallel p(\mathbf{x}_T)) \\ &+ \sum_{t=2}^{T} D_{\text{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) \parallel p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)) \\ &- \log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]\end{aligned} \tag{3}$$

By leveraging the ϵ-model within this framework, diffusion models effectively capture complex data manifolds, achieving state-of-the-art generative performance.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Unified Out-of-Distribution Detection: Traditional OOD detection methods, such as likelihood-based and reconstruction-based approaches, require retraining a new model for each specific inlier data distribution. This results in significant computational costs when switching between different OOD tasks and distributions. Recently, [9] introduced a new concept of Unified OOD detection, where a single unconditional diffusion model is trained, and distributional information can be obtained from inlier distributions that were unseen during training.

The theoretical foundation of this approach builds on the variance-preserving formulation used in DDPM. The difference between each denoising timestep is given in equation 4, rewritten as:

$$d\mathbf{x}_t = -\frac{1}{2}\beta_t \mathbf{x}_t \, dt + \sqrt{\beta_t} \, d\mathbf{w}_t, \quad \mathbf{x}_0 \sim p_0(\mathbf{x}) \tag{4}$$

$$\frac{d\mathbf{x}_t}{dt} = f(\mathbf{x}_t, t) + \frac{g(t)^2}{2\sigma_t^2}\epsilon_p(\mathbf{x}_t, t) \tag{5}$$

In equation 7, we denote ϕ_T and ψ_T as the marginals obtained by evolving two distinct distributions, ϕ_0 and ψ_0, using their respective probability flow ordinary differential equations (ODEs) from equation 5.

$$D_{\mathrm{KL}}(\phi_0 \parallel \psi_0) = \frac{1}{2}\int_0^T \mathbb{E}_{\mathbf{x}_t \sim \phi_t}[\frac{g(t)^2}{\sigma_t^2}\|\epsilon_\phi(\mathbf{x}_t, t) \tag{6}$$

$$- \epsilon_\psi(\mathbf{x}_t, t)\|^2]dt + D_{\mathrm{KL}}(\phi_T \parallel \psi_T) \tag{7}$$

However, the KL divergence remains dependent on the specific model estimators ϵ_ϕ and ϵ_ψ in equation 7. The key observation is that even when executing DDPM forward diffusion using an estimator ϵ_θ trained on a third distribution θ, the sample can still be successfully transformed into a standard Gaussian distribution. This insight motivates the use of ϵ_θ—metrics extracted from an arbitrary diffusion estimator—to perform OOD detection on an inlier distribution ϕ.

The Special Euclidean Group in 3D: The Special Euclidean Group in 3D, denoted as $\mathbb{SE}(3)$, represents the space of rigid body transformations, which consist of both rotations and translations, written as:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

where $R \in \mathbb{SO}(3)$ is a rotation matrix, and $t \in \mathbb{R}^3$ is a translation vector. The group $\mathbb{SO}(3)$ consists of all 3×3 real orthogonal matrices with determinant equal to one:

$$\mathbb{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \mid R^\top R = I, \ \det(R) = 1\} \tag{8}$$

where I is the 3×3 identity matrix. The group $\mathbb{SO}(3)$ represents all possible rotations about the origin in three-dimensional space. The Lie algebra associated with $\mathbb{SO}(3)$ is denoted as $\mathfrak{so}(3)$ and consists of all 3×3 skew-symmetric matrices. A general element $\Omega \in \mathfrak{so}(3)$ can be written as:

$$\Omega = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix}$$

where $\omega = [\omega_1, \omega_2, \omega_3]^\top$ is a vector in $\mathbb{R}^3$. The Lie algebra $\mathfrak{so}(3)$ serves as the tangent space to the manifold $\mathbb{SO}(3)$, providing a locally Euclidean structure that facilitates computations on $\mathbb{SO}(3)$. The exponential map, $\exp : \mathfrak{so}(3) \to \mathbb{SO}(3)$, maps an element from the Lie algebra to the Lie group, enabling the representation of rotations in matrix form. Given $\Omega \in \mathfrak{so}(3)$, the exponential map is:

$$\exp(\Omega) = I + \frac{\sin\theta}{\theta}\Omega + \frac{1-\cos\theta}{\theta^2}\Omega^2 \tag{9}$$

where $\theta = \|\omega\|$ is the rotation angle, and ω is the vector corresponding to Ω. Conversely, the logarithmic map, $\log : \mathbb{SO}(3) \to \mathfrak{so}(3)$, converts a rotation matrix into its corresponding Lie algebra representation. For any $R \in \mathbb{SO}(3)$ that is not the identity matrix, the logarithmic map is given by:

$$\log(R) = \frac{\theta}{2\sin\theta}(R - R^\top), \tag{10}$$

$$\theta = \cos^{-1}\left(\frac{\mathrm{trace}(R) - 1}{2}\right) \tag{11}$$

Here, $\mathfrak{so}(3)$, the *tangent space* of $\mathbb{SO}(3)$, lies within Euclidean space, allowing standard algebraic operations to be applied. This property is particularly useful for designing diffusion models over $\mathbb{SO}(3)$, as it enables efficient computations and parameterizations of rotations.

IV. Method

A. Overview

Here, we present ***D**iffusion-based **O**ut-of-distribution detection on* $\mathbb{SE}(3)$, **DOSE3**. **DOSE3** introduces a *unified* diffusion model for rigid pose trajectories, specifically designed to accommodate the $\mathbb{SE}(3)$ manifold structure. We first detail **DOSE3**'s model architecture for handling ordered sequences. We then introduce $\mathbb{SE}(3)$ *Denoising Diffusion Probabilistic Models* ($\mathbb{SE}(3)$ - DDPM), outlining their training and inference algorithms that incorporate rigid pose structure into the diffusion model. Finally, we explain how to utilize the *diffusion estimator*, a function naturally emerging from $\mathbb{SE}(3)$ - DDPM, to develop an OOD detection statistic for evaluating test samples.

B. Architectural details of DOSE3

The **UNet** architecture, widely adopted in diffusion models for its effective encoder-decoder structure, enables high-fidelity data generation. Originally developed for biomedical image segmentation, UNet's symmetric design with skip connections preserves spatial information through its network layers. While the original UNet employs 2D convolution layers with max pooling and up convolution for dimensional adjustment, we modify this architecture for sequential data diffusion through the following enhancements: (1) Replace all convolution layers with 1D convolutions to process temporal structures in motion trajectories; (2) Introduce attention layers before each up- or down-sampling operation to better capture long-range dependencies in trajectory data, extending beyond the local computations of convolution layers; (3) Implement Residual connections [36] around attention layers, similar to Transformer architecture, to enhance learning capabilities for our complex data format and task. The architecture for each up/down UNet layer in our model is illustrated in fig. 3.

C. Training and Inference of DOSE3

The incorporation of rotation matrices from $\mathbb{SE}(3)$ format introduces manifold space considerations that preclude direct

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

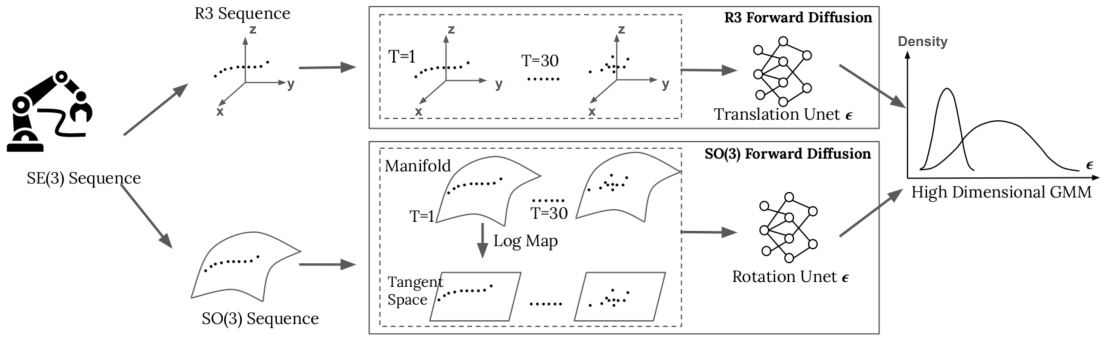


Fig. 2: System diagram of **DOSE3**. Pose trajectories are diffused with rotational components constrained to the $\mathbb{SO}(3)$ manifold, and the resulting diffusion estimators are used to derive OOD statistics through a fitted high-dimensional GMM.

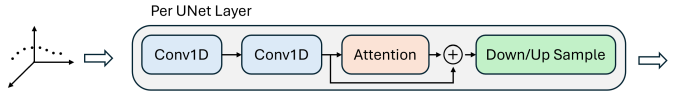


Fig. 3: SE(3) diffusion UNet layer. Trajectories are processed by the 1D convolution modules before being inputted into a transformer.

application of classical diffusion algorithms. We address three primary challenges: (1) The undefined nature of addition and scalar multiplication operations for rotation matrices; (2) The inability to guarantee valid rotation matrices when sampling 3×3 matrices from $\mathcal{N}(0, I)$; (3) The inadequacy of simple $L2$ norm differences for measuring distances/losses between rotation matrices. To overcome these challenges, we introduce the new $\mathbb{SE}(3)$ DDPM algorithm. While we apply standard Euclidean space diffusion to the translational components of $\mathbb{SE}(3)$, we develop specialized techniques for handling manifold diffusion over the $\mathbb{SO}(3)$ rotation space.

We redefine the operators $\in \mathbb{SO}(3)$ as follows. Essentially, we perform all operations after transforming the $\mathbb{SO}(3)$ data from manifold space into Euclidean tangent space by exponential and logarithmic map given by eq. (9) and eq. (11).

$$R_1 \oplus R_2 = R_1 \cdot R_2, \quad k \otimes R_1 = \exp(k \cdot \log(R_1)), \tag{12}$$

Where $k \in \mathbb{R}$, $R_1, R_2 \in \mathbb{SO}(3)$. We then change the noise sampling method from standard Gaussian distribution to the Isotropic Gaussian distribution on $\mathbb{SO}(3)$ ($\mathcal{IG}_{\mathbb{SO}(3)}$) distribution. Shown in equation 13, we first sample v from a standard Gaussian distribution, representing the tangent vector, and then use the exponential map operation to transform it to the $\mathbb{SO}(3)$ space.

$$\mathcal{IG}_{\mathbb{SO}(3)}(\mu, \sigma^2) = \mu \otimes v, \quad v \in \mathbb{R}^3 \sim \mathcal{N}(0, \sigma^2 I) \tag{13}$$

Combining all the metrics and operations defined above, we design the full forward and backward $\mathbb{SO}(3)$ DDPM equations, incorporating the operations on the manifold, as,

$$q(x_t|x_0) = \left(\sqrt{\bar{\alpha}_t} \otimes x_0\right) \oplus \left((1 - \bar{\alpha}_t) \otimes \epsilon\right), \tag{14}$$

$$\text{where } \epsilon \sim \mathcal{IG}_{\mathbb{SO}(3)}(0, I)$$

$$\widehat{x}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}} \otimes \left(x_t \oplus \left(-\sqrt{1 - \bar{\alpha}_t} \otimes \epsilon_\theta(x_t, t)\right)\right),$$

$$\mu_t = \left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} \otimes \widehat{x}_0\right) \oplus \left(\frac{\sqrt{\alpha_t(1 - \bar{\alpha}_{t-1})}}{1 - \bar{\alpha}_t} \otimes x_t\right),$$

$$x_{t-1} = \mu_t \oplus \left(\sqrt{\beta_t} \otimes \epsilon\right) \tag{15}$$

The rotational distance will be adapted as the loss for training in $\mathbb{SO}(3)$. The metric will reflect the average angle difference

Algorithm 1 OOD Detection

Require: DDPM Model, Inlier $\mathbb{SE}(3)$ Dataset I, Query $\mathbb{SE}(3)$ Trajectory q

```
stats ← []
for traj_0 in I do
    sum ← [0, 0, 0, 0, 0, 0]
    for t in [0, T-1] do
        traj_t ← DDPM_forward(traj_0, t)
        sum ← sum + 6D metrics over ϵ_θ(traj_t, t)
    end for
    stats.append(sum)
end for
distribution ← GMM.fit(stats)
q_t ← DDPM_forward(q, t)
metric_q ← 6D metrics over ϵ_θ(q_t, t))
likelihood ← distribution.eval(metric_q)
OOD ← likelihood < threshold
```

on each axis for two rotation matrices as,

$$L_{rot}(R_1, R_2) = \arccos\left(\frac{\text{trace}(R_1^\top R_2) - 1}{2}\right)^2 \tag{16}$$

The final $\mathbb{SE}(3)$ diffusion training takes in batches of trajectories in the format of ordered $\mathbb{SE}(3)$ sequences, maintaining the ordering of the trajectories.

D. OOD Detection

While likelihood-based OOD detection algorithms traditionally rely on generative model likelihood measures, the ELBO shown in equation 3 has proven inadequate for OOD tasks due to its tendency to overestimate OOD sample likelihood [37]. Recent research demonstrates that the diffusion estimator ϵ_θ and its derivatives effectively capture data distribution characteristics and can be obtained from a unified diffusion model without retraining. As shown in Equation 7, the norm of noise estimator ϵ correlates with the divergence between different data distributions [9]. Based on this insight, we define the following OOD statistics group for a diffusion model with noise estimator ϵ_θ, where the operator $\langle x \rangle_p = \frac{1}{N}\sum_{i=0}^{N} x_i^p$. Then, the metric is given as,

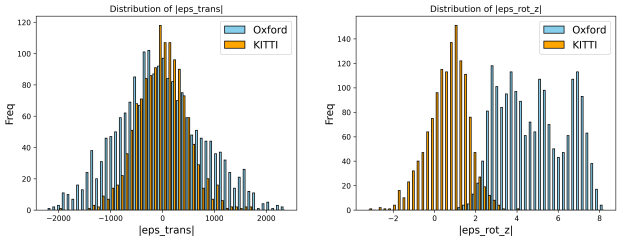
$$MetricGroup(\epsilon_\theta) =$$

$$\left[\sum_t \langle \epsilon_\theta(x_t, t)\rangle_1, \sum_t \langle \epsilon_\theta(x_t, t)\rangle_2, \sum_t \langle \epsilon_\theta(x_t, t)\rangle_3 \right.$$

$$\left. \sum_t \langle \partial\epsilon_\theta(x_t, t)\rangle_1, \sum_t \langle \partial\epsilon_\theta(x_t, t)\rangle_2, \sum_t \langle \partial\epsilon_\theta(x_t, t)\rangle_3 \right], \tag{17}$$

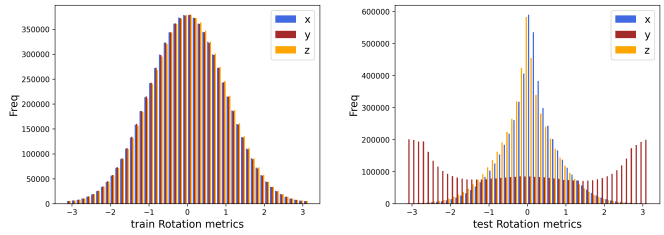
where $MetricGroup(\epsilon_\theta) \in \mathbb{R}^6$. For each sample x_0, we apply the DDPM forward process to obtain the perturbed sample x_t, then compute the metric group to derive final

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.



(a) Translation Metric (b) Rotation Z axis metric

Fig. 4: ϵ distribution retrieved by OOD testing on a model trained from KITTI dataset against the Oxford RoboCar dataset. We observe that the distribution difference is evident on the Z-axis rotation.



(a) Oxford (b) KITTI

Fig. 5: Distribution of elements $\epsilon(x_t, t)$ in rotation tangent space for each dimension when running Oxford Robot Car and KITTI dataset on model trained on Oxford Robot Car

statistics. Given that our $\mathbb{SE}(3)$ diffusion model comprises separate sub-diffusions for $\mathbb{R}^3$ and $\mathbb{SO}(3)$, and rotation metric distributions can vary along x, y, and z directions (as illustrated in figure 5), we establish distinct metric sets for each rotational dimension. This results in 4 sets of statistics per sample: three for individual rotational axes and one for translation, yielding a total of 24 metrics per sample. To process these metrics from the inlier data distribution, we estimate the density over the 24-dimensional joint vector of metric groups. We employ straightforward density estimators such as Gaussian Mixture Models or Kernel Density Estimators, as each metric empirically exhibits approximately Gaussian behavior. During testing, we collect identical metrics for each test sample and infer likelihood from the inlier density estimator, identifying lower-likelihood samples as out of distribution. We establish the OOD sample threshold at the bottom 5 percentile of inlier distribution likelihoods. The complete OOD model fitting and inference procedure is detailed in Algorithm 1.

V. Experiments

A. Experiment Setup

We evaluate **DOSE3**'s validity, performance, and comprehensiveness, we conduct OOD testing using the following tracking datasets $\mathbb{SE}(3)$ datasets:

Oxford RobotCar [11]: This autonomous driving dataset encompasses over 1000 km of driving data from central Oxford, UK. It features multiple sensor modalities, including high-resolution stereo and monocular cameras, 2D and 3D LiDAR scans, and GPS/INS ground truth localization. Our experiments utilize the 3D LiDAR scans and ground-truth poses stored in $\mathbb{SE}(3)$ format.

KITTI [12]: This comprehensive odometry dataset captures autonomous driving scenarios across urban, suburban, and rural environments. The dataset provides stereo and monocular camera imagery, 3D point clouds from a Velodyne LiDAR, and precise GPS/INS measurements. We utilize its pose data represented in $\mathbb{SE}(3)$

iros20-6d-pose-tracking [13]: This data set advances research on the estimation and tracking of 6D objects in dynamic environments. It is specifically designed to support the development and evaluation of algorithms for accurately determining and tracking six degrees of freedom (6D) poses in real-world scenarios.

For the implementation of our diffusion model, we standardize the input length for both $\mathbb{R}^3$ and $\mathbb{SE}(3)$ trajectory diffusion. Each trajectory in the datasets is segmented into fixed-length sub-paths of size 128 during experiments. To standardize the translation data, we first center each trajectory by setting its starting coordinate to the origin, then normalize by dividing by the maximum translation value. This process constrains the translation data to the range [-1, 1], ensuring the model learns trajectory geometry independent of scale. We evaluate **DOSE3** against strong OOD detection methods, including Joint Energy-based Model (JEM) [38] and Glow Model[21] with Likelihood Ratio [2]. We also compare against the classical unified diffusion-based OOD model over $\mathbb{R}^3$ introduced in [9]. These established baselines effectively handle high-dimensional inputs and are widely used for OOD detection.

B. Evaluation of ϵ_θ as an OOD Metric

We analyze the statistical distribution of ϵ_θ from inlier data to assess its effectiveness as an OOD detection metric. Specifically, we investigate how the ϵ_θ distribution of the $\mathbb{SO}(3)$ diffusion contributes to OOD sample identification. In fig. 4, we present a comparative analysis of ϵ distributions between Oxford RobotCar and KITTI datasets, using a model trained on KITTI. Our findings reveal that after translation data normalization, the translation ϵ distributions show substantial overlap across datasets, making them unsuitable as reliable OOD indicators. However, the rotation distribution, especially along the z-axis, demonstrates clear dataset separation. For the KITTI-trained model, we observe that KITTI's rotation distribution is centered at 0, aligning with standard Gaussian noise sampling characteristics. In contrast, the Oxford RobotCar dataset exhibits a notable rightward shift in its distribution, suggesting that matching KITTI samples from Oxford RobotCar input requires a non-Gaussian sampling distribution. The rotational elements ϵ around each axis, when running on a model trained on the Oxford Car dataset, are additionally visualized in fig. 5. We observe that for the KITTI dataset ϵ does not exhibit Gaussian distributions.

C. Quantitative Results

Table I presents the OOD detection performance across datasets using the AUROC metric. All evaluated models underwent unsupervised training exclusively on the KITTI dataset. Our $\mathbb{SE}(3)$ model demonstrates exceptional performance, achieving near-perfect Area under the Receiver Operating Characteristic curve (AUROC) scores across all ID and

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I: AUROC ↑ of OOD Detection. **Bold** denotes the best result. Our models are trained on the KITTI dataset, while each column indicates the inlier data vs outlier data, as inlier vs outlier (O: Oxford, K: KITTI, I:IROS20).

Method	O vs K	K vs O	O vs I	I vs O	K vs I	I vs K
JEM [38]	0.211	0.786	0.437	0.561	0.631	0.336
Glow-LR [21]	0.461	0.556	0.470	0.539	0.529	0.454
Unified $\mathbb{R}^3$-KITTI [9]	0.362	0.770	0.417	0.585	0.793	0.409
DOSE3-KITTI	**1.000**	**0.956**	**1.000**	**0.931**	**1.000**	**0.897**

TABLE II: AUROC ↑ of OOD Detection over different train dataset on sequence length of 128 and 30 diffusion steps (O: Oxford, K: KITTI, I:IROS20)

Method	O vs K	K vs O	O vs I	I vs O	K vs I	I vs K
Unified $\mathbb{R}^3$-Oxford [9]	0.897	0.327	0.890	0.378	0.464	0.532
DOSE3-Oxford	0.934	**0.999**	**1.000**	0.824	**1.000**	0.833
Unified $\mathbb{R}^3$-KITTI [9]	0.362	0.770	0.417	0.585	0.793	0.409
DOSE3-KITTI	**1.000**	0.956	**1.000**	**0.931**	**1.000**	**0.897**

OOD dataset combinations. In contrast, JEM, Glow-LR, and the unified $\mathbb{R}^3$ diffusion model show degraded performance when evaluating KITTI as an OOD dataset, or when trying to identify out-of-distribution samples when the inlier distribution is set as on the unseen Oxford Robot and IROS20 datasets.

Trajectory Dataset used for Training: **DOSE3** strives to develop a single unified model for effective OOD detection. This means we can train on an arbitrary dataset, and then use unseen datasets during training as the inlier dataset. We evaluate both $\mathbb{R}^3$ and $\mathbb{SE}(3)$-based diffusion models trained on two different datasets (Oxford Car and KITTI). Table II presents these results, highlighting two key findings: (1) $\mathbb{SE}(3)$ diffusion consistently demonstrates robust OOD detection capabilities across various training datasets; (2) $\mathbb{SE}(3)$ diffusion successfully performs OOD detection between two previously unseen datasets during training. We observe some performance degradation when training with the Oxford Robot Car dataset. This phenomenon primarily stems from the Oxford Robot Car dataset being slightly less varied than IROS or KITTI. Both IROS and KITTI datasets exhibit broader data distributions, encompassing more varied trajectory shapes. Consequently, when a unified diffusion model trained on the Oxford dataset attempts to distinguish between datasets with different levels of variability, the task becomes more challenging. Nevertheless, these results underscore the advantages of our unified diffusion approach to OOD detection. We do not restrict the model to be trained on a single inlier dataset each time we need to change the inlier distribution. That is, we can use a **DOSE3** model trained on the KITTI dataset to immediate conduct OOD with the Oxford Car dataset as the inlier, thus our method significantly reduces the overall model training time, as it avoids re-training.

Necessity of Rotational Diffusion Information: We compare diffusion models trained on translation-only data versus those trained on complete $\mathbb{SE}(3)$ data to demonstrate the critical role of rotational information. Table II reveals that OOD detection using only $\mathbb{R}^3$ data yields poor results, consistent with the overlapping statistical distributions shown in figure 4. In contrast, **DOSE3** achieves superior performance by incorporating rotational components into the diffusion model. This finding demonstrates that for complex trajectory analysis, orientation and rotation data provide richer discriminative features that vary significantly across different data distributions, thereby serving as robust indicators for OOD detection.

TABLE III: AUROC ↑ of OOD Detection over different sequence length for 30 diffusion steps with model trained on KITTI dataset. With inliers vs outliers labelled in columns (O: Oxford, K: KITTI, I:IROS20).

Seq Length	O vs K	K vs O	O vs I	I vs O	K vs I	I vs K
64	**1.000**	**0.999**	**1.000**	**0.986**	0.999	**0.976**
128	**1.000**	0.956	**1.000**	0.931	**1.000**	0.897
256	0.980	0.961	**1.000**	0.941	**1.000**	0.932
512	**1.000**	**0.999**	**1.000**	0.942	**1.000**	0.923

TABLE IV: AUROC of OOD Detection over different numbers of diffusion steps on sequence length of 128 with model trained on KITTI dataset, along with inference times. With inliers vs outliers labelled in columns (O: Oxford, K: KITTI, I:IROS20).

Diffusion Step	O vs K	K vs O	O vs I	I vs O	K vs I	I vs K	Time (s)
5	0.916	0.969	**1.000**	**0.955**	**1.000**	**0.945**	0.056
10	0.914	**0.975**	**1.000**	0.938	**1.000**	**0.945**	0.104
15	0.948	0.964	**1.000**	0.919	**1.000**	0.912	0.158
30	**1.000**	0.956	**1.000**	0.942	**1.000**	0.897	0.312

D. Ablations

Sequence length of the training data: Table III presents the OOD detection performance for varying trajectory sequence lengths during KITTI dataset pre-training. The results demonstrate that **DOSE3** maintains consistently excellent performance with near-perfect AUROC scores across all ID and OOD pairs, independent of sequence length. This robustness to sequence length variation highlights the model's stability and generalization capabilities.

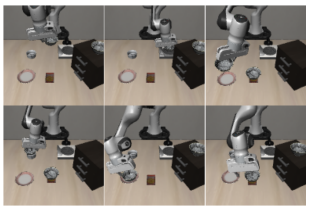
DDPM Forward Steps: Table IV illustrates the relationship between **DOSE3** performance and the number of DDPM steps. The results indicate minimal variation in AUROC scores across different step counts, demonstrating **DOSE3**'s resilience to changes in the number of DDPM steps.

Inference Time vs Diffusion Steps: Obtaining OOD statistics can be done efficiently, with the main computational burden on function calls to the diffusion model for each step. The average times in seconds for inference, on a standard desktop with an NVIDIA RTX 4090 GPU, with 5, 10, 15 and 30 diffusion steps are given in table IV. Inference times increase linearly with number of steps, but even with 30 steps, the inference time for a trajectory is only 0.3s.

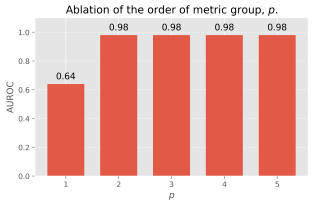
E. OOD Detection on Libero-Spatial Manipulation Tasks

We evaluate the robustness of **DOSE3** on robot manipulator motion trajectories from the dataset *Libero-Spatial* [39]. We seek to differentiate robot end-effector 6D pose trajectories from 3 different tasks, placing a black bowl from the Cookie box (C), Ramekin (R), and Stove (S) onto a plate, and the trajectories from the other tasks were used as training data. We report the AUROC of **DOSE3** in table V. We compare with the diffusion baseline introduced in [9], defined over the (1) position, (2) position and orientation represented by Euler angles, and (3) position and orientation represented by quaternions. We designate these methods as *Diffusion* $\mathbb{R}^3$, *Diffusion* $\mathbb{R}^6$, and *Diffusion Quaternion*. Additionally, we compare our results against a k-nearest neighbour baseline, where half of the inlier datapoints, each containing position and Euler angle orientation, are randomly selected and used as

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.



(a) Example 3d-person view images from the *Libero-Spatial* setups.



(b) We ablate over the order, p, of the OOD metric.

Fig. 6: We perform OOD experiments against baselines and ablate over metric order p, with manipulator trajectories from the *Libero-Spatial* dataset.

	C vs R	C vs S	R vs C	R vs S	S vs C	S vs R
Diffusion $\mathbb{R}^3$	0.60	0.90	0.48	0.90	0.81	0.92
Diffusion $\mathbb{R}^6$	0.79	**1.00**	0.78	**1.00**	0.93	**1.00**
Diffusion Quaternion	0.73	**1.00**	0.70	**1.00**	0.94	0.93
Distance Thresholding	0.74	0.93	0.67	**1.00**	0.89	0.94
DOSE3	**0.90**	**1.00**	**1.00**	**1.00**	**0.98**	**1.00**

TABLE V: AUROC ↑ of OOD detection over the *Libero-Spatial* dataset [39], with inliers vs outliers labelled in columns (C: Cookie Box, R: Ramekin, S: Stove).

the training set. Out-of-distribution samples are then identified as those lying beyond the 95th percentile of the nearest-neighbour distance distribution. Here, we select $k = 10$ and designate this baseline as *Distance Thresholding*.

The experimental results illustrate that both the orientation information and the $\mathbb{SE}(3)$ formulation in **DOSE3**. We observe that the inclusion of orientation information in either the form of Euler angles or Quaternions improves OOD performance, even when the additional dimensions are treated as Euclidean. We observe that when incorporating both position and orientation, the *Distance Thresholding* baseline surpasses the position-only *Diffusion* $\mathbb{R}^3$ model, yet overall performs slightly worse than its diffusion-based counterparts. This suggests that while explicit geometric information aids OOD detection, diffusion-based formulations offer additional performance gains. Unlike baselines that operate purely in Euclidean space, **DOSE3** leverages the structure of the Lie group to compute distances that are geometrically meaningful across both translational and rotational components, providing a more faithful separation between in-distribution and out-of-distribution trajectories.

Ablations on the Metric Group order: DOSE3 leverages a density estimator over the OOD statistics group defined in eq. (17), where we progressively add terms corresponding to increasing p values. We select the include terms of $\langle x \rangle_p$ up to $p = 3$, following the order of the metric in [9]. Here, we ablate over the value of p using the *Stove* and *Cookie box* trajectories as inlier and outlier data, respectively. We observe that at $p = 2$ and beyond, the detection performance is stable at 0.98. This indicates that selecting a reasonable p is sufficient, and the order itself is not very sensitive as long as it is not too small.

F. Training on KITTI and Generalizing to Manipulator Trajectories

We seek to evaluate how well **DOSE3** performs when trained on KITTI and generalized to a relatively small-sized robot manipulator trajectory dataset. We additionally collect

TABLE VI: AUROC ↑ of OOD Detection on robot arms data. Columns indicates inlier vs outlier.

Method	TOR vs CM	CM vs TOR	TOR vs TC	TC vs TOR	CM vs TC	TC vs CM
Unified-$\mathbb{R}^3$-KITTI [9]	0.63	0.75	0.86	**0.90**	0.82	0.83
DOSE3-KITTI	**0.80**	**0.95**	**1.0**	**0.90**	**0.89**	**1.0**

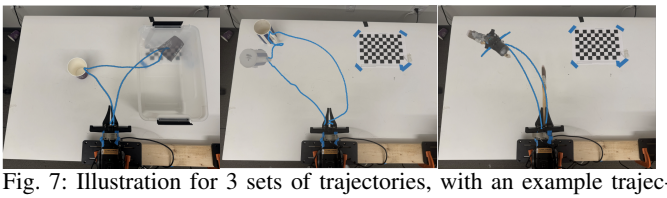


Fig. 7: Illustration for 3 sets of trajectories, with an example trajectory. From left to right: TOR, CM, TC.

trajectories on a robot manipulator, and seek for **DOSE3** to effectively distinguish between the different motion patterns of the robot arms. We collected 3 sets of 15 trajectories each that represent common tasks performed by robot arms, including:

1) **Set TOR: Table-to-Box Object Relocation** Picking up various objects from randomized positions on the table and placing them into a storage box.
2) **Set CM: Cup Manipulation** Using the gripper to pick up a cup, rotate it to simulate pouring and place it upright in a random position on the table.
3) **Set TC: Table Cleaning with a Brush** Gripping a cleaning brush and doing wiping motions on the surface of the table.

For each task set, one representative trajectory is selected and visualized in fig. 7 to illustrate the typical motion pattern associated with that set.

We tabulate the results of detecting the OOD trajectories in table VI. Here, despite **DOSE3** being trained on the KITTI dataset, the model can conduct OOD detection on trajectories from robot manipulation, a completely different domain. Notably, **DOSE3** was directly applied to the new datasets without any retraining or finetuning, preserving the unified framework premise and generalising effectively to new, unseen distributions of 6-DoF trajectories, including non-vehicular, manipulator motion. Here, we compare the AUROC metrics between **DOSE3** and the unified $\mathbb{R}^3$ diffusion model from [9], **DOSE3** model demonstrates near-perfect AUROC scores, consistently outperforming or matching the performance of the $\mathbb{R}^3$ diffusion model. These results indicate that incorporating rotational information significantly enhances the model's ability to detect out-of-distribution motions. This holds even when the inlier distribution is uniquely different from the training distribution of **DOSE3**.

VI. Conclusions

Out-of-Distribution (OOD) detection plays a vital role in robotics, particularly in safety-critical domains where systems must reliably interact with the physical world. In these applications, data typically consists of rigid object pose trajectories that capture both positional and rotational motion. While existing OOD detection approaches operate on assumed Euclidean latent spaces, we present **DOSE3**, a novel unified diffusion-based OOD detection framework specifically designed for $\mathbb{SE}(3)$ trajectory data, and can be widely applied to motion generation [40], [41] and motion prediction [42]. **DOSE3** innovates by directly incorporating manifold operations into

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

the diffusion model and introduces a novel architecture that extends DDPM to handle $\mathbb{SE}(3)$ manifold sequences. Through comprehensive empirical evaluation across diverse real-world safety-critical datasets, we demonstrate **DOSE3**'s robust performance and effectiveness.

REFERENCES

[1] J. Tack, S. Mo, J. Jeong, and J. Shin, "Csi: Novelty detection via contrastive learning on distributionally shifted instances," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 11839–11852, 2020.

[2] J. Ren, P. J. Liu, E. Fertig, J. Snoek, R. Poplin, M. Depristo, J. Dillon, and B. Lakshminarayanan, "Likelihood ratios for out-of-distribution detection," in *Advances in Neural Information Processing Systems* (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), vol. 32, 2019.

[3] H. Choi and E. Jang, "Generative ensembles for robust anomaly detection," 2019.

[4] X. Ran, M. Xu, L. Mei, Q. Xu, and Q. Liu, "Detecting out-of-distribution samples via variational auto-encoder with reliable uncertainty estimation," *Neural Networks*, vol. 145, pp. 199–208, 2022.

[5] T. Denouden, R. Salay, K. Czarnecki, V. Abdelzad, B. Phan, and S. Vernekar, "Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance," 2018.

[6] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 650–656, June 2022.

[7] M. S. Graham, W. H. Pinaya, P.-D. Tudosiu, P. Nachev, S. Ourselin, and J. Cardoso, "Denoising diffusion models for out-of-distribution detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2947–2956, June 2023.

[8] Z. Xiao, Q. Yan, and Y. Amit, "Do we really need to learn representations from in-domain data for outlier detection?," 2021.

[9] A. Heng, A. H. Thiery, and H. Soh, "Out-of-distribution detection with a single unconditional diffusion model," in *Advances in Neural Information Processing Systems*, 2024.

[10] W. Zhi, H. Tang, T. Zhang, and M. Johnson-Roberson, "Unifying representation and calibration with 3d foundation models," *IEEE Robotics and Automation Letters*, 2024.

[11] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," *The International Journal of Robotics Research (IJRR)*, vol. 36, no. 1, pp. 3–15, 2017.

[12] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[13] B. Wen, C. Mitash, B. Ren, and K. E. Bekris, "se(3)-tracknet: Data-driven 6d pose tracking by calibrating image residuals in synthetic domains," *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2020.

[14] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[15] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2014.

[16] Z. Xiao, Q. Yan, and Y. Amit, "Likelihood regret: An out-of-distribution detection score for variational auto-encoder," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 20685–20696, 2020.

[17] E. Nalisnick, A. Matsukawa, Y. Teh, D. Gorur, and B. Lakshminarayanan, "Do deep generative models know what they don't know?," 2019.

[18] D. Hendrycks, M. Mazeika, and T. Dietterich, "Deep anomaly detection with outlier exposure," *Proceedings of the International Conference on Learning Representations*, 2019.

[19] W. Liu, X. Wang, J. Owens, and Y. Li, "Energy-based out-of-distribution detection," *Advances in Neural Information Processing Systems*, 2020.

[20] E. Nalisnick, A. Matsukawa, Y. W. Teh, and B. Lakshminarayanan, "Detecting out-of-distribution inputs to deep generative models using typicality," 2020.

[21] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), vol. 31, 2018.

[22] P. Kirichenko, P. Izmailov, and A. G. Wilson, "Why normalizing flows fail to detect out-of-distribution data," in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds.), vol. 33, pp. 20578–20589, 2020.

[23] A. Sharma, N. Azizan, and M. Pavone, "Sketching curvature for efficient out-of-distribution detection for deep neural networks," in *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2021.

[24] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *arXiv preprint arxiv:2006.11239*, 2020.

[25] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

[26] J. Ho, T. Salimans, A. Gritsenko, W. Chan, M. Norouzi, and D. J. Fleet, "Video diffusion models," *arXiv:2204.03458*, 2022.

[27] N. Chen, Y. Zhang, H. Zen, R. J. Weiss, M. Norouzi, and W. Chan, "Wavegrad: Estimating gradients for waveform generation," 2021.

[28] T. Lai, W. Zhi, T. Hermans, and F. Ramos, "Parallelised diffeomorphic sampling-based motion planning," in *Proceedings of the 5th Conference on Robot Learning*, 2022.

[29] J. Ziegler and C. Stiller, "Spatiotemporal state lattices for fast trajectory planning in dynamic on-road driving scenarios," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1879–1884, 2009.

[30] M. Werling, J. Ziegler, S. Kammel, and S. Thrun, "Optimal trajectory generation for dynamic street scenarios in a frenét frame," in *2010 IEEE International Conference on Robotics and Automation*, pp. 987–993, 2010.

[31] J. Wang, L. Xu, H. Fu, Z. Meng, C. Xu, Y. Cao, X. Lyu, and F. Gao, "Towards efficient trajectory generation for ground robots beyond 2d environment," 2023.

[32] V. Bharilya and N. Kumar, "Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions," *Veh. Commun.*, vol. 46, Apr. 2024.

[33] Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li, "On the continuity of rotation representations in neural networks," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[34] C.-W. Huang, M. Aghajohari, J. Bose, P. Panangaden, and A. Courville, "Riemannian diffusion models," in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, eds.), 2022.

[35] A. Leach, S. M. Schmon, M. T. Degiacomi, and C. G. Willcocks, "Denoising diffusion probabilistic models on SO(3) for rotational alignment," in *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[37] J. Serrà, D. Álvarez, V. Gómez, O. Slizovskaia, J. F. Núñez, and J. Luque, "Input complexity and out-of-distribution detection with likelihood-based generative models," in *International Conference on Learning Representations*, 2020.

[38] W. Grathwohl, K.-C. Wang, J.-H. Jacobsen, D. Duvenaud, M. Norouzi, and K. Swersky, "Your classifier is secretly an energy based model and you should treat it like one," in *International Conference on Learning Representations*, 2020.

[39] B. Liu, Y. Zhu, C. Gao, Y. Feng, qiang liu, Y. Zhu, and P. Stone, "LIBERO: Benchmarking knowledge transfer for lifelong robot learning," in *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[40] W. Zhi, I. Akinola, K. van Wyk, N. Ratliff, and F. Ramos, "Global and reactive motion generation with geometric fabric command sequences," in *IEEE International Conference on Robotics and Automation, ICRA*, 2023.

[41] W. Zhi, T. Zhang, and M. Johnson-Roberson, "Instructing robots by sketching: Learning from demonstration via probabilistic diagrammatic teaching," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.

[42] H. Wang, W. Zhi, G. Batista, and R. Chandra, "Pedestrian trajectory prediction using goal-driven and dynamics-based deep learning framework," *Expert Systems with Applications*, vol. 271, 2025.