DOSE3 : DIFFUSION-BASED OUT-OF-DISTRIBUTION DETECTION ON $\mathbb{SE}(3)$ TRAJECTORIES

Hongzhe Cheng*, **Tianyou Zheng***, **Tianyi Zhang, Matthew Johnson-Roberson, Weiming Zhi**
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{hongzhec, tianyou2, tianyiz4, mkj, wzhi}@andrew.cmu.edu

ABSTRACT

Out-Of-Distribution (OOD) detection, a fundamental machine learning task aimed at identifying abnormal samples, traditionally requires model retraining for different inlier distributions. While recent research demonstrates the applicability of diffusion models to OOD detection, existing approaches are limited to Euclidean or latent image spaces. Our work extends OOD detection to trajectories in *Special Euclidean Group in 3D* ($\mathbb{SE}(3)$), addressing a critical need in computer vision, robotics and engineering applications that process pose sequences of objects in $\mathbb{SE}(3)$. We present ***D**iffusion-based **O**ut-of-distribution detection on* $\mathbb{SE}(3)$ (**DOSE3**), a novel OOD framework that extends diffusion to a unified sample space of $\mathbb{SE}(3)$ pose sequences. We demonstrate **DOSE3**'s superior performance compared to state-of-the-art OOD detection frameworks through extensive validation on multiple benchmark datasets.

1 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 like robotics and autonomous driving, where accurate identification of anomalous motion trajectory (Zhi et al., 2024b) 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 (Tack et al., 2020; Ren et al., 2019; Choi & Jang, 2019; Ran et al., 2022), and reconstruction-based approaches that utilize pretrained generative models to assess sample similarity (Denouden et al., 2018; Wyatt et al., 2022; Graham et al., 2023). However, these methods typically require dataset-specific training, necessitating retraining for different in-distribution (ID) and OOD datasets (Heng et al., 2024). Recent research (Xiao et al., 2021) has addressed this limitation by exploring single discriminative models for OOD detection. Our work similarly aims to develop unified OOD approaches that eliminate retraining requirements.

Current trajectory OOD detection research primarily focuses on Latent Euclidean spaces, 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 (Zhi et al., 2024a; Zhang et al., 2024). 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, creating benchmarks from Oxford RobotCar (Maddern et al., 2017), KITTI (Geiger et al., 2012), and IROS20 (Wen et al., 2020). These datasets enable comprehensive evaluation across varying OOD similarity levels. Our key contributions include: (1) The **DOSE3** framework that *diffuses* over

*Equal Contribution

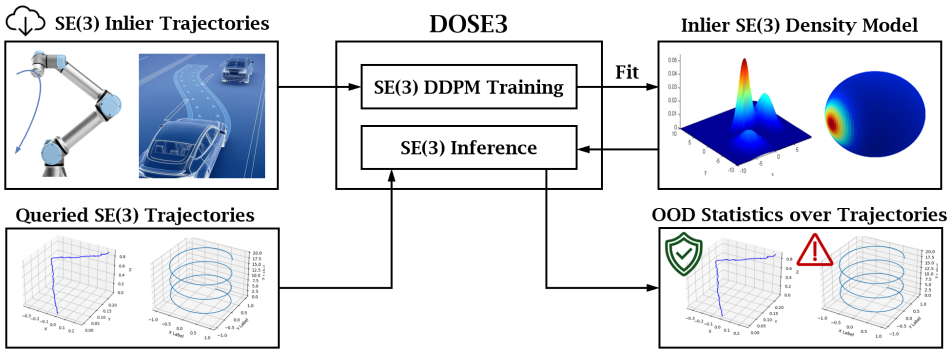


Figure 1: Sequences of rigid poses are abundant in displines that pertain to objects moving in the real world. We propose **DOSE3**, a unified diffusion model over the $\mathbb{SE}(3)$ manifold to detect out-of-distribution pose sequences accurately.

$\mathbb{SE}(3)$ 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-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.

2 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 (Denouden et al., 2018; Tack et al., 2020). More recent research has explored generative models due to their ability to model high-dimensional data and facilitate likelihood estimation (Xiao et al., 2020). However, studies have shown that generative models may assign higher likelihoods to OOD samples than to ID ones (Nalisnick et al., 2019; Hendrycks et al., 2019). To address this issue, various refinements have been proposed, including likelihood ratios (Ren et al., 2019), Watanabe-Akaike Information Criterion (WAIC) (Choi & Jang, 2019), improved noise contrastive priors (Ran et al., 2022), and Energy-based Model (EBM)s (Liu et al., 2020). However, these enhancements remain ineffective in high-dimensional scenarios (Graham et al., 2023). Another approach considers measuring how *typical* a test input is (Nalisnick et al., 2020), but this method suffers from poor performance at the sample level. Normalizing flows (Kingma & Dhariwal, 2018) have also been investigated for OOD detection as they provide direct likelihood estimation, yet they still suffer from overconfidence issues (Kirichenko et al., 2020). Reconstruction-based OOD detection methods, on the other hand, aim to reconstruct input samples and compare them to their reconstructions to measure similarity. Early work used the reconstruction probability of VAEs (An & Cho, 2015; Kingma & Welling, 2014) for anomaly detection. However, later studies found that OOD samples can exhibit similar or even lower reconstruction errors compared to ID samples, reducing the effectiveness of this approach (Denouden et al., 2018).

Diffusion-based OOD Detection: Diffusion models (DMs) have achieved remarkable performance in generative tasks across various modalities, including images (Ho et al., 2020; Song et al., 2021), videos (Ho et al., 2022), and audio (Chen et al., 2021). 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 (Wyatt et al., 2022; Graham et al., 2023). 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 (Xiao et al., 2021). In an effort to construct a unified DM for OOD detection, DiffPath (Heng et al., 2024) 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 (Lai et al., 2022; Ziegler & Stiller, 2009; Werling et al., 2010). However, these models struggle to generalize to complex real-world scenarios that involve three-dimensional interactions (Wang et al., 2023; Bharilya & Kumar, 2024). While there is a growing interest in extending diffusion models to non-Euclidean spaces (Huang et al., 2022; Leach et al., 2022), 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 entire trajectories, enabling a unified OOD detection framework.

3 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.

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

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z} \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{2}$$

where α_t, β_t, and $\bar{\alpha}_t$ 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.

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

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

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, Heng et al. (2024) 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 6, 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 equa-

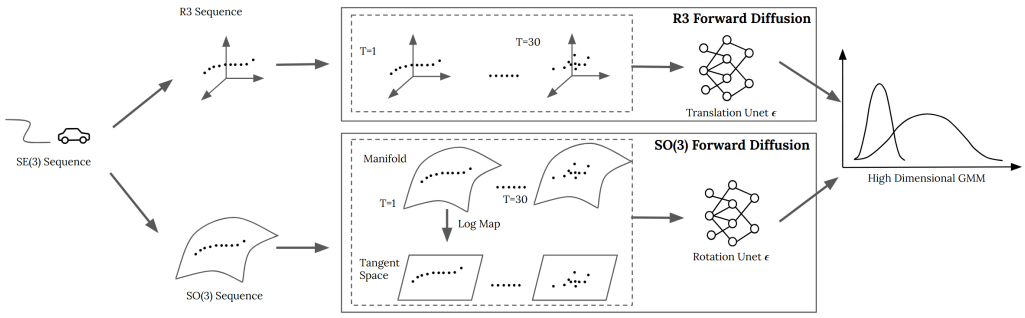


Figure 2: System Diagram of DOSE3 processing flow. Sequences of pose data are diffused, where diffusion over rotational components is constrained to the $\mathbb{SO}(3)$ manifold. The resulting diffusion estimators are used to construct an OOD statistic.

tion 5.

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

However, the KL divergence remains dependent on the specific model estimators ϵ_ϕ and ϵ_ψ in equation 6. 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. The transformation can be 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:

$$\mathbb{SO}(3) = \{R \in \mathbb{R}^{3\times 3} \mid R^\top R = I,\ \det(R) = 1\} \tag{7}$$

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:

$$\exp(\Omega) = I + \frac{\sin\theta}{\theta}\Omega + \frac{1-\cos\theta}{\theta^2}\Omega^2 \tag{8}$$

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:

$$\log(R) = \frac{\theta}{2\sin\theta}(R - R^\top), \qquad \theta = \cos^{-1}\left(\frac{\mathrm{trace}(R) - 1}{2}\right) \tag{9}$$

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.

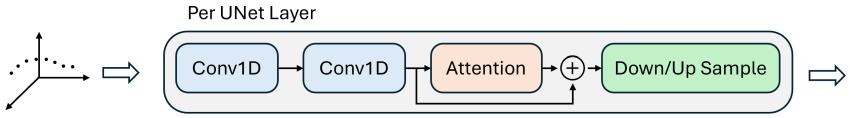


Figure 3: SE(3) diffusion UNet layer. Trajectories are processed by the 1D convolution modules.

4 METHOD

4.1 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.

4.2 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 (He et al., 2016) around attention layers, similar to Transformer architecture, to enhance learning capabilities for our complex data format and task. The resulting architecture for each up/down UNet layer in our model is illustrated in fig. 3.

4.3 TRAINING AND INFERENCE OF **DOSE3**

The incorporation of rotation matrices from $\mathbb{SE}(3)$ format introduces manifold space considerations that preclude direct 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. (8) and eq. (9).

$$R_1 \oplus R_2 = R_1 \cdot R_2, \quad k \otimes R_1 = \exp(k \cdot \log(R_1)), \quad k \in \mathbb{R},\ R_1, R_2 \in \mathbb{SO}(3) \tag{10}$$

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 11, we first sample v from standard Gaussiance distribution, representing the tangent vector, and then use the exponential map operation to transform it to the $\mathbb{SO}(3)$ space.

$$\mathcal{IG}_{\mathbb{SO}(3)}(\mu, \sigma^2) = \mu \otimes v, \quad v \in \mathbb{R}^3 \sim \mathcal{N}(0, \sigma^2 I) \tag{11}$$

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
```

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,

$$
\begin{aligned}
q(x_t \mid x_0) &= \left(\sqrt{\bar{\alpha}_t} \otimes x_0\right) \oplus \left((1-\bar{\alpha}_t) \otimes \epsilon\right), \quad \text{where } \epsilon \sim \mathcal{IG}_{\mathbb{SO}(3)}(0, I) \\
\widehat{x}_0 &= \frac{1}{\sqrt{\bar{\alpha}_t}} \otimes \left(x_t \oplus \left(-\sqrt{1-\bar{\alpha}_t} \otimes \epsilon_\theta(x_t, t)\right)\right), \\
\mu_t &= \left(\frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} \otimes \widehat{x}_0\right) \oplus \left(\frac{\sqrt{\alpha_t(1-\bar{\alpha}_{t-1})}}{1-\bar{\alpha}_t} \otimes x_t\right), \quad x_{t-1} = \mu_t \oplus \left(\sqrt{\beta_t} \otimes \epsilon\right)
\end{aligned}
\tag{12}
$$

The rotational distance will be adapted as the loss for training in $\mathbb{SO}(3)$. The metric will reflect the average angle difference on each axis for two rotation matrices. The equation can be written as

$$
\mathrm{L}_{rot}(R_1, R_2) = \arccos\left(\frac{\mathrm{trace}(R_1^\top R_2) - 1}{2}\right)^2
$$

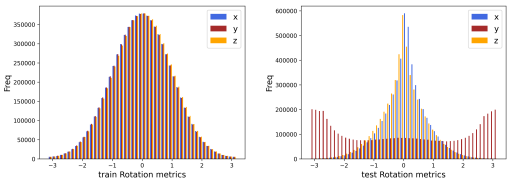
The final $\mathbb{SE}(3)$ diffusion training takes in batches of trajectories in the format of ordered $\mathbb{SE}(3)$ sequences, maintaining the ordering of the trajectories.

4.4 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 (Serrà et al., 2020). 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 6, the norm of noise estimator ϵ correlates with the divergence between different data distributions (Heng et al., 2024). Based on this insight, we define the following OOD statistics group for a diffusion model with noise estimator ϵ_θ, where the operator $\langle x \rangle_p = \frac{1}{N}\sum_{i=0}^{N} x_i^p$.

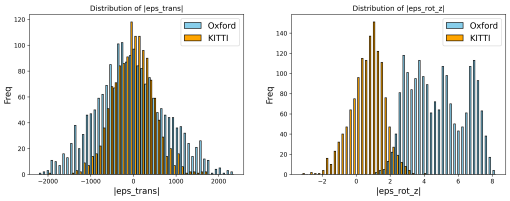
$$
\begin{aligned}
MetricGroup(\epsilon_\theta) = \Bigg[&\sum_t \langle \epsilon_\theta(x_t, t) \rangle_1, \sum_t \langle \epsilon_\theta(x_t, t) \rangle_2, \sum_t \langle \epsilon_\theta(x_t, t) \rangle_3 \\
&\sum_t \langle \partial\epsilon_\theta(x_t, t) \rangle_1, \sum_t \langle \partial\epsilon_\theta(x_t, t) \rangle_2, \sum_t \langle \partial\epsilon_\theta(x_t, t) \rangle_3 \Bigg],
\end{aligned}
\tag{13}
$$

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 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 4), 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



(a) Oxford (b) KITTI

Figure 4: 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



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

Figure 5: 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.

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.

5 EXPERIMENTS

5.1 EXPERIMENT SETUP

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

- **Oxford RobotCar** (Maddern et al., 2017): 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** (Geiger et al., 2012): 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** (Wen et al., 2020): This dataset advances research in 6D object pose estimation and tracking 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 our diffusion model implementation, 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 leading OOD detection methods, including Joint Energy-based Model (JEM)(Grathwohl et al., 2020) and Glow Model(Kingma & Dhariwal, 2018) with Likelihood Ratio (Ren et al., 2019). These established baselines effectively handle high-dimensional inputs and are widely used for OOD detection in image datasets.

5.2 QUANTITATIVE EVALUATION OF ϵ_θ DISTRIBUTION 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. 5, 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

Table 1: AUROC ↑ of OOD Detection. **Bold** denotes the best result (O: Oxford, K: KITTI, I:IROS20).

Method	O/K	K/O	O/I	I/O	K/I	I/K
JEM	0.211	0.786	0.437	0.561	0.631	0.336
Glow-LR	0.461	0.556	0.470	0.539	0.529	0.454
$\mathbb{R}^3$-KITTI	0.362	0.770	0.417	0.585	0.793	0.409
SE3-KITTI (No split)	0.845	0.952	**1.000**	0.398	**1.000**	0.234
SE3-KITTI	**1.000**	**0.956**	**1.000**	**0.931**	**1.000**	**0.897**

Table 2: 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/K	K/O	O/I	I/O	K/I	I/K
$\mathbb{R}^3$-Oxford	0.897	0.327	0.890	0.378	0.464	0.532
$\mathbb{SE}(3)$-Oxford	0.934	**0.999**	**1.000**	0.124	**1.000**	0.433
$\mathbb{R}^3$-KITTI	0.362	0.770	0.417	0.585	0.793	0.409
$\mathbb{SE}(3)$-KITTI	**1.000**	0.956	**1.000**	**0.931**	**1.000**	**0.897**

sampling characteristics. In contrast, the Oxford RobotCar dataset exhibits a notable rightward shift in its distribution, suggesting that reconstructing a KITTI sample from Oxford RobotCar input requires a non-Gaussian sampling distribution.

5.3 Quantitative Results

Table 1 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 AUROC scores across all ID and OOD dataset combinations. In contrast, JEM, Glow-LR, and the $\mathbb{R}^3$ model show degraded performance when evaluating KITTI as an OOD dataset, or when testing on the previously unseen Oxford Robot and IROS20 datasets. Additionally, we evaluate the impact of rotation metric splitting in $\mathbb{SE}(3)$ diffusion. The results indicate that separating the rotation metric space substantially enhances **DOSE3**'s robustness and unified feature representation, particularly improving AUROC scores in scenarios where the training dataset, KITTI, serves as the OOD data.

Trajectory Dataset used for Training: **DOSE3** strives to develop a single unified model for effective OOD detection. We evaluate both $\mathbb{R}^3$ and $\mathbb{SE}(3)$-based diffusion models trained on different datasets. Table 2 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 limitation primarily stems from the dataset's restricted trajectory diversity. Both IROS and KITTI datasets exhibit broader data distributions, encompassing more varied trajectory shapes. Consequently, when an Oxford-trained model attempts to distinguish between datasets with different levels of variability, the task becomes particularly challenging. Nevertheless, these results underscore the advantages of our unified diffusion approach to OOD detection. By requiring training on only a single dataset, our method significantly reduces the overall model training time.

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 2 reveals that OOD detection using only $\mathbb{R}^3$ data yields poor results, consistent with the overlapping statistical distributions shown in figure 5. In contrast, $\mathbb{SE}(3)$ diffusion achieves superior performance by incorporating rotational components. 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.

6 Conclusions

Out-of-Distribution (OOD) detection plays a vital role in machine learning, particularly in safety-critical domains like autonomous driving and robotics 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. **DOSE3** innovates by directly incorporating manifold operations into 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

Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special lecture on IE*, 2(1):1–18, 2015.

Vibha Bharilya and Neetesh Kumar. Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions. *Veh. Commun.*, 46(C), April 2024. ISSN 2214-2096. doi: 10.1016/j.vehcom.2024.100733. URL https://doi.org/10.1016/j.vehcom.2024.100733.

Nanxin Chen, Yu Zhang, Heiga Zen, Ron J. Weiss, Mohammad Norouzi, and William Chan. Wavegrad: Estimating gradients for waveform generation, 2021.

Hyunsun Choi and Eric Jang. Generative ensembles for robust anomaly detection, 2019. URL https://openreview.net/forum?id=B1e8CsRctX.

Taylor Denouden, Rick Salay, Krzysztof Czarnecki, Vahdat Abdelzad, Buu Phan, and Sachin Vernekar. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance, 2018. URL https://arxiv.org/abs/1812.02765.

Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

Mark S. Graham, Walter H.L. Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and Jorge 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.

Will Grathwohl, Kuan-Chieh Wang, Joern-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=Hkxzx0NtDB.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *Proceedings of the International Conference on Learning Representations*, 2019.

Alvin Heng, Alexandre H Thiery, and Harold Soh. Out-of-distribution detection with a single unconditional diffusion model. *arXiv preprint arXiv:2405.11881*, 2024.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arxiv:2006.11239*, 2020.

Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. Video diffusion models. *arXiv:2204.03458*, 2022.

Chin-Wei Huang, Milad Aghajohari, Joey Bose, Prakash Panangaden, and Aaron Courville. Riemannian diffusion models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=ecevn9kPm4.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. 2014.

Durk P Kingma and Prafulla Dhariwal. Glow: Generative flow with invertible 1x1 convolutions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/d139db6a236200b21cc7f752979132d0-Paper.pdf.

Polina Kirichenko, Pavel Izmailov, and Andrew G Wilson. Why normalizing flows fail to detect out-of-distribution data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 20578–20589. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/ecb9fe2fbb99c31f567e9823e884dbec-Paper.pdf.

Tin Lai, Weiming Zhi, Tucker Hermans, and Fabio Ramos. Parallelised diffeomorphic sampling-based motion planning. In *Proceedings of the 5th Conference on Robot Learning*, 2022.

Adam Leach, Sebastian M Schmon, Matteo T. Degiacomi, and Chris G. Willcocks. Denoising diffusion probabilistic models on SO(3) for rotational alignment. In *ICLR 2022 Workshop on Geometrical and Topological Representation Learning*, 2022. URL https://openreview.net/forum?id=BY88eBbkpe5.

Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 2020.

Will Maddern, Geoff Pascoe, Chris Linegar, and Paul Newman. 1 Year, 1000km: The Oxford RobotCar Dataset. *The International Journal of Robotics Research (IJRR)*, 36(1):3–15, 2017. doi: 10.1177/0278364916679498. URL http://dx.doi.org/10.1177/0278364916679498.

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

Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, and Balaji Lakshminarayanan. Detecting out-of-distribution inputs to deep generative models using typicality, 2020. URL https://openreview.net/forum?id=r1lnxTEYPS.

Xuming Ran, Mingkun Xu, Lingrui Mei, Qi Xu, and Quanying Liu. Detecting out-of-distribution samples via variational auto-encoder with reliable uncertainty estimation. *Neural Networks*, 145:199–208, 2022. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2021.10.020. URL https://www.sciencedirect.com/science/article/pii/S0893608021004111.

Jie Ren, Peter J. Liu, Emily Fertig, Jasper Snoek, Ryan Poplin, Mark Depristo, Joshua Dillon, and Balaji Lakshminarayanan. Likelihood ratios for out-of-distribution detection. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/1e79596878b2320cac26dd792a6c51c9-Paper.pdf.

Joan Serrà, David Álvarez, Vicenç Gómez, Olga Slizovskaia, José F. Núñez, and Jordi Luque. Input complexity and out-of-distribution detection with likelihood-based generative models. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=SyxIWpVYvr.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 11839–11852. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/8965f76632d7672e7d3cf29c87ecaa0c-Paper.pdf.

Jingping Wang, Long Xu, Haoran Fu, Zehui Meng, Chao Xu, Yanjun Cao, Ximin Lyu, and Fei Gao. Towards efficient trajectory generation for ground robots beyond 2d environment, 2023. URL https://arxiv.org/abs/2302.03323.

Bowen Wen, Chaitanya Mitash, Baozhang Ren, and Kostas 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. doi: 10.1109/iros45743.2020.9341314. URL http://dx.doi.org/10.1109/IROS45743.2020.9341314.

Moritz Werling, Julius Ziegler, Sören Kammel, and Sebastian 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. doi: 10.1109/ROBOT.2010.5509799.

Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris 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.

Zhisheng Xiao, Qing Yan, and Yali Amit. Likelihood regret: An out-of-distribution detection score for variational auto-encoder. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 20685–20696. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/eddea82ad2755b24c4e168c5fc2ebd40-Paper.pdf.

Zhisheng Xiao, Qing Yan, and Yali Amit. Do we really need to learn representations from in-domain data for outlier detection?, 2021. URL https://arxiv.org/abs/2105.09270.

Tianyi Zhang, Kaining Huang, Weiming Zhi, and Matthew Johnson-Roberson. Darkgs: Learning neural illumination and 3d gaussians relighting for robotic exploration in the dark. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024.

Weiming Zhi, Haozhan Tang, Tianyi Zhang, and Matthew Johnson-Roberson. Unifying representation and calibration with 3d foundation models. *IEEE Robotics and Automation Letters*, 2024a.

Weiming Zhi, Tianyi Zhang, and Matthew Johnson-Roberson. Instructing robots by sketching: Learning from demonstration via probabilistic diagrammatic teaching. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024b.

Julius Ziegler and Christoph 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. doi: 10.1109/IROS.2009.5354448.