A Mathematical Background

DUST-net uses a reparameterization of the space of rigid body transformations that allows distributions over an object's articulation model parameters to be defined naturally. Here, we briefly describe the mathematical foundation leveraged in the proposed distribution over articulation parameters.

A.1 Screw Transformations

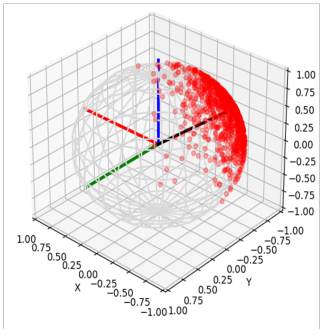
Chasles' theorem states that *"Any displacement of a body in space can be accomplished by means of a rotation of the body about a unique line in space accompanied by a translation of the body parallel to that line"* [32]. Such a line is called a screw axis, S. We represent this line using Plücker coordinates, given as $(\mathbf{l}, \mathbf{m})$ for a $l = \mathbf{p} + x\mathbf{l}$, with moment vector $\mathbf{m} = \mathbf{p} \times \mathbf{l}$, [32, 36]. The constraints $\|\mathbf{l}\| = 1$ and $\langle \mathbf{l}, \mathbf{m} \rangle = 0$ ensure that the degrees of freedom of the line in space are restricted to four. The rigid body displacement in $SE(3)$ as a screw transform is then defined as $\sigma = (\mathbf{l}, \mathbf{m}, \theta, d)$, where the linear displacement d and the rotation θ are connected through the pitch h of the screw axis, $d = h\theta$.

A.2 Stiefel manifold:

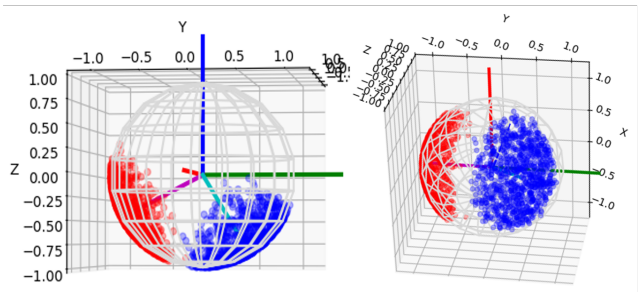
The *Stiefel manifold* $\mathrm{V}_{k,m}$ is the space whose points are sets of k orthonormal vectors in $\mathbb{R}^m$, called k-frames in $\mathbb{R}^m$ $(k \leq m)$ [19]. Points on the Stiefel manifold $\mathrm{V}_{k,m}$ are represented by the set of $m \times k$ matrices X such that $X^T X = I_k$, where I_k is the $k \times k$ identity matrix; thus $\mathrm{V}_{k,m} = \{X_{m,k}; X^T X = I_k\}$. Some special cases of the Stiefel manifold are the unit hypersphere $\mathrm{V}_{1,m}$ in $\mathbb{R}^m$ for $k = 1$, and the orthogonal group $O(m)$ for $m = k$.

A.3 von Mises-Fisher distribution

The von Mises-Fisher distribution (or Langevin distribution) is a unimodal probability distribution on the $(m-1)$ sphere in $\mathbb{R}^m$ (see Figure 6a). A random m-dimensional unit vector $\mathbf{x}$ is said to have the von Mises–Fisher distribution, if its probability distribution function is given by: $f_m(\mathbf{x}|\boldsymbol{\mu}, \kappa) = C_m(\kappa) \exp(\kappa \boldsymbol{\mu}^T \mathbf{x})$, where the concentration parameter $\kappa \geq 0$, the mean direction $\|\boldsymbol{\mu}\| = 1$ and the normalization constant $C_m(\kappa) = \dfrac{\kappa^{\frac{m}{2}-1}}{(2\pi)^{\frac{m}{2}} I_{\frac{m}{2}-1}(\kappa)}$ where I_ν denotes the modified Bessel function of the first kind at order ν [37]. For $m = 3$, the normalization constant reduces to $C_3(\kappa) = \dfrac{\kappa}{4\pi \sinh \kappa} = \dfrac{\kappa\, e^{-\kappa}}{2\pi(1 - e^{-2\kappa})}$.



(a) von Mises-Fisher distribution in $\mathbb{R}^3$. X, Y, Z axes are shown in red, blue and green colors, respectively. Black color represents the mean direction of distribution



(b) Matrix von Mises-Fisher distribution over $\mathrm{V}_{3,2}$, X, Y, Z axes are shown in red, blue and green colors, respectively. Magenta and cyan colors denote vectors corresponding to the first and second column of the matrix $M \in \mathrm{V}_{3,2}$ representing the mode of the distribution

A.4 Matrix von Mises-Fisher distribution

A random matrix X on $\mathrm{V}_{k,m}$ is said to have the matrix von Mises-Fisher distribution (or matrix Langevin distribution), if its density function is given by $\mathcal{F}(\mathbf{X}|m,\mathbf{F}) = \frac{1}{{}_0F_1(\frac{m}{2},\frac{1}{4}\mathbf{F}^T\mathbf{F})}\exp(\mathrm{Tr}(\mathbf{F}^T\mathbf{X}))$, where $\mathbf{F}$ is any $m \times k$ matrix and ${}_0F_1$ is a hypergeometric function with matrix argument [19] (see Figure 6b for an illustration). We can write the general (unique) singular value decomposition (SVD) of $\mathbf{F}$ as $\mathbf{F} = \Gamma\Lambda\Omega^T$, where $\Gamma \in \tilde{\mathrm{V}}_{k,m}$, $\Omega \in O(k)$, $\Lambda = diag(\lambda_1, ..., \lambda_k)$, $\lambda_1 \geq ... \geq \lambda_k \geq 0$, $\tilde{\mathrm{V}}_{k,m}$ denotes the set of matrices $\Gamma \in \mathrm{V}_{k,m}$ with the property that all the elements of the first row of the matrix Γ are positive, and $O(k)$ denoting the orthogonal group in k dimensions. It can be shown that ${}_0F_1(\frac{m}{2},\frac{1}{4}\mathbf{F}^T\mathbf{F}) = {}_0F_1(\frac{m}{2},\frac{1}{4}\Lambda^2)$. For more details, we refer to [19].

B Joint distribution over model parameters

A screw transform, represented as a tuple $\langle \mathsf{S}, \theta, d \rangle$, corresponds to a point on the manifold $\mathbb{S} \times SO(2) \times \mathbb{R}^+$, where $\mathbb{S} := \mathrm{V}_{2,3} \times \mathbb{R}^+$, $\mathrm{V}_{2,3}$ is the Stiefel manifold of $2-$frames in $\mathbb{R}^3$, $SO(2)$ denotes the circle group or the special orthogonal in two dimensions, and $\mathbb{R}^+$ denotes the set of positive real numbers. The unified representation proposed by Jain et al. [14] considers the motion of an articulated object as a sequence of screw transforms that share a common screw axis S. Hence, the extended tuple $\langle \mathsf{S}, \theta_{1:n-1}, d_{1:n-1} \rangle$, representing the articulation model for an object, corresponds to a point on the manifold $\mathbb{S} \times [SO(2)]^{n-1} \times [\mathbb{R}^+]^{n-1}$. We can define a joint distribution over the articulation model parameters by defining the probability density function for the distribution as the exponentiated distance of a point from the modal point of the distribution, and subsequently restricting the density function to the manifold [19]. However, calculating the normalization factor for this distribution is challenging. For example, a direct extension of the von Mises-Fisher distribution to define a distribution on $\mathrm{V}_{2,3} \times \mathbb{R}$ yields a density function with a normalizing factor that requires integrating a generalized hypergeometric function, which, to the best of our knowledge, is not computationally tractable to compute [38, 39]. Therefore, to define a distribution over the articulation model parameters that is tractable to learn, we make certain assumptions and propose an approximate joint distribution over the model parameters in this work.

Given a sequence of n depth images $\mathcal{I}_{1:n}$ of object part motion, the joint probability distribution over the articulation model parameters $p(\mathsf{S}, \theta_{1:n-1}, d_{1:n-1} \mid \mathcal{I}_{1:n})$ can be written as a product of a distribution over the screw axis parameters and a conditional distribution over the joint configuration parameters:

$$p(\mathsf{S}, \theta_{1:n-1}, d_{1:n-1} \mid \mathcal{I}_{1:n}) = p(\mathsf{S} \mid \mathcal{I}_{1:n})\; p(\theta_{1:n-1}, d_{1:n-1} \mid \mathsf{S}, \mathcal{I}_{1:n}) \tag{3}$$

We first approximate the distribution over the screw axis parameters S as a product of two marginal distributions: one over the orientation vector tuple $\langle \mathbf{l}, \hat{\mathbf{m}} \rangle \in \mathrm{V}_{2,3}$ and another over the moment vector magnitude $\|\mathbf{m}\| \in \mathbb{R}^+$,

$$p(\mathsf{S} \mid \mathcal{I}_{1:n}) \approx p(\langle \mathbf{l}, \hat{\mathbf{m}} \rangle \mid \mathcal{I}_{1:n})\; p(\|\mathbf{m}\| \mid \mathcal{I}_{1:n}) \tag{4}$$

This approximation is motivated by the fact that calculating statistics over manifolds can be computationally intractable in a general setting [19, 37, 40]. This approximation enables us to define the probability density function over the screw axis parameters using standard distributions over manifolds whose properties are well studied in the literature, such as the matrix von Mises-Fisher distributions over Stiefel manifolds [19, 37].

Calculating the conditional distribution over joint configurations, $p(\theta_{1:n-1}, d_{1:n-1} \mid \mathsf{S}, \mathcal{I}_{1:n})$, exactly would require us to evaluate hypergeometric functions over the complete manifold in which the screw transforms lie. Hypergeometric functions in the matrix argument result in an infinite series in terms of zonal polynomials, which becomes combinatorially expensive to calculate with the increasing number of terms [40]. To maintain the numerical tractability of the solution, we approximate the probability density function of the conditional distribution as a Dirac delta function centered at the expected value of the distribution over the screw axis parameters $\bar{\mathsf{S}}$:

$$\begin{aligned} p(\theta_{1:n-1}, d_{1:n-1} \mid \mathsf{S}, \mathcal{I}_{1:n}) &\approx\; \delta_{\bar{\mathsf{S}}}[p(\theta_{1:n-1}, d_{1:n-1} \mid \mathsf{S}, \mathcal{I}_{1:n})] \\ &=\; p(\theta_{1:n-1}, d_{1:n-1} \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \end{aligned} \tag{5}$$

where $\bar{\mathsf{S}} = \int_{\mathbb{S}} \mathsf{S}\, p(\mathsf{S} \mid \mathcal{I}_{1:n})$.

As we noted earlier, the unified parameterization of the articulation model parameters corresponds to a sequence of rigid body transforms (or screw transforms). Each of these rigid body transforms can be treated as an independent frame transformation between the object parts. Leveraging this fact, we approximate the conditional distribution over the joint configurations as a product of marginals over screw transforms at each time step:

$$p(\theta_{1:n-1}, d_{1:n-1} \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) = \prod_{i=1}^{n-1} p(\theta_i, d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \tag{6}$$

In this work, we approximate the conditional distribution over the joint configurations, $p(\theta_i, d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n})$, as a product of marginals over the rotation and displacement parameters to further simplify the parameterization of the joint distribution over articulation model parameters:

$$p(\theta_i, d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \approx p(\theta_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n})\, p(d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \tag{7}$$

While this approximate distribution cannot capture the correlations between joint configurations, it was found to be sufficiently expressive to enable DUST-Net to outperform the state-of-the-methods for articulation model estimation with a significant margin (see Section 5). In the future, DUST-Net may be extended to use multivariate distributions instead, which can capture the correlations between joint configurations as well.

Combining these together, in this work, we propose to approximate the joint distribution over articulation model parameters as:

$$\begin{aligned} p(\mathsf{S}, \theta_{1:n-1}, d_{1:n-1} \mid \mathcal{I}_{1:n}) &\approx p(\mathsf{S} \mid \mathcal{I}_{1:n}) \prod_{i=1}^{n-1} p(\theta_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \prod_{i=1}^{n-1} p(d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \\ &\approx p(\langle \mathbf{l}, \hat{\mathbf{m}} \rangle | \mathcal{I}_{1:n})\, p(\|\mathbf{m}\| \,|\mathcal{I}_{1:n}) \prod_{i=1}^{n-1} p(\theta_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \prod_{i=1}^{n-1} p(d_i \mid \bar{\mathsf{S}}, \mathcal{I}_{1:n}) \end{aligned} \tag{8}$$

where the exact parameterization of each of these probability distribution functions is discussed in section 4 of the main text.

C Hypergeometric function ${}_pF_q$

A general hypergeometric function ${}_pF_q$ in the matrix argument can be written as an infinite series in terms of zonal polynomials, which are multivariate symmetric homogeneous polynomials and form a basis of the space of symmetric polynomials [19]. Given an $m \times m$ symmetric, positive-definite matrix Y, the hypergeometric function ${}_pF_q$ of matrix argument Y is defined as

$${}_pF_q\left(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix} \,\middle|\, Y\right) := \sum_{n=0}^{\infty} \sum_{\nu \in \mathcal{P}_n} \frac{(a_1)_\nu \cdots (a_p)_\nu}{(b_1)_\nu \cdots (b_q)_\nu} \cdot \frac{\mathcal{C}_\nu(Y)}{n!}, \tag{9}$$

where

- $\mathcal{P}_n$ is the set of all ordered integer partitions of n
- $(a)_\nu$ is the generalized Pochhammer symbol, defined as

  $$(a)_\nu = (a)_{(\nu_1, \ldots, \nu_k)} := \prod_{i=1}^{k} \left(a - \frac{i-1}{2}\right)_{\nu_i};$$

  , where, $(a)_{\nu_i} = a(a+1)...(a+\nu_i - 1), (a)_0 = 1$,
- and $\mathcal{C}_\nu(Y)$ denotes the zonal polynomial of Y, indexed by a partition ν, which is a symmetric homogeneous polynomial of degree n in the eigenvalues $y_1, \ldots, y_m$ of Y, satisfying

  $$\sum_{\nu \in \mathcal{P}_n} \mathcal{C}_\nu(Y) = (\operatorname{tr} Y)^n = (y_1 + \cdots + y_m)^n. \tag{10}$$

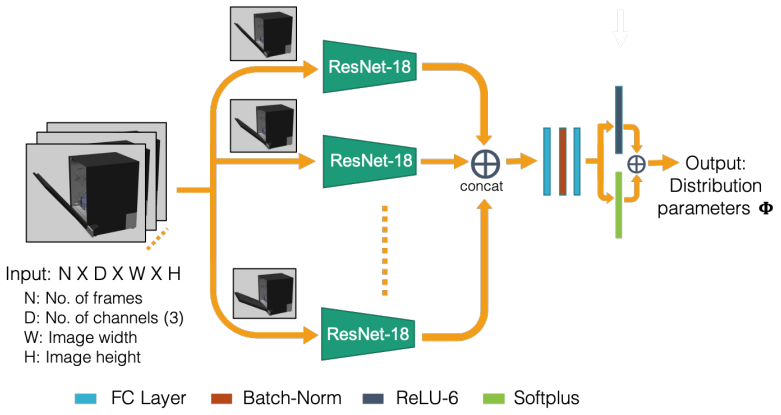


Figure 7: DUST-net architecture

Using zonal polynomials, we can define the hypergeometric function ${}_0F_1(\frac{3}{2}, \frac{1}{4}\Lambda^2)$ defining the normalization factor of the matrix von Mises-Fisher distribution over Stiefel manifold $\mathcal{V}_{3,2}$ as

$$ {}_0F_1(\frac{3}{2}, \frac{1}{4}\Lambda^2) := \sum_{n=0}^{\infty} \sum_{\nu \in \mathcal{P}_n} \frac{1}{(\frac{3}{2})_\nu} \frac{C_\nu(\Lambda)}{n!}, \tag{11} $$

where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2)$, $\mathcal{P}_n$ is the set of all ordered integer partitions of n, $(a)_\nu$ is the generalized Pochhammer symbol, and $C_\nu(\Lambda)$ denotes the zonal polynomial of Λ indexed by a partition ν. This series converges for all input matrices for a general hypergeometric function ${}_pF_q$ if $p \leq q$, which holds in our case [19]. Recently, Jiu and Koutschan [40] investigated the zonal polynomials in detail and developed a computer algebra package to calculate these polynomials in SageMath. We use this package to calculate the the hypergeometric function ${}_0F_1(\frac{3}{2}, \frac{1}{4}\Lambda^2)$. However, as the number of terms in the series grows combinatorially with n, we truncate the series at $n = 25$ for computational reasons. Through our experimental analysis, we found that this truncated series is a good approximation of ${}_0F_1$ as the series converges to a finite value, if the singular values of the F, i.e. λ_1 and λ_2 remain below a maximum value $\lambda_{max} = 50$.

D Network Architecture

Figure 7 shows the detailed network architecture for DUST-net. DUST-net uses an off-the-shelf convolutional network, ResNet-18, to extract task-relevant visual features from the input images, which are later passed through a two-layer MLP to predict a set of parameters Φ for the distribution $p(\mathsf{S}, \theta_{1:n-1}, d_{1:n-1} \mid \mathcal{I}_{1:n}, \Phi)$. We use ReLU activations for the hidden fully-connected layers. The first four output parameters (out of 40) of the last linear layer of MLP correspond to the parameters (α, β, γ) and ω, representing the matrices Γ and Ω respectively, which lie in ranges $[0, 2\pi)$, $[0, \pi)$, $[0, 2\pi)$, and $[0, 2\pi)$ respectively. We pass the first four values of the output of the last linear layer through a ReLU-6 layer [41] to correctly map the predicted values with their respective ranges. The rest of the parameters are required to be non-negative. We pass the remaining output values of the last linear layer through a Softplus layer for non-negative output.

Figure 8: Object classes used from the simulated articulated object dataset [11]. Object classes: cabinet, drawer, microwave, and toaster (left to right)

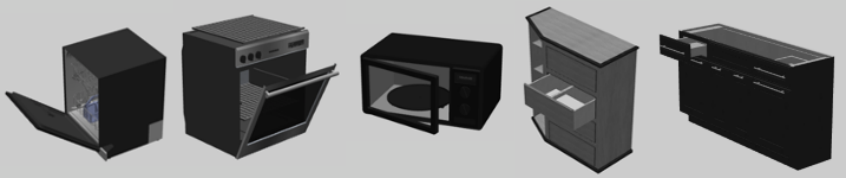


Figure 9: Object classes used from the PartNet-Mobility dataset [20–22]. Object classes: dishwasher, oven, microwave, drawer- 1 column, and drawer- multiple columns (left to right)

Figure 10: Real world objects used to evaluate DUST-net's performance. Object classes: microwave, drawer, and toaster (left to right)

E Experimental details

E.1 Datasets

Objects used in the experiments from each of the dataset are shown in the Figures 8 and 9. We sampled a new object geometry and a joint location for each training example in the simulated articulated object dataset, as proposed by [11]. For the PartNet-Mobility dataset, we considered 11 microwave (8 train, 3 test), 36 dishwasher (27 train, 9 test), 9 oven (6 train, 3 test), 26 single column drawer (20 train, 6 test), and 14 multi-column drawer (10 train, 4 test) object models. For both datasets, we sampled object positions and orientations uniformly in the view frustum of the camera up to a maximum depth dependent upon the object size. The objects and depth images are rendered in Mujoco [34]. We apply random frame skipping and pixel dropping to simulate noise encountered in real world sensor data. We consider three household objects — a microwave, a drawer, and a toaster oven, in the real world objects dataset for evaluating DUST-net's performance. The objects are shown in Figure 10.

To generate the labels for screw displacements, we follow the same procedure as used by Jain et al. [14]. Considering one of the objects, o_i, as the base object, we calculate the screw displacements between temporally displaced poses of the second object o_j with respect to it. Given a sequence of n images $\mathcal{I}_{1:n}$, we calculate a sequence of $n-1$ screw displacements ${}^1\boldsymbol{\sigma}_{o_j} = \{{}^1\sigma_2, ...{}^1\sigma_n\}$, where each ${}^1\sigma_k$ corresponds to the relative spatial displacement between the pose of the object o_j in the first image $\mathcal{I}_1$ and the images $\mathcal{I}_{k,\ k\in\{2...n\}}$. Note ${}^1\boldsymbol{\sigma}_{o_j}$ is defined in the frame $\mathcal{F}_{o_j^1}$ attached to the pose of the object o_j in the first image $\mathcal{I}_1$. We then transform ${}^1\boldsymbol{\sigma}_{o_j}$ to the camera frame by defining the 3D line motion matrix $\tilde{D}$ between the frames $\mathcal{F}_{o_j^1}$ and $\mathcal{F}_{o_i}$ [42], and transforming the common screw axis ${}^1\mathsf{S}$ to the target frame $\mathcal{F}_{o_i}$. The configurations ${}^1q_k$ remain the same during frame transformations. The 3D line motion matrix $\tilde{D}$ between two frames can be constructed using the rotation matrix R and a translation vector $\mathbf{t}$ between two frames $\mathcal{F}_A$ and $\mathcal{F}_B$, as:

$$\begin{bmatrix} {}^B\mathbf{l} \\ {}^B\mathbf{m} \end{bmatrix} = {}^B\tilde{D}_A \begin{bmatrix} {}^A\mathbf{l} \\ {}^A\mathbf{m} \end{bmatrix}, \quad \text{where,} {}^B\tilde{D}_A = \begin{bmatrix} R & \mathbf{0} \\ [\mathbf{t}]_\times R & R \end{bmatrix}, [\mathbf{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \tag{12}$$

where $[\mathbf{t}]_\times$ denotes the skew-symmetric matrix corresponding to the translation vector $\mathbf{t}$, and $({}^A\mathbf{l}, {}^A\mathbf{m})$ and $({}^B\mathbf{l}, {}^B\mathbf{m})$ represents the line l in frames $\mathcal{F}_A$ and $\mathcal{F}_B$, respectively [42].

	MAAD / SL	MAAD		Screw Loss	MAAD	SL	MAAD	SL	Precision				
	$\mathbf{l}$	$\hat{\mathbf{m}}$	$\|\mathbf{m}\|$	$D(\mathsf{S}_{GT}, \mathsf{S}_{pred})$	θ_i	$\theta_\mathbf{l}$	d_i	$d_\mathbf{l}$	$\lambda_\mathbf{l}$	$\lambda_{\hat{\mathbf{m}}}$	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d
vm-SoftOrtho	**0.139**	**0.154**	0.068	0.956	**0.012**	**0.117**	0.003	0.006	**56.2**	**55.8**	9.8	47.9	89.5
Direct **F**	0.240	0.261	0.062	0.104	0.010	0.208	0.002	0.006	8.4	7.9	9.8	48.5	75.3
ScrewNet	0.846	0.929	0.486	0.475	0.115	0.217	0.111	0.118	-	-	-	-	-
Abbatematteo et al. [11]	0.194	-	-	0.111	0.223	-	0.045	-	-	-	-	-	-
DUST-net	0.151	0.163	**0.052**	**0.059**	**0.012**	0.122	**0.002**	**0.006**	53.8	54.0	**18.3**	**128.1**	**219.1**
ScrewNet (Local)	0.178	0.443	0.068	0.033	0.057	0.118	0.015	0.015	-	-	-	-	-

Table 1: Mean error values on the MAAD and Screw Loss(SL) metrics for the simulated articulated objects dataset [11]. Point estimates for DUST-net correspond to the modes of the distributions predicted by DUST-net. Angular values $\{\mathbf{l}, \hat{\mathbf{m}}, \theta_i, \theta_\mathbf{l}\}$ and distances $\{\|\mathbf{m}\|, D, d_i, d_\mathbf{l}\}$ are reported in radian and meter, respectively. Numerical values are reported for the uncertainty parameters $\{\lambda_i, \beta_j\}$. Symbol $-$ represents value not reported.

	MAAD / SL	MAAD		Screw Loss	MAAD	SL	MAAD	SL	Precision				
	$\mathbf{l}$	$\hat{\mathbf{m}}$	$\|\mathbf{m}\|$	$D(\mathsf{S}_{GT}, \mathsf{S}_{pred})$	θ_i	$\theta_\mathbf{l}$	d_i	$d_\mathbf{l}$	$\lambda_\mathbf{l}$	$\lambda_{\hat{\mathbf{m}}}$	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d
vm-SoftOrtho	0.284	0.243	0.221	1.137	0.030	0.086	0.012	0.027	26.9	31.1	5.7	54.5	60.9
Direct **F**	**0.214**	**0.212**	0.257	0.219	0.030	0.064	0.012	**0.024**	8.1	7.3	4.9	59.5	70.9
ScrewNet	0.846	0.929	0.486	0.475	0.115	0.217	0.111	0.118	-	-	-	-	-
Abbatematteo et al. [11]	0.989	-	-	**0.095**	0.141	-	0.085	-	-	-	-	-	-
DUST-net	0.220	0.219	**0.178**	0.189	**0.029**	**0.063**	**0.012**	0.029	**49.3**	**48.3**	**7.7**	**72.0**	**131.9**
ScrewNet (Local)	0.260	1.23	0.314	0.151	0.060	0.106	0.040	0.009	-	-	-	-	-

Table 2: Mean error values on the MAAD and Screw Loss(SL) metrics for the PartNet-Mobility dataset [20–22]. Point estimates for DUST-net correspond to the modes of the distributions predicted by DUST-net. Angular values $\{\mathbf{l}, \hat{\mathbf{m}}, \theta_i, \theta_\mathbf{l}\}$ and distances $\{\|\mathbf{m}\|, D, d_i, d_\mathbf{l}\}$ are reported in radian and meter, respectively. Numerical values are reported for the uncertainty parameters $\{\lambda_i, \beta_j\}$. Symbol $-$ represents value not reported.

F Further Results

F.1 Accuracy of Point Estimates

Detailed numerical results for the synthetic articulated objects dataset and the PartNet-Mobility dataset are shown in Tables 1 and 2, respectively. Results demonstrate that under both metrics, the estimates obtained from DUST-net are considerably more accurate than those obtained from the state-of-the-art methods. DUST-net also correctly estimates very high distribution concentration parameters for the true, noise-free labels. The first baseline, vm-SoftOrtho, performs comparably with DUST-net on both datasets when only MAAD estimates are considered. However, Tables 1 and 2 show that it produces a very high distance ($\approx$ 1m) between the predicted and ground-truth screw axes. This error arises due to the soft-orthogonality constraint used by vm-SoftOrtho, as DUST-net and the second baseline method, both of which handle the constraint implicitly, do not report high errors on that metric. Meanwhile, the second baseline, Direct F, performs comparably with DUST-net on both metrics for both datasets, but fails to capture the uncertainty over parameters with the required accuracy.

F.2 Uncertainty Estimation

The detailed numerical results from the second set of experiments are shown in Table 3. In the noise-less case, the singular values of the matrix von Mises-Fisher distribution increases until they reach their maximum allowed value at $\lambda_{max} = 50$, while the precision parameters $\beta_j, j \in \{\|\mathbf{m}\|, \theta, d\}$ for truncated normal distributions over remaining parameters become arbitrarily large.

	λ_1	λ_2	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d	λ_1	λ_2	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d	λ_1	λ_2	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d	λ_1	λ_2	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d
Label Noise	No noise					15	15	50	50	50	12	12	50	50	50	10	10	50	50	50
SynArt	53.8	53.9	18.3	128.0	219.0	8.2	8.2	14.6	53.7	51.9	6.8	6.8	10.5	41.6	49.6	3.8	3.8	10.3	41.9	47.4
PartNet	49.3	48.3	7.7	72.0	132.0	6.4	6.3	9.4	29.5	29.2	4.9	4.7	8.9	34.0	37.9	3.2	3.1	9.4	31.2	32.1

Table 3: Testing variation of DUST-net's confidence over predicted articulation model parameters with input noise. DUST-net's confidence over its predicted parameters decreases monotonically as input noise is increased showing that DUST-net's predicted distribution captures the network's confidence over the predicted articulation parameters effectively.

		MAAD / SL	MAAD		Screw Loss	MAAD	SL	MAAD	SL	Precision				
		$\mathbf{l}$	$\hat{\mathbf{m}}$	$\|\mathbf{m}\|$	$D(\mathsf{S}_{GT}, \mathsf{S}_{pred})$	θ_i	$\theta_{\mathbf{l}}$	d_i	$d_{\mathbf{l}}$	$\lambda_{\mathbf{l}}$	$\lambda_{\hat{\mathbf{m}}}$	$\beta_{\|\mathbf{m}\|}$	β_θ	β_d
Toaster	ScrewNet	2.42	2.48	0.74	0.76	0.45	1.26	0.01	0.00	-	-	-	-	-
Oven	DUST-net	**0.17**	**0.31**	**0.52**	**0.59**	**0.44**	**0.64**	**0.01**	**0.01**	2.5	0.1	11.6	10.8	75.5
Microwave	ScrewNet	0.79	0.81	**0.13**	0.52	1.19	0.54	0.01	0.01	-	-	-	-	-
	DUST-net	**0.41**	**0.42**	0.22	**0.43**	**0.46**	**0.40**	**0.00**	**0.00**	0.7	0.6	19.7	14.3	39.9
Drawer	ScrewNet	0.69	**0.24**	0.49	**0.24**	**0.72**	0.97	0.08	**0.08**	-	-	-	-	-
	DUST-net	**0.42**	0.50	**0.32**	0.74	0.75	**0.56**	**0.07**	**0.08**	0.2	0.1	12.3	31.6	55.2

Table 4: Mean error values on the MAAD and Screw Loss metric for estimation of articulation model parameters for real-world objects when network was trained solely using simulated data. ScrewNet predictions are reported in the camera frame. Angular values $\{\mathbf{l}, \hat{\mathbf{m}}, \theta_i, \theta_{\mathbf{l}}\}$ and distances $\{\|\mathbf{m}\|, D, d_i, d_{\mathbf{l}}\}$ are reported in radian and meter, respectively. Numerical values are reported for the uncertainty parameters $\{\lambda_i, \beta_j\}$. Symbol $-$ represents value not reported.

F.3 Real objects

The numerical results from the sim-to-real transfer experiments are shown in Table 4. Results report that while DUST-net outperforms ScrewNet in estimating the model parameters for real-world objects, the estimated parameters are not yet accurate enough to be used directly for manipulating these objects. However, a noteworthy insight from the results is that DUST-net also reported very low confidence over the predicted parameters. This clearly delineates why it is beneficial to estimate a distribution over the articulation model parameters instead of only point estimates, as discussed earlier in the section 5.3.