

Are Euler angles a useful rotation parameterisation for pose estimation with Normalizing Flows?

Giorgos Sfikas¹

gsfikas@uniwa.gr

Konstantina Nikolaidou²

konstantina.nikolaidou@ltu.se

Foteini Papadopoulou¹

fpapadopoulou@uniwa.gr

George Retsinas³

gretsinas@central.ntua.gr

Anastasios L. Kesidis¹

akesidis@uniwa.gr

¹ Dpt. of Surveying and Geoinformatics Engineering, University of West Attica, Athens, Greece

² Dpt. of Computer Science, Luleå University of Technology, Luleå, Sweden

³ School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

Abstract

Object pose estimation is a task that is of central importance in 3D Computer Vision. Given a target image and a canonical pose, a single point estimate may very often be sufficient; however, a probabilistic pose output is related to a number of benefits when pose is not unambiguous due to sensor and projection constraints or inherent object symmetries. With this paper, we explore the usefulness of using the well-known Euler angles parameterisation as a basis for a Normalizing Flows model for pose estimation. Isomorphic to spatial rotation, 3D pose has been parameterized in a number of ways, either in or out of the context of parameter estimation. We explore the idea that Euler angles, despite their shortcomings, may lead to useful models in a number of aspects, compared to a model built on a more complex parameterisation.

1 Introduction

Estimating the pose of objects from a single RGB image is an important and challenging problem in computer vision [1, 10, 20]. It is closely connected to several scientific challenges, such as navigation in 3D scenes, augmented and virtual reality, robotic manipulation, and autonomous driving [11, 33]. In its more general form, the problem of estimating object pose encompasses 6 degrees of freedom (DOF); 3DOF for spatial position and 3DOF for spatial rotation. In this work, we focus on 3D object pose estimation, i.e., estimation of object spatial rotation [1]. Furthermore, we are interested in obtaining a probability density function instead of just a single point estimate, as in standard rotation regression (e.g. [17]). The motivation for this choice is that a probabilistic formulation of the problem constitutes

an elegant way to describe both ambiguities due to inherent uncertainty of pose given a single image related to the characteristics of central projection, as well as uncertainty arising due to the geometry of the object itself. A deterministic prediction will be inadequate given the multimodal nature of the problem, resulting in oversimplified solutions that ignore symmetry and uncertainty. Hence, in general there is an irreducible, aleatoric component of uncertainty that is incorrect to disregard. Furthermore, the probabilistic approach subsumes the non-probabilistic case as the latter can always be described as a Dirac PDF around the point estimate. Several parametric distributions have been proposed for probabilistic modeling of rotations, including the von Mises distribution for Euler angles [18, 27], the Bingham distribution for quaternions [3, 9], and the Matrix Fisher distribution for rotation matrices [23]. While these models are statistically grounded, their unimodal nature limits their ability to accurately capture symmetric objects, which are common in real-world environments [18].

In the proposed approach, we use Normalizing Flows (NFs) as our model for density estimation. NFs are a powerful and flexible framework for learning complex distributions through invertible mappings from simple base distributions. Although NFs have been successfully applied in Euclidean spaces [5], their extension to the non-Euclidean manifold $SO(3)$ remains underexplored. Part of existing approaches, such as ReLie [6] and ProHMR [15], attempt to project latent Euclidean flows onto $SO(3)$. Other methods [5, 19, 21] develop flows on general Riemannian manifolds without exploiting the geometric structure of $SO(3)$, often resulting in limited representational capacity and suboptimal performance [18]. Liu *et al.* [18] suggest constructing a flow as a composition of Möbius transforms and the newly introduced quaternion affine layer. Their Möbius transform works as a modification of the original use for spheres in [29], where a coupling approach uses one column of a rotation matrix as input for the conditioner, while the coupling transformer acts on a second column. As the output of the transform is not necessarily a valid rotation matrix, in each flow step, transform parameters are projected so that the result stays in $SO(3)$. The quaternion affine layers act via a linear transform over a \mathbb{R}^4 representation of the rotation quaternion. Again, the result is projected to the rotation manifold. *In this paper, we explore the following idea:* Given the complexity of the rotation matrix parameterisation in the context of estimation with NFs, can it be beneficial to employ a simpler parameterisation of rotation? We hypothesize that a simpler parameterisation may constitute a simpler objective for our learning model. Ideally, we would want a lower dimensionality parameterisation, and one that would not require per-flow step reprojections to the rotation manifold as in [18]. Euler angles are a parameterisation that fits this description [7]; they are the simplest parameterisation possible, using 3 unconstrained angles of rotation about fixed axes, a number that matches the intrinsic degrees of freedom of spatial rotation. Euler angles have been (rightfully) critiqued due to well-known shortcomings (cf. Section 3). However, they are very much in use in specific contexts, for example, in Photogrammetry [32]. We test our assumption with experimental trials on a number of datasets that cover diverse object geometries that may encompass complex symmetries. We show that the proposed model will lead to better density estimation compared to other models, including NFs that use rotation matrices for parameterisation [18]. We also check the effectiveness of our model on cases where data lie on subsets of $SO(3)$ that are related to gimbal lock singularities for Euler angles. Qualitative results suggest that Euler angles are related to different inductive biases that may often be suboptimal (cf. e.g. Figure 1). However, overall, we obtain slightly better fits than rotation matrix parameterisation NFs. Perhaps not unsurprisingly, an exception to this rule is datasets with ground truth heavily concentrated around points of Euler angle singularity.

The remainder of the paper is structured as follows. Section 2 discusses extensions of

flow models to non-Euclidean domains, including $\mathcal{SO}(3)$. The proposed alternative is presented in Section 3, including a succinct introduction on Euler angles and their advantages and shortcomings as a parameterisation of rotation. Experimental results are presented in Section 4, and we close the paper with our conclusions and future work in Section 5.

2 Normalizing Flows on circular and spherical domains

This section reviews using NFs on non-Euclidean domains like \mathbb{S} or $\mathcal{SO}(3)$. Much of the framework has been laid down in [29], while [18] have adapted the toolset for use with the rotational manifold.

Elements. The basic idea in NFs is as follows. Consider first a simple Gaussian model for density estimation: $p(x; \mu, \Sigma) \propto \exp\{-\frac{1}{2}\|g_{\mu, \Sigma}(x)\|^2\}$, where $g_{\mu, \Sigma}(x) = \Sigma^{-1/2}(x - \mu)$. In NFs, the affine $g_{\bullet}(x)$ is replaced by a non-linear transformation, and in particular, a special class of invertible neural networks. Of course, a closed-form Maximum Likelihood (ML) solution is out of the question if $g_{\theta}(\cdot)$ is a neural network, but we can still use the arsenal of gradient-based optimisation to obtain reasonable estimates (much like in the rest of Deep Learning). Furthermore, invertibility and differentiability of $g_{\theta}(\cdot)$ ensure that processes like sampling and evaluation with respect to the learned distribution are tractable and conceptually very straightforward. For example, sampling amounts to simply drawing a sample from a standard Gaussian and transforming it according to the inverse transformation $g_{\theta}^{-1}(\cdot)$. So, the likelihood model in NFs is: $p(x; \theta) \propto \exp\{-\frac{1}{2}\|(g_{\theta}(x))\|^2\}$, where $g_{\theta}(x)$ is defined as a non-linear bijection.¹ The major benefit of non-linearity over $g_{\theta}(x)$ is that we can model any distributional form (given constraints for the form of g) [14].

Conditional flows. NFs can be used in a supervised setting, and similar to the paradigm of GANs or Diffusion models, here, too, we have a conditional counterpart of the basic model. In terms of density estimation, we aim to learn $p(x|z)$ instead of $p(x)$. In the current application context of pose estimation, condition z is an input image (or image plus canonical pose), and the expected outcome is a distribution over likely poses.

Modeling of joint or more complex distributions follows similar considerations, whereas $p(x_1, x_2)$ can be written as $p(x_2|x_1)p(x_1)$, and we need a way to model conditional PDFs with flows. Coupling flows and autoregressive flows are two alternatives that may be used to model conditional flows.

Handling circular and spherical topologies. NFs on bounded and circular domains like circles (\mathbb{S}) or Tori (\mathbb{T}) require special considerations. First, the Gaussian as a base distribution is inadequate, as it implies an unbounded, Euclidean topology, for any $n \in \mathbb{N}$ and \mathbb{R}^n . Second, flow layers also need to be carefully designed so that the circular character of these topologies is respected. For example, transformations that are defined over \mathbb{S} (the circle) should take into account the periodicity of the domain.

¹The larger part of the Normalizing flow literature follow the convention that the normalizing transform g is the “inverse” direction and g^{-1} is the “forward” direction [26, Ch.16]. The current convention is used to simplify formulae and follow the implementation of [18]. Hence, here g represents the normalizing direction and g^{-1} represents the generative direction [26, Ch.16.1.2]

Rezende *et al.* [29] have pioneered working with NFs on circles and tori, and have introduced a number of suitable flow layers to this end. A number of options that satisfy the conditions set by [29] among which the Möbius transformations [13]. Other options are circular splines and non-compact projections, the latter of which is a complex form of the Möbius transform [29, Appendix H]. The Möbius transformation [13] is defined as follows. Given a D -dimensional sphere \mathbb{S}^D , we can consider it as the locus of points with unitary distance from the origin in \mathbb{R}^{D+1} . Let $w \in \mathbb{R}^{D+1}$ be a vector with $\|w\| < 1$, which will act as the parameter of the transformation. Then a point $x \in \mathbb{R}^{D+1}$ is mapped according to the Möbius transform with parameter w as follows:

$$g_w(x) = \frac{1 - \|w\|^2}{\|x - w\|^2}(x - w) - w. \quad (1)$$

It can be easily checked that $w = 0$ corresponds to an identity transformation. The part of the sphere that is close to the parameter point w is expanded as a result, and the rest is contracted.

One crucial property that, in the context of flow expressivity, proves to be a drawback is that the group of Möbius transformations is closed under composition [13, Theorem 2]. There are two workarounds concerning this point: either use a convex combination of two or more Möbius transforms, or compose a flow using other flow layers other than Möbius transforms in-between them. Both solutions are used in practice. Unlike the Möbius transform itself, there is no closed-form for the inverse of the convex combination of Möbius transforms. However, it can be numerically inverted with precision ϵ using bisection search within $\mathcal{O}(\log \epsilon^{-1})$ iterations.

Defining flows on $\mathcal{SO}(3)$. In Liu *et al.* [18], a NF for $\mathcal{SO}(3)$ is defined. This is defined as a cascade of Möbius layers and Quaternion affine layers. In order to capture the geometric constraints of the $\mathcal{SO}(3)$ manifold, the Möbius transforms are defined in a very specific manner, acting in each flow layer over a single column of the rotation matrix $R = [r_1 \ r_2 \ r_3]$, where r_i are the columns of R . As the column to be transformed, say r_2 , must adhere to being unit-norm, it is treated as a point on the two-dimensional sphere \mathbb{S}^2 . At the same time, r_2 must be orthogonal to r_1 and r_3 . The strategy of [18] is to use one of the remaining columns as the conditioner, meaning that it will act as input to the coupling NN, which will output Möbius transform parameters. The transformer column is ensured to be orthogonal to the conditioner column by applying a suitable projection over the Möbius transforms parameters ($w = w' - r_1[r_1 \cdot w']$), where w and w' are output and input parameters, respectively. The other column, say r_3 , will be computed as the outer product of the rest ($r_3 = r_1 \times r_2$). With this, all rotation matrix constraints are met. In each subsequent Möbius layer, the role of each column, i.e., whether it is a conditioner, a transformer, or it is computed via the outer product, is changed in order to ensure that no single column has any special bias. Quaternion affine layers act by defining a linear transformation over a real-valued representation of the rotation quaternion. Judging by the presented results of [18], between Möbius and Quaternion affine, Möbius transforms arguably contribute the more “added value” to the model.

3 Proposed Euler Angles-based Flow

Euler angles. NFs work by defining a transformation over data, the characteristics and motivation of which we have already described. In the context of the current application, our data are elements of the $\mathcal{SO}(3)$ manifold. Various ways have been put forward to pa-

parameterize $\mathcal{SO}(3)$, each one with its advantages and disadvantages. In section 2 we saw that previous work uses the Rotation matrix [18].

Perhaps the simplest parameterisation available is by using “Euler angles”. Their motivation is related to a celebrated theorem by Euler, which states that any two independent orthonormal coordinate frames can be related by a sequence of rotations about coordinate axes, where no two successive rotations may be about the same axis [16, Section 4.3]. Crucially, *not more than three* successive rotations are necessary to describe any arbitrary rotation. Also, the axes of rotation may be chosen arbitrarily. A straightforward choice of axes is the coordinate frame axes, X, Y, Z . Hence, given a fixed sequence of axes, we only require three angles to describe a rotation, which are called Euler angles [7, 16]. A usual convention in Photogrammetry is to denote rotations about X, Y, Z with letters ω, ϕ, κ respectively (for example, [12, 32]). These define a set of rotation matrices:

$$R_\omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & \sin \omega \\ 0 & -\sin \omega & \cos \omega \end{bmatrix}, R_\phi = \begin{bmatrix} \cos \phi & 0 & -\sin \phi \\ 0 & 1 & 0 \\ \sin \phi & 0 & \cos \phi \end{bmatrix}, R_\kappa = \begin{bmatrix} \cos \kappa & \sin \kappa & 0 \\ -\sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

and the full rotation is written as a composition of the three composing rotations. Corresponding to the sequence of rotation around X, Y , then Z , we have $R = R_\kappa R_\phi R_\omega$. An inverse relation, taking us from the rotation matrix to Euler angles, can be elaborated. However, here manifests the problem of “gimbal lock”: there exist areas that are not matched to a set of Euler angles in a unique manner. For example, for $\phi = \frac{\pi}{2}$ we can only determine the difference of Euler angles $\omega - \kappa$, but not each one of them separately:

$$R = \begin{bmatrix} 0 & \cos(\omega - \kappa) & \sin(\omega - \kappa) \\ 0 & -\sin(\omega - \kappa) & \cos(\omega - \kappa) \\ 1 & 0 & 0 \end{bmatrix} \quad (3)$$

A similar result holds for $\phi = -\frac{\pi}{2}$. In our implementation, we redefine these relations to be bijective: Euler angles to rotation matrices are forced to be unambiguous definitions. For example, when converting rotation matrices to Euler angles and $\cos \phi = 0$, we use the convention $\kappa = 0$ and the remaining degree of freedom is assigned effectively to ω . Note that the measure described here by no means “solve” gimbal lock in Euler angles. Our point of view is that we may be able to take advantage of the positive aspect related to Euler angles (i.e., their simplicity to implement), under the premise that the greater part of the $\mathcal{SO}(3)$ manifold does not suffer from this singularity.

Defining flows with Euler angles. The proposed flow uses the Möbius transform as its backbone. Compared to a flow that uses rotation matrices (cf. sec. 2, the implementation is arguably much more straightforward. In each flow step, our input is the three Euler angles $x = (\omega, \phi, \kappa)$. The domain of each is taken to be a unit sphere \mathbb{S}^1 , so ω, ϕ, κ all take values $\in [0, 2\pi]$. Equivalently, we have $x \in \mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$. Our strategy is to use a coupling flow (cf. Section 2), where in each flow step, one subset of the Euler angles will correspond to the conditioner, and the other corresponds to the transformer. Assuming angle κ is being transformed in a given step, we have:

$$\omega' = \omega, \phi' = \phi, \kappa' = \sum_{i=1}^K \rho_i \left[\frac{1 - \|w_i\|^2}{\|\kappa - w_i\|^2} (\kappa - w_i) - w_i \right],$$

Table 1: Comparisons of the proposed Euler angles-based model (“Ours”) versus prior models. The quality of fit to pose data is evaluated using a log-likelihood measure.

log likelihood \uparrow	avg.	peak	cone	cube/fisher24	line
Riemannian [21]	5.82	13.47	8.82	1.02	-0.026
ReLie [6]	-	-	5.32	3.27	-6.97
IPDF [25]	4.38	7.30	4.75	4.33	1.12
Mixture MF [23]	6.04	10.52	8.36	4.52	0.77
Moser Flow [30]	6.28	11.15	8.22	4.42	1.38
Möbius [18]	7.28	13.93	8.99	4.81	1.38
Möbius+Affine Flow [18]	7.28	13.93	8.99	4.81	1.38
Euler angles flow (ours)	8.07	15.09	9.35	5.71	2.14

where $\sum_{i=1}^K \rho_i = 1$ and $\rho_i \geq 0$ for all $i \in [1, K]$. As parameters of the transformer, we have: $\{\rho_i, w_i\}_{i=1}^K = \phi[\cos\omega, \sin\omega, \cos\phi, \sin\phi]$, where $\phi[\cdot, \cdot, \cdot, \cdot]$ is defined as a standard (non-invertible) neural network, which will output parameters for the K combined Möbius kernels. Akin to a minimal version of positional encoding [22], we have mapped conditioner Euler angles as $\theta \mapsto (\cos\theta, \sin\theta)$, in order to allow the network $\phi[\cdot, \cdot, \cdot, \cdot]$ to use the fact that the domain of the angles is periodic. The output parameters are fed to the coupling neural network. In effect, the above equations describe a coupling transform where the transformer is defined via a convex combination of Möbius transforms over \mathbb{S}^1 . Of course, this formulation will allow transformation only for the Euler angle κ . In order to allow transformations for the other two Euler angles, we apply permutations of the role of each angle, in a round-robin manner.

The end result is a flow model that aims to perform estimation in a distinctly simpler manner than the approach based on rotation matrix parameterisations (cf. Section 2, [18]). The dimensionality of the target quantity is lower, where we have a single angle in each step, and unconstrained, meaning that no projections are required. We argue that this scheme is more convenient for learning and inference, both in the sense of a learning objective, as well as in terms of computational load.

4 Experimental Results

We conduct experiments to evaluate our proposed method, deploying Euler angles on NFs. We have run experiments on a collection of different datasets that cover different aspects of the task. Datasets “Synthetic” and “Gimbal” are used to gauge performance on unconditional density estimation on rotation data. Datasets “SYMSOL I” and “ModelNet10-SO3” are benchmarks that involve image input data, covering both objects that lend to inherently multimodal as well as unimodal pose ground truths.

Synthetic dataset. Each subset of this dataset is made up of a number of samples off the $\mathcal{SO}(3)$ manifold. It was introduced in [18], and the quality of the fit to the data is measured in terms of log-likelihood, as shown in Table 1. We use 24 Euler angle-based Möbius flow layers, each of which uses $K = 64$ Möbius kernels. A batch size of 1024 and learning rate equal to 10^{-4} is used. Models are trained for 50k iterations. Samples from learned distributions can be examined in fig. 1 and 2. We use the visualisation tool of [25]. The two models – the rotation matrix-based [18] and the proposed Euler angles flow model

– exhibit distinctly different inductive biases. We can discern two kinds of differences. First, the rotation matrix model handles cases where the PDF is made up of disjoint islands of mass. Second, the Euler angles model tends to estimate variance more faithfully. Overall, these differences stack up to a slight advantage for the Euler angles implementation, as per the numerical results of Table 1.

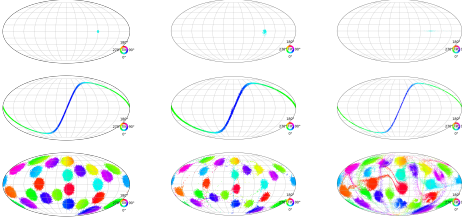


Figure 1: Results for “synthetic” dataset. From top row to bottom row: Ground truth and results for subsets “peak”, “cone”, “cube”. From left column to right column: Ground truth distribution, samples generated by the model proposed in [18], and samples generated by the proposed Euler angles-based flow model. The rotation matrix of [18] tends to model better disjoint “islands” of mass; however, the proposed Euler angles model is better at estimating variance. (see also comment on fig. 2).

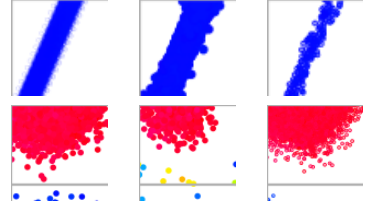


Figure 2: Detail from Figure 1. From left column to right column: Ground truth distribution, samples generated by the model proposed in [18], and samples generated by the proposed Euler angles-based flow model. Note the difference in inductive biases between the two models: Liu *et al.* [18] tends to misestimate variance by a considerable margin; the proposed Euler angles flow is closer to the correct value.

Gimbal dataset. We introduce an additional dataset for unsupervised density estimation that we call *Gimbal*, in order to gauge the efficiency of the proposed Euler flows exactly with areas of $\mathcal{SO}(3)$ that are problematic for our parameterisation. Regardless of the choice of axes in an Euler angles parameterisation, there will always exist parts of $\mathcal{SO}(3)$ where gimbal lock takes place (cf. discussion in Section 3). Under the convention we use, this happens when $\phi = \frac{\pi}{2}$ or $\phi = -\frac{\pi}{2}$. We produce random elements of $\mathcal{SO}(3)$ by sampling as follows: $\omega \sim \mathcal{N}(0, \sigma^2)$, $\phi \sim \frac{1}{2}\mathcal{N}(\frac{\pi}{2}, \sigma_\phi^2 \sigma^2) + \frac{1}{2}\mathcal{N}(-\frac{\pi}{2}, \sigma_\phi^2 \sigma^2)$, $\kappa \sim \mathcal{N}(0, \sigma^2)$, where we set $\sigma_\phi^2 = 0.1$. We set $\sigma^2 = \{1, 0.1\}$. Following the layout of the “synthetic” dataset discussed in the previous subsection, we split the sampled rotations into a training and test set that are made up of 60k and 12k iterations, respectively. We have compared the proposed model versus the model of Liu *et al.* [18], using the setups described in subsection 4. Numerical results are reported in Table 2. We can deduce that the more data is concentrated around areas of singularity, the harder it is for the Euler angles model to achieve a good fit. It is

Table 2: Results for “gimbal” dataset. Testing of the proposed Euler angles flow model in areas of singularity. Log-likelihood is used to evaluate the quality of fit.

log likelihood \uparrow	$\sigma^2=1$	$\sigma^2=0.1$
Möbius + Affine Flow [18]	4.50	13.31
Euler angles flow	2.74	9.46

worth noting that here we see a different picture than the one of the results of the previous synthetic experiment (Table 1), as there is a stark difference in log-likelihood, but now Euler angles are the model that fares less competitively. This is perhaps not surprising, as the data of this set cover deliberately the areas of $SO(3)$ related to Euler angles singularity.

SYMSOL I dataset. We perform experiments on SYMSOL I [25], which comprise sets of objects with pose ground truths that are inherently multimodal due to object symmetries. SYMSOL I contains high-order symmetry shape images such as tetrahedron, cube, cone, and cylinder. We train our model for 900k iterations and otherwise follow the training regime for the Möbius and Quaternion affine-based models of Liu *et al.* [18]. For our implementation, we use the same flow step sequence where we have disabled Quaternion affine flow layers. Otherwise, the default parameters for running the code were used. The results are presented in Table 3. Our method achieves the best log-likelihood performance for the cone and cylinder shapes and the second-best for the cube shape, resulting in the second-best average log-likelihood performance. As a comment on the results, it seems that there is a negative trend on objects which have ground truth pdfs with numerous disjoint islands of mass. Such are objects like the icosahedron or the tetrahedron. This is perhaps related to a similar effect in the “Synthetic” dataset (cf. Figure 1 & discussion in Section 5).

Table 3: Numerical results on SYMSOL I. Exponents M and $M + A$ correspond to using [18] with only Möbius layers or the full model. Best and second-best performance are highlighted with boldface and underline, respectively.

	log likelihood \uparrow					
	avg.	cone	cube	cyl.	ico.	tet.
[4]	0.81	2.45	-2.15	1.34	-0.16	2.56
[8]	1.86	6.13	0.00	3.17	0.00	0.00
[28]	0.42	-1.05	1.79	1.01	-0.10	0.43
[25]	6.39	6.74	7.10	6.55	3.57	7.99
[18] ^M	9.41	<u>10.52</u>	9.68	<u>10.00</u>	<u>5.35</u>	<u>11.51</u>
[18] ^{M+A}	10.38	<u>10.05</u>	11.64	<u>9.54</u>	8.26	12.43
Ours	<u>9.66</u>	12.36	<u>10.20</u>	11.64	4.15	9.97

Table 4: Numerical results of ModelNet10-SO3. Indices U and F denote the base distribution used for the flow in each case. Best and second-best performance is highlighted, with boldface and underline, respectively.

	Acc@ 15° \uparrow	Acc@ 30° \uparrow	Med. (°) \downarrow
[4]	0.562	0.694	32.6
[28]	0.456	0.528	49.3
[24]	0.693	0.757	17.1
[25]	0.719	0.735	21.5
[18] _U	0.760	0.774	14.6
[18] _F	<u>0.744</u>	<u>0.768</u>	12.2
Ours _U	0.742	0.753	19.5
Ours _F	0.736	0.754	<u>12.7</u>

ModelNet10-SO3 dataset. We further conduct experiments on the ModelNet10-SO3 dataset [17] that is widely used for regressing rotations from single images. The dataset is comprised of the uniformly rotated renderings of CAD models from the ModelNet10 dataset [34]. We present the performance using both the uniform and pre-trained Fisher distributions in terms of Acc@ 15°, Acc@ 30°, and Median Error in Table 4. As we can see, our proposed approach achieves competitive accuracies, close to the highest performing methods, while we achieve the second-best median error after [18] with the corresponding Fisher distribution.

4.1 Training speed comparison

Training speed comparison. We report a training speed comparison over the proposed Euler flows alternative, versus two of the versions presented in Liu *et al.* [18]. Speed tests were run on an RTX 2080Ti GPU. Tests on unsupervised density estimation (synthetic) use default settings, and conditional estimation uses a batch size of 2. Results are shown in Table 5. The speed of the proposed model can be attributed to the simpler architecture of the flow, owing to the minimal character of the Euler angles representation.

Table 5: Training speed comparison: Reported figures are milliseconds per training iteration, on average. Exponents M and $M + A$ correspond to using [18] with only Möbius layers or the full model. See text for details.

Time per iteration (ms) ↓	Unconditional	Conditional
[18] ^M	483	384
[18] ^{M+A}	966	734
Ours	387	335

5 Conclusion and Future Work

As a one-sentence answer to the titular question of this paper, “Are Euler angles a useful parameterisation of rotation for pose estimation with Normalizing Flows?” *Yes*, despite their shortcomings, they can be used to construct a useful model. Our experiments corroborate this point of view, and numerical results are routinely in the ballpark of the state of the art. However, our analysis has shown that there are several important nuances that must be taken into account if we are to decide whether we should use this model in a practical scenario versus a model that does not use this parameterisation. There do exist caveats, some of which were not completely surprising. In particular, the proposed model does not fare well in areas related to gimbal lock singularity. In a practical scenario, this, of course, must be taken into account; as with the use of Euler angles, the key factor here is whether or not we have prior knowledge that our angles are related to areas of singularity (cf. the case of aerophotogrammetry). A point that we believe is worth mentioning is related to the reasons for which we have slightly better results in many cases. This appears to be related to two factors: a) a different inductive bias than the one related to the rotation matrix model; b) a learning-wise “easier” objective. The Euler angles model seems to have some difficulty modeling those cases where there are disjoint masses in the ground truth. This is a well-known issue with Normalizing Flows in general, and there is current research that may lead to considerable improvements in this respect [2]. Another point that is noteworthy is that Euler angles lend to a much less costly model, as it can be trained faster than both alternatives of the rotation matrix that we compared against. Further, we have seen that the Euler angles model outperforms routinely other non-NF-based models for rotation estimation. Finally, let us note that the current implementation using Euler angles is not the only way that one can construct a flow with them, as there are several hyperparameters in place. First and foremost is the Möbius transform itself, or the way we implement coupling (one dimension for the transformer versus two for the conditioner), putting in place phase-translating flows to improve expressivity, and so on. Other considerations could include integration with more capable transformer backbones or fusing with other types of probabilistic models [31, 35]. This could be the subject of future work.

Acknowledgements

The publication/registration fees were partially covered by the University of West Attica.

References

- [1] Vassileios Balntas, Andreas Doumanoglou, Caner Sahin, Juil Sock, Rigas Kouskouridas, and Tae-Kyun Kim. Pose guided RGBD feature learning for 3D object pose estimation. In *Proceedings of the IEEE international conference on computer vision*, pages 3856–3864, 2017.
- [2] Harry Bevens, Will Handley, and Thomas Gessey-Jones. Piecewise normalizing flows. In *International Conference on Machine Learning (ICML)*, 2024.
- [3] Haowen Deng, Mai Bui, Nassir Navab, Leonidas Guibas, Slobodan Ilic, and Tolga Birdal. Deep bingham networks: Dealing with uncertainty and ambiguity in pose estimation. *International Journal of Computer Vision*, pages 1–28, 2022.
- [4] Jiankang Deng, Jiong Guo, and Stefanos Zafeiriou. Rotation-aware face alignment via bingham distribution guided rotation regression. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2021.
- [5] Luca Falorsi and Patrick Forré. Neural ordinary differential equations on manifolds. *arXiv preprint arXiv:2006.06663*, 2020.
- [6] Luca Falorsi, Pim de Haan, Tim R Davidson, and Patrick Forré. Reparameterizing distributions on lie groups. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3244–3253. PMLR, 2019.
- [7] Wolfgang Förstner and Bernhard P Wrobel. *Photogrammetric computer vision*, volume 6. Springer, 2016.
- [8] Igor Gilitschenski, Lennart Kühn, Mustafa Mukadam, Alan Willsky, and Daniela Rus. Deep orientation uncertainty learning based on bingham distributions. In *International Conference on Learning Representations (ICLR)*, 2019.
- [9] Igor Gilitschenski, Roshni Sahoo, Wilko Schwarting, Alexander Amini, Sertac Karaman, and Daniela Rus. Deep orientation uncertainty learning based on a bingham loss. In *International Conference on Learning Representations (ICLR)*, 2019.
- [10] Sabera Hoque, Md Yasir Arafat, Shuxiang Xu, Ananda Maiti, and Yuchen Wei. A comprehensive review on 3d object detection and 6d pose estimation with deep learning. *IEEE Access*, 9:143746–143770, 2021.
- [11] Sabera Hoque, Shuxiang Xu, Ananda Maiti, Yuchen Wei, and Md Yasir Arafat. Deep learning for 6d pose estimation of objects: a case study for autonomous driving. *Expert Systems with Applications*, 223:119838, 2023.
- [12] George E Karras and Elli Petsa. Metric information from uncalibrated single images. In *Proceedings XVII CIPA Symposium October*, volume 99, 1999.

- [13] Shogo Kato and Peter McCullagh. Möbius transformation and a cauchy family on the sphere. *arXiv preprint arXiv:1510.07679*, 2015.
- [14] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020.
- [15] Nikos Kolotouros, Georgios Pavlakos, Dinesh Jayaraman, and Kostas Daniilidis. Probabilistic modeling for human mesh recovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11605–11614, 2021.
- [16] Jack B. Kuipers. *Quaternions and rotation sequences: a primer with applications to orbits, aerospace, and virtual reality*. Princeton university press, 1999.
- [17] Shuai Liao, Efstratios Gavves, and Cees GM Snoek. Spherical regression: Learning viewpoints, surface normals and 3d rotations on n-spheres. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9759–9767, 2019.
- [18] Yulin Liu, Haoran Liu, Yingda Yin, Yang Wang, Baoquan Chen, and He Wang. Delving into discrete normalizing flows on $SO(3)$ manifold for probabilistic rotation modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21264–21273, 2023.
- [19] Aaron Lou, Derek Lim, Isay Katsman, Leo Huang, Qingxuan Jiang, Ser Nam Lim, and Christopher M De Sa. Neural manifold ordinary differential equations. In *Advances in Neural Information Processing Systems*, volume 33, pages 17548–17558, 2020.
- [20] Siddharth Mahendran, Haider Ali, and René Vidal. 3d pose regression using convolutional neural networks. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 2174–2182, 2017.
- [21] Emile Mathieu and Maximilian Nickel. Riemannian continuous normalizing flows. *Advances in Neural Information Processing Systems*, 33:2503–2515, 2020.
- [22] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [23] David Mohlin, Josephine Sullivan, and Gérald Bianchi. Probabilistic orientation estimation with matrix fisher distributions. *Advances in Neural Information Processing Systems*, 33:4884–4893, 2020.
- [24] Thomas Mohlin, Johan Karlsson, and Christopher Zach. Deep rotation regression with matrix fisher distributions. In *European Conference on Computer Vision (ECCV) Workshops*, 2020.
- [25] Kieran Murphy, Carlos Esteves, Varun Jampani, Srikumar Ramalingam, and Ameesh Makadia. Implicit representation of probability distributions on the rotation manifold. In *International Conference on Machine Learning*, 2021.
- [26] Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023. URL <http://udlbook.com>.

- [27] Sergey Prokudin, Peter Gehler, and Sebastian Nowozin. Deep directional statistics: Pose estimation with uncertainty quantification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [28] Sergey Prokudin, Christoph Lassner, and Peter Gehler. Deep directional statistics: Pose estimation with von mises-fisher distribution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [29] Danilo Jimenez Rezende, George Papamakarios, Sébastien Racaniere, Michael Albergo, Gurtej Kanwar, Phiala Shanahan, and Kyle Cranmer. Normalizing flows on tori and spheres. In *International Conference on Machine Learning*, pages 8083–8092. PMLR, 2020.
- [30] Noam Rozen, Aditya Grover, Maximilian Nickel, and Yaron Lipman. Moser flow: Divergence-based generative modeling on manifolds. *Advances in neural information processing systems*, 34:17669–17680, 2021.
- [31] Giorgos Sfikas, Christophoros Nikou, Nikolaos Galatsanos, and Christian Heinrich. Majorization-minimization mixture model determination in image segmentation. In *CVPR 2011*, pages 2169–2176. IEEE, 2011.
- [32] Efstratios Stylianidis. Measurements: Introduction to photogrammetry. *Photogrammetric Survey for the Recording and Documentation of Historic Buildings*, pages 139–195, 2020.
- [33] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint arXiv:1809.10790*, 2018.
- [34] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3D ShapeNets: A Deep Representation for Volumetric Shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015.
- [35] Shuangfei Zhai, Ruixiang Zhang, Preetum Nakkiran, David Berthelot, Jiatao Gu, Huangjie Zheng, Tianrong Chen, Miguel Angel Bautista, Navdeep Jaitly, and Josh Susskind. Normalizing flows are capable generative models. *arXiv preprint arXiv:2412.06329*, 2024.