PDQ-Net: Deep Probabilistic Dual Quaternion Network for Absolute Pose Regression on $SE(3)$

Abstract

Accurate absolute pose regression is one of the key challenges in robotics and computer vision. Existing direct regression methods suffer from two limitations. First, some noisy scenarios such as poor illumination conditions are likely to result in the uncertainty of pose estimation. Second, the output n-dimensional feature vector in the Euclidean space $\mathbb{R}^n$ cannot be well mapped to $SE(3)$ manifold. In this work, we propose a deep dual quaternion network that performs the absolute pose regression on $SE(3)$. We first develop an antipodally symmetric probability distribution over the unit dual quaternion on $SE(3)$ to model uncertainties and then propose an intermediary differential representation space to replace the final output pose, which avoids the mapping problem from $\mathbb{R}^n$ to $SE(3)$. In addition, we introduce a backpropagation method that considers the continuousness and differentiability of the proposed intermediary space. Extensive experiments on the camera re-localization task on the Cambridge Landmarks and 7-Scenes datasets demonstrate that our method greatly improves the accuracy of the pose as well as the robustness in dealing with uncertainty and ambiguity, compared to the state-of-the-art.

1 INTRODUCTION

Accurate absolute pose regression on $SE(3)$ is an age-old problem in the fields of robotics [Mur-Artal and Tardós, 2017, Wang et al., 2018, Campos et al., 2021] and computer vision [Kendall et al., 2015, Deng et al., 2019]. Early research focuses on geometry-based approaches due to their reliability and accuracy in some static environments. However, the geometry-based approaches cannot function well in some specific scenarios such as poor illumination and the textureless case.

In recent years, the deep learning technique has provided us with an alternative vehicle to regress the absolute pose. In these approaches, a proper output for representing the pose is particularly important when dealing with more complicated environments. A number of deep models have been presented to output an n-dimensional feature vector that is used to directly denote the pose, where the rotation is mostly represented by the Euler angle, unit quaternion, or rotation matrix, and translation is substituted by a 3-dimensional vector [Wang et al., 2018, Sattler et al., 2019].

In spite of some advancements, these learning-based techniques suffer from one or two limitations. First, some work cannot capture and model pose uncertainties in practical scenarios, which affects the estimation accuracy. Second, an n-dimensional feature vector on a Euclidean space $\mathbb{R}^n$ is taken as the common output by existing work [Shotton et al., 2013, Sattler et al., 2019, Xue et al., 2019], which however cannot be well mapped to $SE(3)$ manifold since the pose in $SE(3)$ is not homeomorphic to the Euclidean space $\mathbb{R}^n$ [LaValle, 2006]. For example, the angle 0 and 2π in $\mathbb{R}^n$ can map to the same rotation in $SE(3)$, which fails to satisfy the one-to-one mapping feature in homeomorphism theory.

In this paper, we propose a probabilistic deep dual quaternion network that regresses the absolute pose on $SE(3)$ from a single RGB image. Compared to existing learning-based approaches, our model has three advanced features. First, we take pose uncertainties into consideration by introducing an antipodally symmetric distribution over the unit dual quaternion on $SE(3)$. In this way, the pose regression problem can be converted into a deep probabilistic problem. Second, we propose to estimate the pose indirectly by presenting an intermediary differential representation space as the output of our deep probabilistic model. Afterward, the rotation and translation can be derived from the intermediary representation space by modeling a quadratically-constrained quadratic program (QCQP) problem and a Gaus-

Submitted to the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022). **To be used for reviewing only**.

sian process respectively, which avoids the issue of mapping from $\mathbb{R}^n$ to $SE(3)$. Third, we introduce a backpropagation method that considers the continuousness and differentiability of the proposed intermediary space. Extensive experiments on the camera re-localization task on the Cambridge Landmarks and 7-Scenes datasets demonstrate that our method outperforms the state-of-the-art, in terms of pose accuracy and robustness.

In summary, our work makes the following contributions:

- We develop an antipodally symmetric probability distribution over the unit dual quaternion on $SE(3)$ to model the pose uncertainty, which is the key factor in improving the accuracy of pose regression.
- We propose a deep dual quaternion model to regress the pose indirectly, which effectively addresses the mapping problem from $\mathbb{R}^n$ to $SE(3)$. Additionally, a backpropagation method is introduced to the proposed model.
- We implement our deep probabilistic model on the Cambridge Landmarks and 7-Scenes datasets. Extensive experiment results show that our method greatly improves the accuracy of the pose as well as the ability in dealing with uncertainty and ambiguity, compared to the state-of-the-art.

2 RELATED WORK

Absolute pose estimation combining with the deep learning technique becomes a hot topic in robotics and computer vision field. The related techniques roughly fall into two categories: direct pose regression and indirect pose regression.

Direct pose regression aims to directly regress the absolute pose from sequential RGB images by utilizing various well-designed end-to-end networks [Wang et al., 2018, Clark et al., 2017, Kendall et al., 2015, Shotton et al., 2013, Chen et al., 2021]. In these approaches, most of them follow the same pipeline: features are extracted using a defined network such as the PoseNet [Kendall et al., 2015], the BranchNet [Wu et al., 2017], which are then embedded into a high-dimensional vector that lies in the Euclidean space $\mathbb{R}^n$. Then Sattler et al. [2019] pointed out that this embedding layer typically corresponds to the output of the second-to-last layer in direct pose regression methods. The last layer performs a linear projection from the embedding space to the space of poses. However, these methods commonly suffer from one of two issues. First, the uncertainty of poses may result in the degradation of the accuracy of predicted poses since some of them are not robust enough to capture uncertainties [Kendall and Cipolla, 2016]. Second, the output n-dimensional feature vector lying on a Euclidean space $\mathbb{R}^n$ may not be well mapped to $SE(3)$ [LaValle, 2006].

A strategy to overcome these problems is to regress the absolute pose indirectly, which is achieved by regressing an intermediary vector to replace the output feature vector [Deng et al., 2020, Bui et al., 2020]. However, it is usually hard to find such a representation space. Poursaeed et al. [2018] introduced a Siamese model for uncalibrated cameras to regress a fundamental matrix that serves as an intermediary representation of camera poses, but it fails to capture uncertainties of poses. Recently, deep probabilistic models have been developed by regressing essential parameters of the probabilistic distribution. [Gilitschenski et al., 2020] introduced a deep Bingham model for the object orientation estimation on $SO(3)$ by regressing the orthogonal matrix of the Bingham distribution, where pose uncertainties are modeled as a Bingham distribution. [Mohlin et al., 2020] similarly developed a deep matrix Fisher distribution for object rotation estimation on $SO(3)$ by regressing the parameter matrix of the Fisher distribution. But it is generally hard to find such statistic distributions on $SE(3)$ manifold to measure pose uncertainties.

To this point, we develop an antipodally symmetric probability distribution over the unit dual quaternion on $SE(3)$ to model pose uncertainties. Based thereon, we present a deep probabilistic distribution to indirectly regress the absolute pose on $SE(3)$.

3 UNIT DUAL QUATERNION DISTRIBUTION

This section gives the definition of the unit dual quaternion distribution on $SE(3)$. For this purpose, we first briefly revisit the concept of dual quaternions and then give the description of the unit dual quaternion distribution.

3.1 DUAL QUATERNION

In this work, a quaternion $\mathbf{q}$ is defined as $\mathbf{q} = q_0 + q_1\mathbf{i} + q_2\mathbf{j} + q_3\mathbf{k}$, the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ is the standard basis of the three-dimensional Euclidean space $\mathbb{R}^3$. For convenience, we bring the vector $\mathbf{q} = [q_0, \mathbf{q}_{vec}] \in \mathbb{R}^4$ to denote the quaternion. The multiplication between two arbitrary quaternions can be done with a matrix-vector form that is given by

$$\mathbf{p} \odot \mathbf{q} = \mathbf{R}_q \mathbf{p} = \begin{bmatrix} q_0 & -\mathbf{q}_{vec}^T \\ \mathbf{q}_{vec} & -\mathbf{q}_{vec}^{\times} + q_0 \mathbf{I}_3 \end{bmatrix} \mathbf{p},$$

where $[\mathbf{a}]^{\times}$ denotes the skew-symmetric matrix formed from the vector $\mathbf{a}$, and $\mathbf{I}$ refers to the identity matrix.

The norm of a quaternion is defined as $\sqrt{\mathbf{q} \odot \mathbf{q}^*}$, with $\mathbf{q}^* = [q_0, -\mathbf{q}_{vec}]$ being the conjugate of $\mathbf{q}$. And quaternions with a unit norm are called unit quaternions, which are used for denoting the pure rotation on the unit hypersphere $\mathbb{S}^3 \subset \mathbb{R}^4$.

Dual quaternion consists of the real part quaternion and dual part quaternion which is a convenient tool for encapsulating

the rotation and translation,

$$\mathbf{v} = \mathbf{q}_r + \epsilon \mathbf{q}_d, \epsilon \neq 0, \epsilon^2 = 0, \tag{1}$$

where $\mathbf{q}_r$ is the unit quaternion for indicating the rotation, and $\mathbf{q}_d$ is the dual part quaternion for representing the composition of the rotation quaternion and translation quaternion $\mathbf{q}_t = [0, t_x, t_y, t_z]^T$, with $\mathbf{q}_d = 0.5\mathbf{q}_t \odot \mathbf{q}_r \in \mathbb{R}^4$.

Since the dual part $\mathbf{q}_d$ is orthogonal to the real part on the hypersphere space $\mathbb{S}^3$, we further get the unit dual quaternion manifold $\mathbb{DH}_1 := \{[\mathbf{q}_r^T, \mathbf{q}_d^T]^T | \|\mathbf{q}_r\| = 1, \mathbf{q}_r \in \mathbb{S}^3, \mathbf{q}_r^T \mathbf{q}_d = 0\} \subset \mathbb{R}^8$. Furthermore, the translation vector $\mathbf{t} = [t_x, t_y, t_z]$ can be recovered from $\mathbb{DH}_1$ according to [Li et al., 2021], which can be written as

$$\mathbf{t} = 2[\mathbf{R}_{qr}]_{1:3}^T \mathbf{q}_d, \tag{2}$$

where $[\mathbf{R}_{qr}]_{1:3}$ refers to the last three columns of the right multiplication matrix $\mathbf{R}_{qr}$.

3.2 EXPONENTIAL DISTRIBUTION OF UNIT DUAL QUATERNION

Unit dual quaternions $\mathbf{v}$ and $-\mathbf{v}$ denote the same transformation because of the property of unit quaternions, namely $\mathbf{q}_r = -\mathbf{q}_r$. We assume there exists an antipodally symmetric exponential distribution of unit dual quaternions $\mathbf{v} \in \mathbb{DH}_1$. Previous work [Gilitschenski et al., 2020] offered an exponential distribution for representing the orientation and position on $SE(2)$ manifold but failed to be applied on $SE(3)$. In this work, we bridge this gap by developing the unit dual quaternion distribution on $SE(3)$.

Definition 1. *A vector $\mathbf{v} \in \mathbb{DH}_1 \subset \mathbb{R}^8$ can be modeled as an antipodally symmetric distribution if its probability density function has the following form*

$$f(\mathbf{v}) = \frac{1}{N(\mathbf{F})} \exp(\mathbf{v}^T \mathbf{F} \mathbf{v}), \tag{3}$$

where $N(\mathbf{F})$ refers to the normalization constant of the proposed distribution. And the real symmetric matrix $\mathbf{F} \in \mathbb{R}^{8\times 8}$ is the parameter matrix.

We split the vector $\mathbf{v} = [\mathbf{q}_r^T, \mathbf{q}_d^T]^T$ with $\mathbf{q}_r \in \mathbb{S}^3$ and $\mathbf{q}_d \in \mathbb{R}^4$. Meanwhile, we decompose the symmetric parameter matrix $\mathbf{F}$ as follows,

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 \\ \mathbf{F}_2^T & \mathbf{F}_3 \end{bmatrix}, \mathbf{F}_i \in \mathbb{R}^{4\times 4}, i = 1, 2, 3.$$

Then the exponential distribution (3) can be rewritten as

$$\begin{aligned} f(\mathbf{v}) = \frac{1}{N(\mathbf{F})} \exp \big(& \underbrace{\mathbf{q}_r^T (\mathbf{F}_1 - \mathbf{F}_2 \mathbf{F}_3^{-1} \mathbf{F}_2^T) \mathbf{q}_r}_{Bingham-like} + \\ & \underbrace{(\mathbf{q}_d + \mathbf{F}_3^{-1} \mathbf{F}_2^T \mathbf{q}_r)^T \mathbf{F}_3 (\mathbf{q}_d + \mathbf{F}_3^{-1} \mathbf{F}_2^T \mathbf{q}_r))}_{Gaussian-like} . \end{aligned} \tag{4}$$

In this distribution, we have the following theorem for parameter matrix $\mathbf{F}$, where the proof can be found in the Supplementary Material.

Theorem 1. *Considering the antipodally symmetric distribution (4), the sub-block matrix $\mathbf{F}_1 \in \mathbb{R}^{4\times 4}$ is real symmetric, and $\mathbf{F}_3 \in \mathbb{R}^{4\times 4}$ is real symmetric and negative definite.*

4 SYMMETRIC MATRIX F ON $SE(3)$

This section gives a thorough analysis of the parameter matrix $\mathbf{F}$ in the unit dual quaternion distribution on $SE(3)$.

4.1 SYMMETRIC MATRIX F

As shown in Equation (4), $\mathbf{F}$ is decomposed into three submatrices, which are significant components of the Bingham-like distribution and the Gaussian-like distribution.

4.1.1 The Bingham-like Matrix

The Bingham distribution was introduced [Bingham, 1974] as an extension of the Gaussian distribution, which lies on the surface of the unit hypersphere,

$$f(\mathbf{q}; \mathbf{M}, \mathbf{Z}) = \frac{1}{N(\mathbf{Z})} \exp\left(\mathbf{q}^T \mathbf{M} \mathbf{Z} \mathbf{M}^T \mathbf{q}\right), \tag{5}$$

where $\mathbf{Z} \in \mathbb{R}^{4\times 4}$ is a diagonal matrix with an ascending entries $z_1 \leq z_2 \leq z_3 \leq z_4 \leq 0$, the matrix $\mathbf{M} \in \mathbb{R}^{4\times 4}$ is an orthogonal matrix, and $N(\mathbf{Z})$ is the normalization constant. Usually, we enforce the last entry of $\mathbf{Z}$ as a zero value using the property of the Bingham distribution, namely $f(\mathbf{q}_r; \mathbf{M}, \mathbf{Z}) = f(\mathbf{q}_r; \mathbf{M}, \mathbf{Z} + c\mathbf{I})$, by setting $c = -z_4$.

Here we show that the sub-vector $\mathbf{q}_r \in \mathbb{S}^3$ follows the Bingham distribution, the essential matrix $\mathbf{M} \in \mathbb{R}^{4\times 4}$ and $\mathbf{Z} \in \mathbb{R}^{4\times 4}$ can be computed according to the Theorem 2, the proof can be found in the Supplementary Material.

Theorem 2. *The parameter matrix $\mathbf{F} \in \mathbb{R}^{8\times 8}$ is able to be decomposed into an orthogonal matrix $\mathbf{M} \in \mathbb{R}^{4\times 4}$ and a diagonal matrix $\mathbf{Z} \in \mathbb{R}^{4\times 4}$ via the eigendecomposition of $\mathbf{F}_1 - \mathbf{F}_2 \mathbf{F}_3^{-1} \mathbf{F}_2^T$.*

4.1.2 The Gaussian-like Matrix

Similarly, the parameter matrix $\mathbf{F}$ is also critical in the marginal distribution of $\mathbf{q}_d$ which is the Gaussian distribution with the mean vector being $-\mathbf{F}_3^{-1} \mathbf{F}_2^T \mathbf{q}_r$ and covariance matrix being $-\frac{1}{2}\mathbf{F}_3^{-1}$,

$$\begin{aligned} & f(\mathbf{q}_d | \mathbf{q}_r) \propto \\ & \quad \exp\left(\mathbf{q}_d - (-\mathbf{F}_3^{-1} \mathbf{F}_2^T \mathbf{q}_r))^T \mathbf{F}_3 (\mathbf{q}_d - (-\mathbf{F}_3^{-1} \mathbf{F}_2^T \mathbf{q}_r)\right). \end{aligned} \tag{6}$$

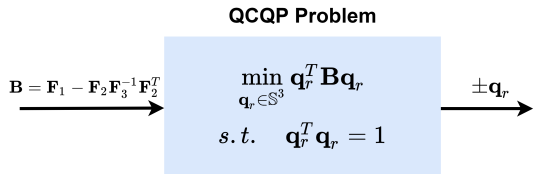

Figure 1: The differentiable QCQP for representing the rotation.

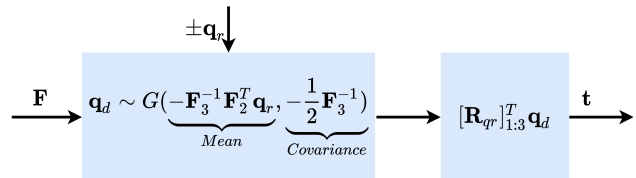

Figure 2: The Gaussian process for representing the translation.

4.2 F AS THE INTERMEDIARY REPRESENTATION SPACE

As $\mathbf{F}$ is an essential element of the developed unit dual quaternion distribution, in this subsection, we show that it can be considered as the intermediary representation space for regressing the pose.

4.2.1 Rotation Representation

For convenience, we set the upper case $\mathbf{B} = \mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T$, which is taken as the representation for rotation quaternions. According to the properties of matrix theory, the matrix $\mathbf{B} \in \mathbb{R}^{4\times4}$ is real symmetric with a simple minimum eigenvalue, which is written as

$$\mathbf{B} = \begin{bmatrix} b_1 & b_2 & b_3 & b_4 \\ b_2 & b_5 & b_6 & b_7 \\ b_3 & b_6 & b_8 & b_9 \\ b_4 & b_7 & b_9 & b_{10} \end{bmatrix}.$$

Then we regard the computation of the rotation quaternion $\mathbf{q}_r$ as an optimization problem which is defined as a quadratically-constrained quadratic program (QCQP) problem that arises in [Yang and Carlone, 2019].

Definition 2 (QCQP problem)**.** *Let matrix $\mathbf{B} \in \mathbb{R}^{4\times4}$ be a symmetric matrix, which can be parameterized with the vector $\mathbf{b} \in \mathbb{R}^{10}$. The QCQP problem related to $\mathbf{B}$ is shown in Figure 1, which is written as*

$$\begin{aligned} &\min_{\mathbf{q}_r \in \mathbb{S}^3} \mathbf{q}_r^T \mathbf{B} \mathbf{q}_r \\ &s.t. \quad \mathbf{q}_r^T \mathbf{q}_r = 1. \end{aligned} \tag{7}$$

Note that the solution to this problem in Figure 1 is to calculate the eigenvector corresponding to the minimum eigenvalue of $\mathbf{B}$.

4.2.2 Translation Representation

The parameter matrix $\mathbf{F}$ also offers a new standpoint for representing the translation vector. As known, the marginal distribution of $\mathbf{q}_d$ is a Gaussian distribution which is shown in Equation (6). Hence, the $\mathbf{q}_d$ can be represented by $\mathbf{m} = -\mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r$, with the uncertainty being measured by the covariance matrix $\mathbf{G} = -\frac{1}{2}\mathbf{F}_3^{-1}$, the related Gaussian process is illustrated in Figure 2. Finally the translation can be recovered from $\mathbf{q}_d$ using Equation (2).

5 DEEP LEARNING AND F

In this section, we intend to develop a new probabilistic dual quaternion network to indirectly regress the pose. First, we show $\mathbf{F}$ is a smooth representation of the pose. Next, we develop a backpropagation method using an implicit function theorem. Subsequently, a new uncertainty metric is proposed to measure uncertainties. Finally, we give the structure of the proposed network.

5.1 SMOOTH FEATURE OF F

A smooth representation for $SE(3)$ is important for learning-based methods when concerning the backpropagation procedure. Here we mainly consider the smooth feature of $\mathbf{F}$ on $SO(3)$ manifold since the translation can be computed after rotation quaternions are estimated.

We utilize the concept of *continuous representation* presented in [Zhou et al., 2019] to give a specific analysis to it. Considering the surjective map between the representation space and original space which is shown in Figure 4, we set the matrix $\mathbf{B}$ as the representation space and the rotation quaternions as the original space. Zhou et al. [2019] demonstrated that the mapping function (f, h) is a representation if f is a left inverse of h. Conversely, the representation is continuous if h is continuous. Moreover, a continuous representation is possible if the dimension of the embedding space is greater than five. In this context, the representation space $\mathbf{B}$ can be simplified with the 10-dimensional vector $\mathbf{b} \in \mathbb{R}^{10}$, then we introduce the *Smooth Global Section Theorem* to show that the representation space is a smooth and continuous mapping to $SO(3)$, where the proof can be found in [Peretroukhin et al., 2020].

Theorem 3 (Smooth Global Section)**.** *Consider the surjective map f: $\mathbb{R}^{10} \to SO(3)$ such that $f(\mathbf{B})$ returns the rotation matrix $\mathbf{R}$ defined by the two antipodal unit quaternions $\pm\mathbf{q}_r$ by minimizing the QCQP problem. There exists a smooth and global mapping, or section, h: $SO(3) \to \mathbb{R}^{10}$ such that $f(h(\mathbf{R})) = \mathbf{R}$.*

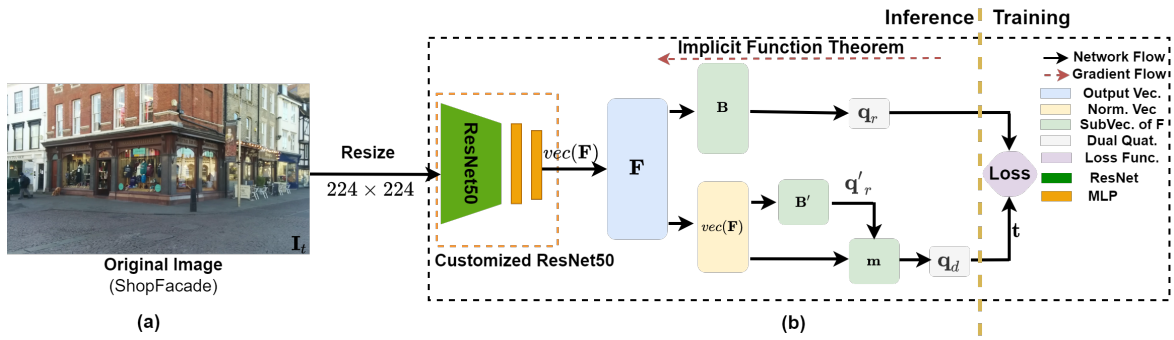

Figure 3: The proposed deep dual quaternion network structure. The letters (a), (b) correspond to the input image, and the proposed deep dual quaternion network. We start from the current frame input I_t which is further fed into the proposed network. The output of the network is a vector $vec(\mathbf{F})$ with 36 elements which then consists of the matrix $\mathbf{F}$. After learning the parameter matrix $\mathbf{F}$, the rotation quaternions $\mathbf{q}_r$ and translation vector $\mathbf{t}$ are computed using the theory of dual quaternion distribution. During training, an implicit function theorem is applied to the backpropagation procedure.

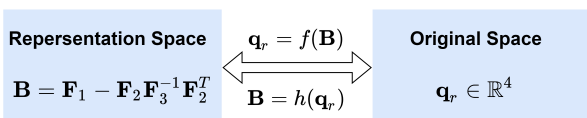

Figure 4: A surjective map between the representation space and the original space.

5.2 GRADIENT COMPUTATION

Regarding the continuousness of the proposed intermediary representation space $\mathbf{F}$, we introduce a gradient computation method for the backpropagation procedure. Recall that the matrix $\mathbf{B} \in \mathbb{R}^{4\times4}$ is real symmetric, which can be simplified with a 10-dimensional vector $\mathbf{b} = vec(\mathbf{B})$. Magnus [1985] demonstrated that $\mathbf{q}_r$ will be differentiable at $\mathbf{b}$ provided that the minimum eigenvalue λ_1 of $\mathbf{B}$ is simple. Hence, the gradient is implemented using the *implicit function theorem*,

$$\frac{\partial \mathbf{q}_r}{\partial \mathbf{b}} = \mathbf{q}_r \otimes (\lambda_1 \mathbf{I} - \mathbf{B})^+, \tag{8}$$

where $\otimes$ denotes the Kronecker product, $(\cdot)^+$ refers to the Moore-Penrose pseudo-inverse.

5.3 UNCERTAINTY MEASUREMENT

The introduction of the uncertainty metric is of great significance to measure pose uncertainties. Intuitively, we consider the proposed unit dual quaternion distribution as the composition of the Bingham distribution $B(\mathbf{q}_r; \mathbf{B})$ and the Gaussian distribution $G(\mathbf{q}_d; -\mathbf{F_3^{-1}F_2^T q_r}, -\frac{1}{2}\mathbf{F}_3^{-1})$.

For the rotation uncertainty, the Bingham belief is a proper choice [Peretroukhin et al., 2020]. According to the property of the Bingham distribution in Equation (5), the rotation uncertainty is written as

$$U_q(\mathbf{Z}) = \sum_i^4 z_i = z_1 + z_2 + z_3 - 3z_4, \tag{9}$$

where $z_i \leq 0, i = 1, 2, 3, 4$.

Likewise, we also use the Gaussian belief to measure the translation uncertainty. Here we decompose the covariance matrix $\mathbf{G} = -\frac{1}{2}\mathbf{F}_3^{-1}$ using the eigendecomposition method, and then the translation uncertainty can be written as

$$U_t(\mathbf{G}) = \sum_i^4 \lambda_i = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4, \tag{10}$$

where $\lambda_i, i = 1, 2, 3, 4$ denotes eigenvalues of $\mathbf{G}$.

5.4 NETWORK STRUCTURE

The structure of the proposed deep probabilistic dual quaternion network is shown in Figure 3. We set the ResNet-50 as the backbone network where the parameters are initialized from pre-trained ImageNet weights. Then two fully connected layer are appended to the ResNet's activations to regress the parameter matrix $\mathbf{F}$. Since the parameter matrix $\mathbf{F}$ is real symmetric, we can encode it with a 36-dimensional feature vector. Subsequently, we decompose the $\mathbf{F}$ with two branches, namely rotation branch and translation branch. For the rotation branch, we estimate rotation quaternions by solving the QCQP problem shown in Figure 1. For the translation branch, we surprisingly find that directly using the estimated rotation quaternions is not sufficient to get a proper translation vector. In this case, a tiny trick is adopted that we normalize the output vector $vec(\mathbf{F})$ to generate a pseudo matrix $\mathbf{B}'$ which is shown in Figure 5. Then another rotation quaternion is estimated to compose the mean vector $\mathbf{m} = -\mathbf{F_3^{-1}F_2^T q_r}$. During the training stage, an implicit

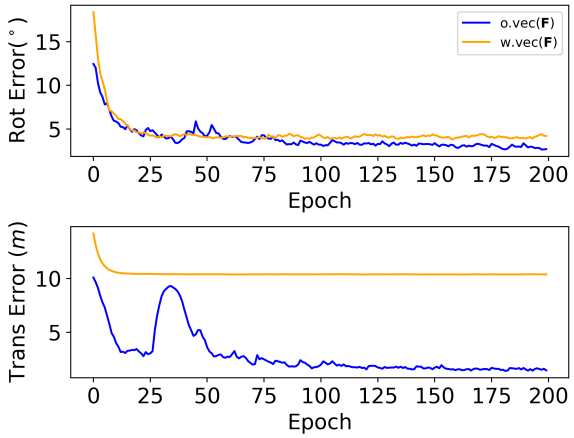

Figure 5: The difference of the use of the normalization of $vec(\mathbf{F})$ during the learning phase. We take the ShopFacade scene in the Cambridge Landmarks dataset for example. The orange line indicates the pose error curves without normalizing the $vec(\mathbf{F})$. While the blue line denotes the pose error curves after normalizing the $vec(\mathbf{F})$. o.$vec(\mathbf{F})$/w.$vec(\mathbf{F})$: without/with normalization of the $vec(\mathbf{F})$.

function theorem is applied to the backpropagation procedure. Finally, the translation vector $\mathbf{t}$ can be recovered from $\mathbf{m}$.

Additionally, we adopt the loss function present in [Kendall et al., 2015] as our loss metric,

$$L = \|\mathbf{t}_e - \mathbf{t}_g\|_2 + \alpha\|\mathbf{q}_e - \mathbf{q}_g\|_2, \tag{11}$$

where $\mathbf{t}_e$ and $\mathbf{q}_e$ denote the inferred translation and rotation, while $\mathbf{t}_g$ and $\mathbf{q}_g$ are the labeled translation and rotation. Moreover, we set the scale factor $\alpha = 100$ on the 7-Scenes dataset and $\alpha = 300$ on the Cambridge Landmarks dataset.

6 EXPERIMENTS

In this section, we perform the absolute pose regression on the task of camera re-localization on two public datasets, namely Cambridge Landmarks [Kendall et al., 2015] and 7-Scenes datasets [Shotton et al., 2013], which consists of RGB frames with associated ground truth camera poses and provides training as well as test sequences. First, we give a comparison with state-of-the-art pose regression methods to show the inferred pose accuracy[1].Then we evaluate our deep probabilistic model on the noisy Cambridge Landmarks dataset to show its robustness in dealing with uncertainty and ambiguity.

[1]The final results are measured by the median rotation error and the translation error.

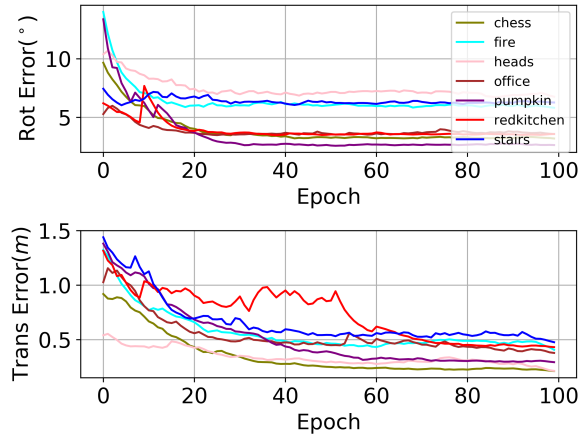

Figure 6: The rotation and translation error curves of each scene during test stages on the 7-Scenes dataset.

6.1 TRAINING DETAILS

We run our experiments in the Pytorch framework [Paszke et al., 2019]. We use the Adam optimizer [Kingma and Ba, 2015] and begin with a learning rate of 10^{-4}, and gradually decrease the learning rate exponentially with the multiplicative factor being 0.9. We use a batch size of 16 and train for 100 epochs for the 7-Scenes dataset and 200 epochs for the Cambridge Landmarks dataset. All input frames are resized to 224×224.

6.2 RESULTS

6.2.1 Normal Scenes

7-Scenes Dataset. We test our model on all 7 scenes on the 7-Scenes dataset. Since the majority of the scenes do not show highly ambiguous environments, we regard them to be non-ambiguous. The final test curves of 7 scenes can be found in Figure 6. Clearly, the overall rotation error is less than 7° and the overall translation error is less than 0.5m.

In order to demonstrate the pose accuracy of our model, we make a comparison with other pose regression methods including PoseNet and its variant Dense PoseNet[Kendall et al., 2015], MapNet and its variant MapNet++[Sattler et al., 2019], BPN[Brahmbhatt et al., 2018], VidLoc[Clark et al., 2017], UBN and MBN-MB[Deng et al., 2020]. The quantitative and qualitative results are listed in Table 1.

From Table 1, the evaluated results on the 7-Scenes dataset show that our method outperforms state-of-the-art methods on the rotation accuracy. On the other hand, our method also achieves competitive results in terms of the translation accuracy.

Cambridge Landmarks Dataset. To further demonstrate the pose accuracy on different scenes, we also implement

Table 1: Evaluation on the 7-Scenes dataset. The results are reported with the median translation error(m) and the median rotation error(°).

Scene	Chess	Fire	Heads	Office	Pumpkin	RedKitchen	Stairs
PoseNet	0.32m/8.12°	0.47m/14.4°	0.29m/12.0°	0.48m/7.68°	0.47m/8.42°	0.59m/8.64°	0.47m/13.8°
Dense PoseNet	0.32m/6.60°	0.47m/14.0°	0.30m/12.2°	0.48m/7.24°	0.49m/8.12°	0.58m/8.34°	0.48m/13.1°
MapNet	0.08m/3.25°	0.27m/11.69°	0.18m/13.2°	0.17m/5.15°	0.22m/4.02°	0.23m/4.93°	0.30m/12.08°
MapNet++	0.10m/3.17°	0.20m/9.04°	0.13m/11.1°	0.18m/5.38°	0.19m/3.92°	0.20m/5.01°	0.30m/13.4°
BPN	0.37m/7.24°	0.43m/13.7°	0.31m/12.0°	0.48m/8.04°	0.61m/7.54°	0.58m/7.54°	0.48m/13.1°
VidLoc	0.18m/-	0.26m/-	0.14m/-	0.26m/-	0.36m/-	0.31m/-	0.26m/-
UBN	0.10m/4.97°	0.27m/12.87°	0.12m/14.05°	0.20m/7.52°	0.23m/7.11°	0.19m/8.25°	0.28m/13.1°
MBN-MB	0.10m/4.35°	0.28m/11.86°	0.12m/12.76°	0.19m/6.55°	0.22m/6.9°	0.21m/8.08°	0.31m/9.98°
Ours	0.20m/2.9°	0.30m/5.63°	0.19m/6.53°	0.30m/3.51°	0.28m/2.6°	0.40m/3.6°	0.42m/6.23°

Table 2: Evaluation on the Cambridge Landmarks dataset. The results are reported with the median translation error(m) and the median rotation error(°).

Scene	Kings College	Hospital	ShopFacade	St.Mary Church
PoseNet	1.92m/5.40°	2.31m/5.38°	1.46m/8.08°	2.65m/8.48°
Dense PoseNet	1.66m/4.86°	2.62m/4.90°	1.41m/7.18°	2.45m/7.96°
MapNet	1.07m/1.89°	1.94m/3.91°	1.49m/4.22°	2.0m/4.53°
BPN	1.74m/4.06°	2.57m/5.12°	1.25/7.54°	2.11m/8.38°
UBN	0.88/1.77°	1.93m/3.71°	0.8m/4.74°	1.84m/6.19°
MBN-MB	0.83m/2.08°	2.16m/3.64°	0.92m/4.93°	1.37m/6.03°
Ours	1.20m/0.84°	2.46m/1.72°	1.10m/2.51°	2.40m/2.63°

our approach on the Cambridge Landmarks dataset. We select the Kings College, Hospital, ShopFacade and St.Mary Church as our evaluation scenes. Again, we plot the four different test curves in Figure 7. The final converge results show that the overall rotation error is less than 3° and overall translation error is no more than about 3m.

Next, we also list our final pose accuracy in Table 2 to make a comparison with state-of-the-art methods. Similar pose accuracy is reported that our method can achieve a more accurate pose in the Cambridge Landmarks dataset.

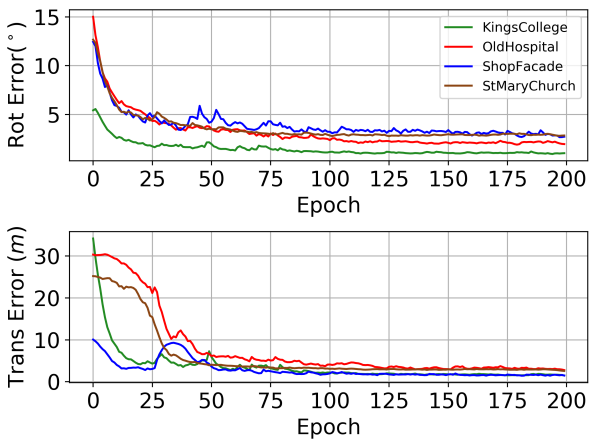

Figure 7: The rotation and translation error curves of each scene during test stages on the Cambridge Landmarks dataset.

6.2.2 Noisy Scenes

To further demonstrate the performance of our method in dealing with uncertainty and ambiguity, we conduct our model on four noisy scenarios, namely the Kings College, Hospital, ShopFacade, and St.Mary Church, where these scenes are processed by manually adding the Gaussian blur kernel, randomly changing the brightness, contrast, saturation of all frames, and both to simulate different ambiguous environments.

Without retraining the proposed deep probabilistic dual quaternion network, we directly feed the processed frames into the trained model to predict the camera pose. The quantitative results are shown in Table 3. The results show that the pose errors have some minor changes in noisy scenes including the Kings College, Hospital, ShopFacade and St.Mary Church.

Then we measure uncertainties of our model under normal and noisy environments, which is shown in Figure 8. Intuitively, the pose errors in the blur environment denoted by purple points have a similar distribution compared to that in the original environment denoted by the red points under the both rotation uncertainty and translation uncertainty measurement. Nevertheless, the pose errors in the brightness change environment denoted by blue points have some minor differences especially for the rotation uncertainty. Likewise, the pose errors in the blur and brightness change environment denoted by the orange points have a similar tendency, but we believe that it is reasonable in noisy

Table 3: The quantitative results of the proposed deep probabilistic dual quaternion network in the noisy scenes on the Cambridge Landmarks dataset.

Scene	Kings College	Hospital	ShopFacade	St.Mary Church
Normal	1.20m/0.84°	2.46m/1.72°	1.10m/2.51°	2.4m/2.63°
Blur	1.42/0.92°	2.59m/1.74°	1.12m/2.53°	2.60m/2.65°
Brightness	1.55m/1.00°	3.07m/2.14°	1.30m/2.61°	3.00m/2.60°
Blur & Brightness	1.76m/1.22°	3.29m/2.31°	1.38m/2.86°	3.13m/2.69°

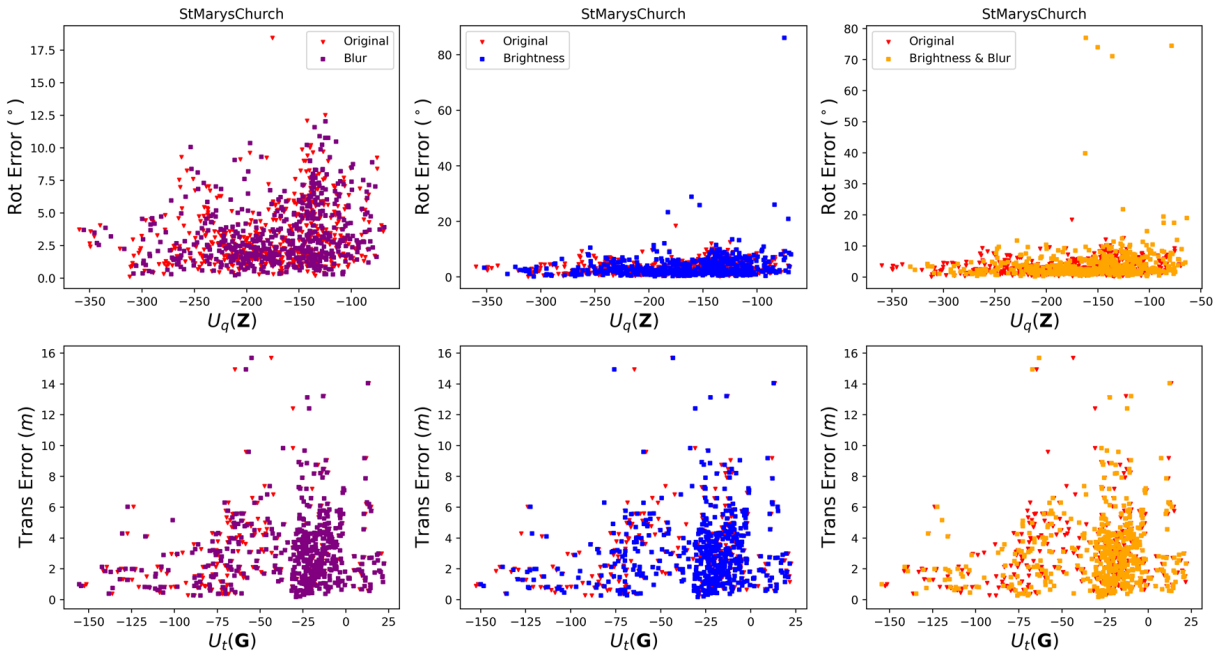

Figure 8: Uncertainty evaluation on the Cambridge Landmarks dataset. The left column shows the pose errors under the pose uncertainty metric in the blur environment, where the radius of the Gaussian blur kernel is 3.8. The middle column shows the pose errors under the pose uncertainty metric in the random brightness change environment, where the maximum brightness factor is 0.6, the maximum contrast factor is 0.6 and the maximum saturation factor is 0.5. The right column indicates pose errors under the both blur and brightness change environment. Note: we only plot the St.Mary Church scene, full information can be found in the Supplementary Material.

environments. More importantly, there are only a few points out of the original distribution in aforementioned two scenes which have limited effects to the overall pose accuracy. As a result, the experiment results in noisy environments suggest that our model is robust to deal with uncertainty and ambiguity.

7 CONCLUSION

We design a deep probabilistic dual quaternion network that addresses the absolute pose regression problem on $SE(3)$. Unlike existing work, we take pose uncertainties into consideration by introducing an antipodally symmetric distribution over the unit dual quaternion. To address the mapping problem from the Euclidean space $\mathbb{R}^n$ to $SE(3)$ manifold, we present an intermediary differential representation space $\mathbf{F}$ as the output of our model to indirectly regress poses. Additionally, we introduce a backpropagation method for batch optimization. Experiment results on the camera relocalization task on the 7-Scenes dataset and the Cambridge Landmarks dataset show that our method outperforms state-of-the-art methods on the pose accuracy. Moreover, extensive experiments on the noisy scenes on the Cambridge Landmarks dataset show that our method has the ability to deal with uncertainty and ambiguity.

In the future, we will explore the pose regression problem leveraging our representation with a negative log-likelihood loss function to improve the reliability and robustness of our model in real-world applications.

References

Christopher Bingham. An antipodally symmetric distribution on the sphere. *Annals of Statistics*, 2:1201–1225, 1974.

Samarth Brahmbhatt, Jinwei Gu, Kihwan Kim, James Hays, and Jan Kautz. Geometry-aware learning of maps for camera localization. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2616–2625, 2018.

Mai Bui, Tolga Birdal, Haowen Deng, Shadi Albarqouni, Leonidas Guibas, Slobodan Ilic, and Nassir Navab. 6d camera relocalization in ambiguous scenes via continuous multimodal inference. In *European Conference on Computer Vision*, pages 139–157. Springer, 2020.

Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE Transactions on Robotics*, 37 (6):1874–1890, 2021.

Kefan Chen, Noah Snavely, and Ameesh Makadia. Wide-baseline relative camera pose estimation with directional learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3257–3267, 2021.

Ronald Clark, Sen Wang, Andrew Markham, Niki Trigoni, and Hongkai Wen. Vidloc: A deep spatio-temporal model for 6-dof video-clip relocalization. In *2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6856–6864, 2017.

Haowen Deng, Mai Bui, Nassir Navab, Leonidas Guibas, Slobodan Ilic, and Tolga Birdal. Deep bingham networks: Dealing with uncertainty and ambiguity in pose estimation. *arXiv preprint arXiv:2012.11002*, 2020.

Xinke Deng, Arsalan Mousavian, Yu Xiang, Fei Xia, Timothy Bretl, and Dieter Fox. Poserbpf: A rao-blackwellized particle filter for 6d object pose estimation. *Robotics: Science and Systems XV*, 2019.

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*, 2020.

Alex Kendall and Roberto Cipolla. Modelling uncertainty in deep learning for camera relocalization. In *2016 IEEE International Conference on Robotics and Automation*, pages 4762–4769, 2016.

Alex Kendall, Matthew Koichi Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. *2015 IEEE International Conference on Computer Vision*, pages 2938–2946, 2015.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations*, 2015.

Steven M. LaValle. *The Configuration Space*, page 105–152. Cambridge University Press, 2006.

Kailai Li, Florian Pfaff, and Uwe D. Hanebeck. Unscented dual quaternion particle filter for se(3) estimation. *IEEE Control Systems Letters*, 5:647–652, 2021.

Jan R. Magnus. On differentiating eigenvalues and eigenvectors. *Journal of Optimization Theory and Applications*, 1985.

David Mohlin, Josephine Sullivan, and Gérald Bianchi. Probabilistic orientation estimation with matrix fisher distributions. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4884–4893. Curran Associates, Inc., 2020.

Raúl Mur-Artal and Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5): 1255–1262, 2017.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Valentin Peretroukhin, Matthew Giamou, David M. Rosen, W. Nicholas Greene, Nicholas Roy, and Jonathan Kelly. A Smooth Representation of SO(3) for Deep Rotation Learning with Uncertainty. In *Proceedings of Robotics: Science and Systems*, Jul. 12–16 2020.

Omid Poursaeed, Guandao Yang, Aditya Prakash, Qi Fang, Hanqing Jiang, Bharath Hariharan, and Serge J. Belongie. Deep fundamental matrix estimation without correspondences. In *ECCV Workshops*, 2018.

Torsten Sattler, Qunjie Zhou, Marc Pollefeys, and Laura Leal-Taixe. Understanding the limitations of cnn-based absolute camera pose regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3302–3312, 2019.

Jamie Shotton, Ben Glocker, Christopher Zach, Shahram Izadi, Antonio Criminisi, and Andrew W. Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2930–2937, 2013.

Sen Wang, Ronald Clark, Hongkai Wen, and Agathoniki Trigoni. End-to-end, sequence-to-sequence probabilistic visual odometry through deep neural networks. *The International Journal of Robotics Research*, 37:513 – 542, 2018.

Jian Wu, Liwei Ma, and Xiaolin Hu. Delving deeper into convolutional neural networks for camera relocalization. In *2017 IEEE International Conference on Robotics and Automation*, pages 5644–5651, 2017.

Fei Xue, Xin Wang, Shunkai Li, Qiuyuan Wang, Junqiu Wang, and Hongbin Zha. Beyond tracking: Selecting memory and refining poses for deep visual odometry. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8567–8575, 2019.

Heng Yang and Luca Carlone. A quaternion-based certifiably optimal solution to the wahba problem with outliers. *2019 IEEE/CVF International Conference on Computer Vision*, pages 1665–1674, 2019.

Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5738–5746, 2019.

PDQ-Net: Deep Probabilistic Dual Quaternion Network for Absolute Pose Regression on $SE(3)$ (Supplementary Material)

In this part, we mainly give the supplementary material for the original paper. First, we show the proofs of several features of the unit dual quaternion distribution on $SE(3)$. Then we give the supplementary analysis of the proposed model on different noisy scenes in the Cambridge Landmarks dataset to demonstrate the robustness of our method.

A PROOF

Theorem 1. *Considering the antipodally symmetric distribution $f(\mathbf{v})$, the sub-block matrix $\mathbf{F}_1 \in \mathbb{R}^{4\times4}$ is real symmetric, $\mathbf{F}_3 \in \mathbb{R}^{4\times4}$ is real symmetric and negative definite.*

$$f(\mathbf{v}) = \frac{1}{N(\mathbf{F})}\exp\underbrace{(\mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r}_{Bingham-like} + \underbrace{(\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)^T\mathbf{F}_3(\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r))}_{Gaussian-like}. \tag{1}$$

Proof. We expand the original unit dual quaternion probability density function as follows,

$$f(\mathbf{v}) = \frac{1}{N(\mathbf{F})}\exp\underbrace{\left(\begin{bmatrix}\mathbf{q}_r\\ \mathbf{q}_d\end{bmatrix}^T\begin{bmatrix}\mathbf{F}_1 & \mathbf{F}_2\\ \mathbf{F}_2^T & \mathbf{F}_3\end{bmatrix}\begin{bmatrix}\mathbf{q}_r\\ \mathbf{q}_d\end{bmatrix}\right)}_{A}. \tag{2}$$

We denote the exponential part as A, then we expand it and have

$$\begin{aligned} A &= \left(\begin{bmatrix}\mathbf{q}_r\\ \mathbf{q}_d\end{bmatrix}^T\begin{bmatrix}\mathbf{F}_1 & \mathbf{F}_2\\ \mathbf{F}_2^T & \mathbf{F}_3\end{bmatrix}\begin{bmatrix}\mathbf{q}_r\\ \mathbf{q}_d\end{bmatrix}\right)\\ &= \mathbf{q}_r^T\mathbf{F}_1\mathbf{q}_r + \mathbf{q}_d^T\mathbf{F}_2^T\mathbf{q}_r + \mathbf{q}_r^T\mathbf{F}_2\mathbf{q}_d + \mathbf{q}_d^T\mathbf{F}_3\mathbf{q}_d. \end{aligned} \tag{3}$$

Clearly, the sub-matrices $\mathbf{F}_1$ and $\mathbf{F}_3$ have to be real symmetric since they are critical for keeping the antipodally symmetric feature of the target probability density function. Then we apply a tiny trick to Equation (3) and we have

$$\begin{aligned} A =& \mathbf{q}_r^T\mathbf{F}_1\mathbf{q}_r - \mathbf{q}_r^T\mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r + \mathbf{q}_d^T\mathbf{F}_3\mathbf{q}_d\\ &+ \mathbf{q}_d^T\mathbf{F}_3\mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r + \mathbf{q}_r^T\mathbf{F}_2\mathbf{F}_3^{-T}\mathbf{F}_3\mathbf{q}_d\\ &+ \mathbf{q}_r^T\mathbf{F}_2\mathbf{F}_3^{-T}\mathbf{F}_3\mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r\\ =& \mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r\\ &+ (\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)^T\mathbf{F}_3(\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r). \end{aligned} \tag{4}$$

Next, we take an integration of $\exp(A)$ over the unit dual quaternion manifold $\mathbb{DH}_1 \subset \mathbb{R}^8$,

$$\begin{aligned} N(\mathbf{F}) &= \int_{\mathbb{DH}_1} f(\mathbf{v})d\mathbf{v}\\ &= \int_{\mathbb{S}^3}\int_{\mathbb{R}^4} \mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r\\ &\quad + (\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)^T\mathbf{F}_3(\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)d\mathbf{q}_d d\mathbf{q}_r. \end{aligned} \tag{5}$$

We find that the inner integration corresponds to the unnormalized Gaussian density function, then we have

$$N(\mathbf{F}) \propto \int_{\mathbb{S}^3} \mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r d\mathbf{q}_r. \tag{6}$$

Hence, the matrix $-\frac{1}{2}\mathbf{F}_3^{-1}$ can be regarded as the covariance matrix of the Gaussian distribution, in which $\mathbf{F}_3$ is negative definite while $\mathbf{F}_2$ is arbitrary. □

Theorem 2. *The parameter matrix $\mathbf{F} \in \mathbb{R}^{8\times8}$ is able to be decomposed into an orthogonal matrix $\mathbf{M} \in \mathbb{R}^{4\times4}$ and a diagonal matrix $\mathbf{Z} \in \mathbb{R}^{4\times4}$ via the eigendecomposition of $\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T$.*

Proof. As shown in Equation (5), the marginal distribution of $\mathbf{q}_d$ is the unnormalized Gaussian distribution. Here we

take an integration to it.

$$
\begin{aligned}
f(\mathbf{q}_r) \propto \exp(\int_{\mathbb{R}^4} & \mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r \\
& + (\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)^T\mathbf{F}_3(\mathbf{q}_d + \mathbf{F}_3^{-1}\mathbf{F}_2^T\mathbf{q}_r)d\mathbf{q}_d) \\
& \propto \exp(\mathbf{q}_r^T(\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T)\mathbf{q}_r).
\end{aligned} \tag{7}
$$

To this point, Equation (7) is a Bingham-like distribution, in which $\mathbf{q}_r \in \mathbb{S}^3$ is the unit quaternion, then

$$\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T = \mathbf{M}\mathbf{Z}\mathbf{M}^T.$$

As a result, the orthogonal matrix $\mathbf{M}$ and the diagonal matrix $\mathbf{Z}$ can be obtained via the eigendecomposition of $\mathbf{F}_1 - \mathbf{F}_2\mathbf{F}_3^{-1}\mathbf{F}_2^T$ □

B EXPERIMENT

Noisy Scenes We train our model on the original Cambridge Landmarks dataset for 200 epochs. We use the Adam optimizer and begin with a learning rate of 10^{-4}, and gradually decrease the learning rate exponentially with the multiplicative factor being 0.9, where the learning curve is shown in Figure 1. The batch size is 16 and all input images are resized to 224×224.

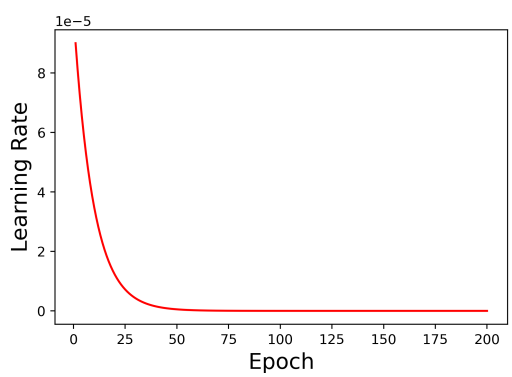

Figure 1: The learning curve for the Cambridge Landmarks dataset.

Then we apply the trained model to four scenes with three different noises. We select the Kings College, Hospital, ShopFacade and St.Mary Church as our evaluation scenes, which is shown in Figure 2(a). First, we manually add the Gaussian blur kernel to all frames in the above four different scenes, where the radius of the Gaussian blur kernel is 3.8 shown in Figure 2(b). Then we randomly change the brightness, the contrast, and the saturation on the second noisy scene which can be found in Figure 2(c), we set the maximum brightness factor is 0.6, the maximum contrast factor is 0.6 and the maximum saturation factor is 0.5 in this case. Finally, we add the above two noises, i.e. blur kernel and random brightness, contrast, and saturation, to the third noisy scene which can be found in Figure 2(d).

Next, we feed noisy frames into the trained model and the uncertainties of our model in all scenes are shown in Figure 3, where the red points are pose errors under the uncertainty of the original scene, the purple points are pose errors in the blur scene, the blue points refer to pose errors in the brightness change scene and the orange points are pose errors in the blur and brightness change environment.

Figure 2: Visualization of four different scenes under different noise conditions. The letter (a) refers to the original scenes. The letters (b),(c),(d) correspond to the different noise added to the original scene. And the four different scenes, Kings College, Hospital, ShopFacade and St.Mary Church, can be found from the first row to the last row.

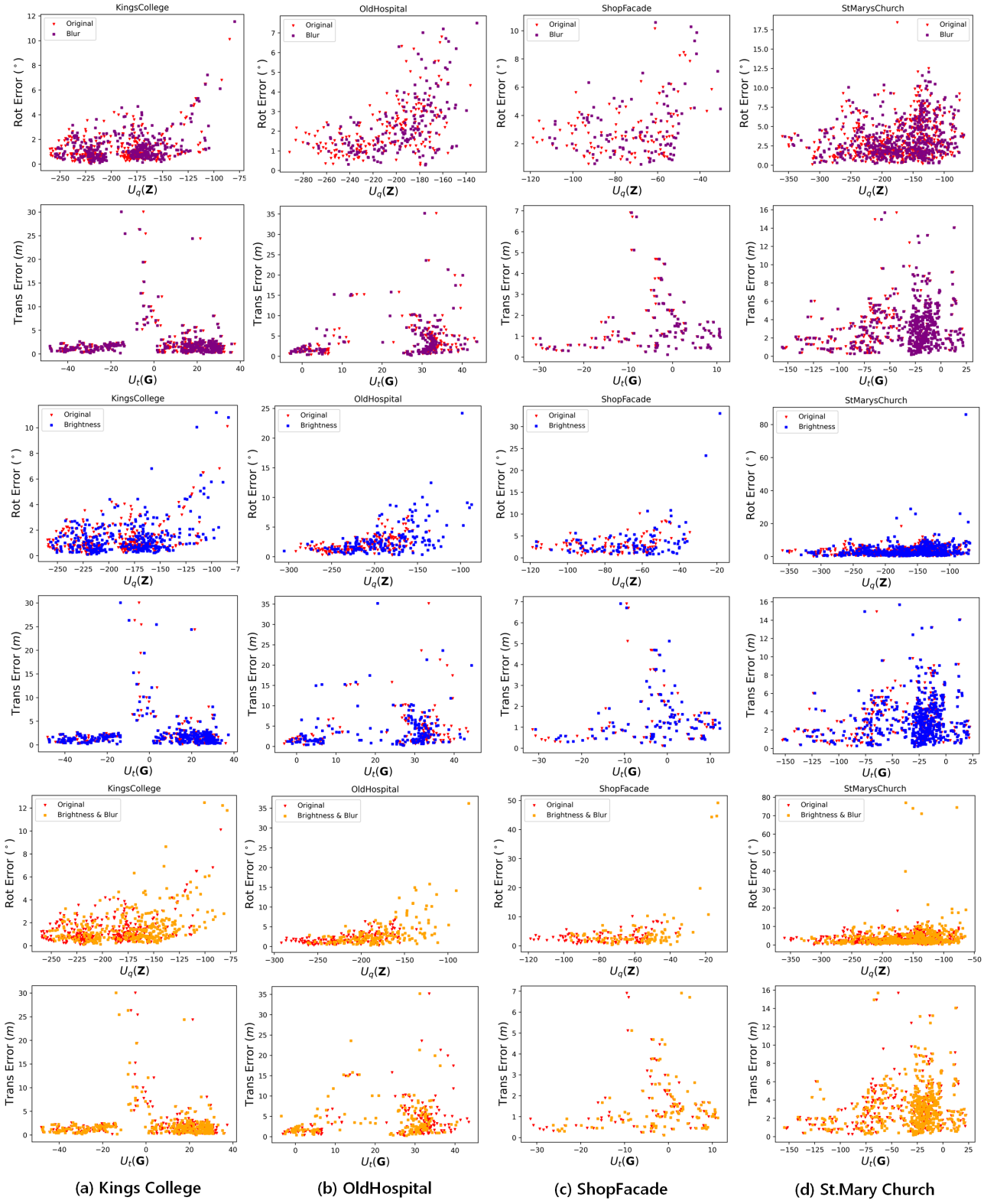

Figure 3: Uncertainty evaluation on the Cambridge Landmarks dataset. The letters (a), (b),(c),(d) correspond to the different scenes of the Cambridge Landmarks dataset. The pose errors under different uncertainty metrics of the model in each scene are shown in the corresponding row, where the odd rows show the rotation uncertainty of the corresponding noisy scene and the even rows show the translation uncertainty of the corresponding noisy scene.