Deep Pose Estimation for Image-Based Registration

Benjamin Hou[1], **Nina Miolane**[2], **Bishesh Khanal**[1,3], **Matthew C.H. Lee**[1,4], **Amir Alansary**[1], **Steven McDonagh**[1], **Joseph Hajnal**[3], **Daniel Rueckert**[1], **Ben Glocker**[1], **Bernhard Kainz**[1]
[1]Imperial College London, [2]INRIA & Stanford, [3]King's College London, [4]HeartFlow
bh1511@imperial.ac.uk

Abstract

Pose estimation is an omnipresent problem in medical image analysis. Deep learning methods often parameterise a pose with a representation that separates rotation and translation, as commonly available frameworks do not provide means to calculate loss on a manifold. In this paper, we propose a general Riemannian formulation of the pose estimation problem and train CNNs directly on $SE(3)$ equipped with a left-invariant Riemannian metric. At each training step; the loss is calculated as the squared Riemannian geodesic distance, with the gradients required for back-propagation calculated with respect to the predicted pose $\hat{p}$ on the tangent space of the manifold $SE(3)$ at $\hat{p}$. We thoroughly evaluate the effectiveness of our loss function by comparing its performance with popular and most commonly used existing methods, and show that it can improve registration accuracy for image-based 2D to 3D registration.

1 Introduction

Intensity-based registration and landmark matching are the de-facto standards to align data from multiple image sources into a common co-ordinate system. Applications that require intensity-based registration include *e.g.*, atlas-based segmentation, motion-compensation, tracking, or clinical analysis of the data visualised in a standard co-ordinate system. A pose, i.e. a rigid transformation in 3D, is an element of the Lie group $SE(3)$, the Special Euclidean group in 3D, and has two components; a rotation component of group $SO(3)$ and a translation component of $\mathbb{R}^3$. $SE(3)$ has the following convenient matrix representation (called the homogeneous representation):

$$SE(3) = \left\{ X \quad | \quad X = \left[\begin{array}{c|c} R & t \\ \hline 0 & 1 \end{array} \right], t \in \mathbb{R}^3, R \in SO(3) \right\} \tag{1}$$

In usual implementations of $SE(3)$, the rotation can be parameterised in any form as long as the $SO(3)$ group structure is implicitly imposed. The rotation can be stored as Euler angles, as quaternion, as axis-angle or as rotation matrix. This needs to be considered carefully, especially when designing deep learning applications, as the numerical properties of each parameterisation can hamper efficacy.

Popular deep learning frameworks available today, such as Caffe, TensorFlow, Theano, PyTorch, do not provide the means to regress on $SE(3)$, as the common loss metrics provided are cross-entropy for probabilities or a p-norms for distances. In literature, the $SE(3)$ pose has being parameterised in many different forms, with most methods utilising the L2-norm as the loss metric. E.g., Kendall et al. [4] uses the L2-norm to regress parameters on the Lie algebra $\mathfrak{se}(3)$ directly, with a β parameter to weight the contribution between rotation and translation. Methods that do not couple together the rotation and translation parameters neglect the intrinsic structure of $SE(3) = SO(3) \ltimes \mathbb{R}^3$, which can lead to unpredictable behaviours. As for the non-linear structure of $SO(3)$, this can be observed visually with quaternions, e.g., the Euclidean distance of two quaternions can be small, despite the rotation being large. Hence, it is desirable to have a loss function that respects the structure and non-linearity of $SE(3)$ as a Lie group, and thus as a manifold.

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

2 Method

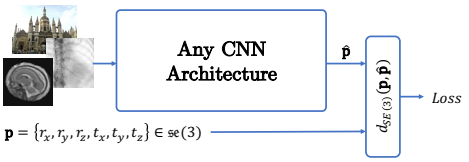

Figure 1: CNN architecture using a Riemannian geodesic distance as the loss on $SE(3)$.

The core of our method is to implement a new loss layer: we define the loss as the squared geodesic distance on $SE(3)$ equipped with a left-invariant Riemannian metric, Figure 1. The network architecture is therefore structure agnostic, as long as the regressor head outputs a vector of six values.

Left-invariant Riemannian metric on $SE(3)$: A Riemannian metric on $SE(3)$ is a smooth collection of positive definite inner products on each tangent space of $SE(3)$. Then, $SE(3)$ becomes a Riemannian manifold. With a left-invariant metric, it is enough to define an inner product on the tangent space at the identity of $SE(3)$, and then "propagate" it: the metric is s.t. $\forall u, v \in T_{p_1}SE(3)$ and $\forall p_1, p_2 \in SE(3)$: $< DL_{p_1}(p_2)u, DL_{p_1}(p_2)v > |_{L_{p_1}p_2} =< u, v > |_{p_2}$ where L_{p_1} is the left translation by p_1: $L_{p_1}(p_2) = p_1 \circ p_2$, and $DL_{p_1}(p_2)$ its differential at p_2. Defining an inner product Z at $p_2 =$ identity enables us to get a metric Z_{p_1} at the tangent space of any pose p_1 of $SE(3)$ [6], and thus to compute inner products and norms of tangent vectors at p_1.

Loss and gradient: We use the loss function: $\text{loss}(\mathbf{p}, \hat{\mathbf{p}}) = \text{dist}^Z_{SE(3)}(\mathbf{p}, \hat{\mathbf{p}})^2 = \left\|\text{Log}^Z_{\hat{\mathbf{p}}}(\mathbf{p})\right\|^2_{Z_{\hat{\mathbf{p}}}}$ where $\text{dist}^Z_{SE(3)}$ is the geodesic distance and $\text{Log}_{\hat{\mathbf{p}}}$ is the Riemannian logarithm at $\hat{\mathbf{p}}$ i.e. a tangent vector at $\hat{\mathbf{p}}$. We use a left-invariant Riemannian metric, thus: $\text{loss}(\mathbf{p}, \hat{\mathbf{p}}) = ||DL_{\hat{\mathbf{p}}^{-1}}.\text{Log}^Z_{\hat{\mathbf{p}}}(\mathbf{p})||^2_Z$, where we now have a tangent vector at the identity and we can use the inner product Z to compute its squared norm. If we take Z being the canonical inner product at identity, this is the L2-norm but on the tangent vector transported from $\hat{\mathbf{p}}$ to identity using the differential $DL_{\hat{\mathbf{p}}^{-1}}$. The backward gradient corresponding to the loss seen as a function of $\hat{\mathbf{p}}$ is $\nabla_{\hat{\mathbf{p}}}\text{loss}(\mathbf{p}, \hat{\mathbf{p}}) = -2 \cdot \text{Log}^Z_{\hat{\mathbf{p}}}(\mathbf{p})$ [7] which is a tangent vector at $\hat{\mathbf{p}}$.

Implementation: The inputs to the loss layer are the poses $\mathbf{p}$ and $\hat{\mathbf{p}}$ for ground truth and prediction respectively. We represent a pose with `geomstats` implementation [5] *i.e.* as the Riemannian Logarithm of canonical left-invariant metric on $SE(3)$ s.t. $p = \{r, t\} = \{r_x, r_y, r_z, t_x, t_y, t_z\} \in \mathbb{R}^6$. With this parameterisation, the rotation r is in axis-angle parameterisation, the inner product Z is a 6x6 positive definite matrix and the differential $DL_{\hat{p}}$ of the left translation is the 6x6 jacobian matrix: $J_{\hat{p}} = \begin{pmatrix} \frac{\partial L_{\hat{p}}{}^r}{\partial r} & \frac{\partial L_{\hat{p}}{}^r}{\partial t} \\ \frac{\partial L_{\hat{p}}{}^t}{\partial r} & \frac{\partial L_{\hat{p}}{}^t}{\partial t} \end{pmatrix}$. We denote $v_t = \text{Log}^Z_{\hat{\mathbf{p}}}(\mathbf{p})$ which is a tangent vector at $\hat{\mathbf{p}}$ in this parameterisation. The loss is calculated by $\text{loss}(\mathbf{p}, \hat{\mathbf{p}}) = v_t^T * J^T_{\hat{p}^{-1}} * Z * J_{\hat{p}^{-1}} * v_t$ where $*$ is the matrix multiplication and the Riemannian logarithm v_t is given by `geomstats`. The gradient is calculated by: $\nabla_{\hat{\mathbf{p}}}\text{loss}(\mathbf{p}, \hat{\mathbf{p}}) = -2 * J^T_{\hat{p}^{-1}} * Z * J_{\hat{p}^{-1}} * v_t$.

3 Experiments and Results

We evaluate our novel loss function on two existing datasets: **(Exp1)** the common C-Arm X-Ray to Computed Tomography (CT) alignment problem with data from [2]. **(Exp2)**, the pose estimation dataset for motion compensation in fetal Magnetic Resonance Imaging (MRI) from [1]. In each experiment, we benchmark existing $SE(3)$ parameterisation strategies with the respective loss function used. PoseNet: direct regression of parameters on the Lie algebra $\mathfrak{se}(3)$ using L2-norms. Anchor Points: a re-parameterisation of $SE(3)$ in Euclidean space, where three statically defined points in 3D space defines a plane. Each Anchor Point is regressed independently using the L2-norm. Finally, our $SE(3)$ loss function, i.e., the geodesic distance on the Riemannian manifold.

Exp1: It can be seen that the average error for Euler parameters and translation parameters (for both healthy and pathological patients) are similar to each other, and shown insignificant by Student's t-test. However, there is a noticeable trend in average geodesic distance errors. Student's t-test showed significant difference between $SE(3)$ loss compared to PoseNet and Anchor Points for both datasets (marked by *). This shows that the geodesic metric is able to quantify properties that the metric expressed in Euler-translation parameters cannot.

Table 1: Mean Error of Loss Functions on DRR (Digitally Reconstructed Radiographs)

	R_x	R_y	R_z	t_x	t_y	t_z	G.D.
PoseNet	7.960	3.136	7.547	62.650	57.315	45.852	15201.845
Anchor Points	**7.274**	**2.511**	**7.059**	59.292	**54.889**	**40.576**	15115.858
$SE(3)$	8.243	3.697	7.924	**58.647**	55.477	44.189	**14170.722***
Healthy Patient Dataset							
	R_x	R_y	R_z	t_x	t_y	t_z	G.D.
PoseNet	10.653	5.788	10.760	69.107	72.238	57.726	23495.708
Anchor Points	**8.540**	**4.060**	**8.553**	65.521	**68.543**	54.133	21725.921
$SE(3)$	10.511	6.789	11.913	**62.588**	68.747	**54.110**	**19624.246***
Pathological Patient Dataset							

Exp2: In this experiment, we replicate the experiment and evaluation method from [1]. We evaluated our loss regressor for 2D/3D registration used during motion compensation of fetal MRI data in canonical organ space.

Table 2: Mean Error of Loss Functions on Fetal Brain Images

	CC	MSE	PSNR	SSIM	G.D.
PoseNet	0.8199	1046.4	18.6509	0.5448	18.1708
Anchor Points	0.8378	935.0	19.3564	0.5845	15.7504
$SE(3)$	**0.8732***	**724.9713***	**20.7484***	**0.6470***	**10.0836***

Our $SE(3)$ loss function shows drastic improvement in all image similarity metrics. This is confirmed by Student's t-test which shows significant difference, and is crucial for Slice-to-Volume applications as the metric for slice alignment is derived from the metrics used above [3].

Discussion and Conclusion A pose is a combination of rotation and translation, therefore it seems reasonable that a CNN predicting a pose should use a metric that accounts for both of them simultaneously. However, one should compare metrics with a target application. Metrics are perceptually a method of measurement with its own set of rules, *e.g.*, imperial vs. metric system for quantifying distances. Choosing a metric for a target application is not always straight forward and often a question of required precision, *e.g.*, one would not measure the diameter of a pinhead with a meter rule, nor measure distance between cities with a caliper.

We have shown that our loss function, using a Riemannian geodesic distance on $SE(3)$ is better suited for medical registration tasks as shown in both experiments. Exp1 shows each test case yielding no significant difference on Euler and translation parameters, with significant difference on geodesic parameters. This suggests that Euler-translation parameters separately are not able to fully quantify the properties of $SE(3)$. In Exp2, our loss function was able to significantly improve the image similarity metrics, as used by Slice-to-Volume motion compensation algorithms.

References

[1] Hou, B., et al.: 3D Reconstruction in Canonical Co-ordinate Space from Arbitrarily Oriented 2D Images. IEEE Trans. Med. Imaging PP(99), 1–1 (2018)

[2] Hou, B., et al.: Predicting slice-to-volume transformation in presence of arbitrary subject motion. In: MICCAI'17. pp. 296–304 (2017)

[3] Kainz, B., et al.: Fast Volume Reconstruction from Motion Corrupted Stacks of 2D Slices. IEEE Trans. Med. Imag. 34(9), 1901–13 (2015)

[4] Kendall, A., et al.: Posenet: A convolutional network for real-time 6-DOF camera relocalization. In: ICCV. pp. 2938–2946 (2015)

[5] Miolane, N.: Geomstats: Computations and statistics on manifolds with geometric structures. (Feb 2018), `https://github.com/ninamiolane/geomstats`

[6] Miolane, N., Pennec, X.: Computing Bi-Invariant Pseudo-Metrics on Lie Groups for Consistent Statistics. Entropy 17(4), 1850–1881 (Apr 2015)

[7] Pennec, X.: Probabilities and statistics on riemannian manifolds: Basic tools for geometric measurements. In: NSIP. pp. 194–198. Citeseer (1999)