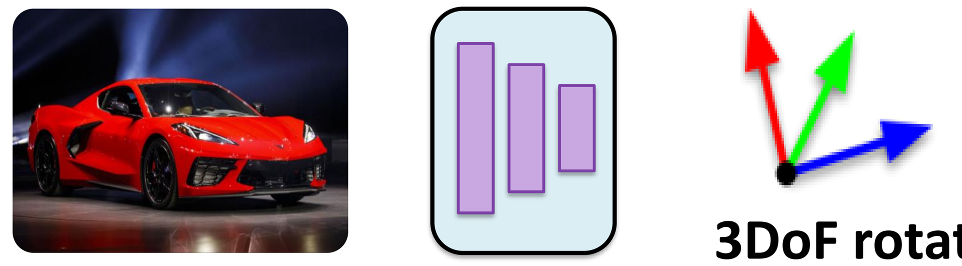
Projective Manifold Gradient Layer for Deep Rotation Regression

Jiayi Chen[1], Yingda Yin[1], Tolga Birdal[2], Baoquan Chen[1], Leonidas Guibas[2], He Wang[1]
[1]Peking University [2]Stanford University

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Introduction

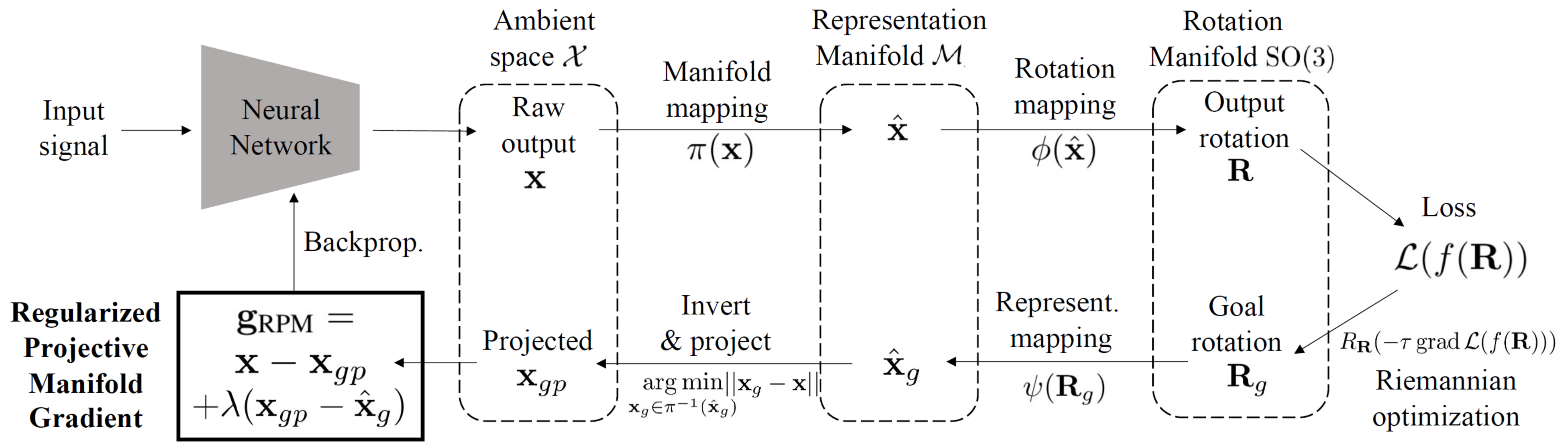
➢ **Task**: to improve the accuracy of deep rotation regression



➢ **Motivation: SO(3) is a non-Euclidean manifold while network outputs are in a Euclidean ambient space.**
- The forward pass thus always involves **projection** onto manifold.
- However, naïve backward pass simply backprop. based on the chain rule without considering the **many-to-one** nature of the projection.

➢ **Our work:**
- proposes a **manifold-aware gradient layer** to replace naïve backward pass while maintaining forward pass unchanged
- **significantly improve** rotation regression on a broad range of tasks (supervised/unsupervised rotation est. from images/point clouds) and rotation representations <u>at no cost of speed and memory</u>.

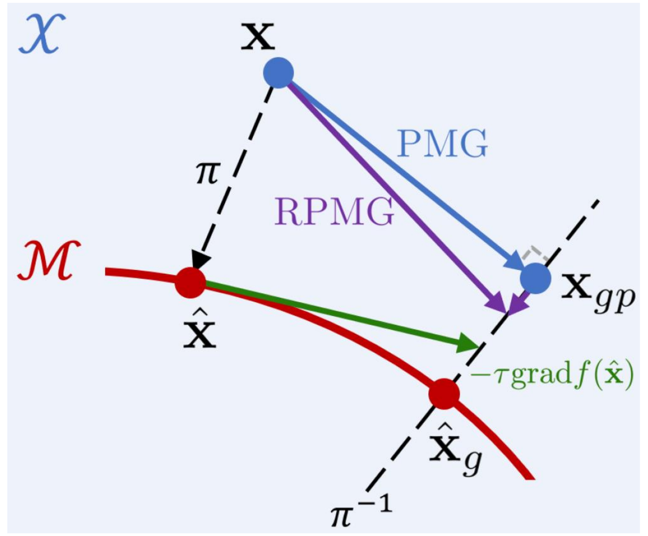
Method Overview



➢ **Forward – same as previous works**

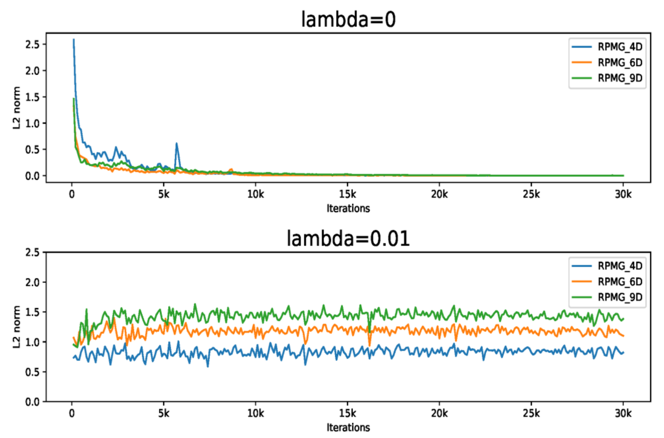
Rotation Repres.	Ambient Space	Manifold Mapping	Representation Manifold	Rotation Mapping
Quaternion	$\mathbb{R}^4$	Length norm	S^3	quat.-rot. conversion
6D [1]	$\mathbb{R}^6$	Gram-Schmidt process	Grassmann Manif.	cross product
9D [2]	$\mathbb{R}^9$	Symmetric SVD	SO(3)	identity mapping

➢ **Backward**
- Find a goal R_g using **Riemannian gradient**: $\mathbf{R}_g \leftarrow R_{\mathbf{R}}(-\tau \operatorname{grad} \mathcal{L}(f(\mathbf{R})))$ especially useful when R_{GT} is not available (self-supervised case).
- Inverse image $\pi^{-1}(\widehat{x_g})$: one-to-many but usually an analytical solution is available after some relaxations.
- Project: To find the element which is closest to the raw output in $\pi^{-1}(\widehat{x_g})$, by solving $\mathbf{x}_{gp} = \underset{\pi(\mathbf{x}_g)=\hat{\mathbf{x}}_g}{\operatorname{argmin}} ||\mathbf{x} - \mathbf{x}_g||_2$
- **Projective manifold gradient:**
$$\mathbf{g}_{PM} = \mathbf{x} - \mathbf{x}_{gp}$$
- **Key insight:**
  - *a multi-ground-truth problem* for x
  - the lowest redundancy in the gradient
  - similar to min-of-N strategy

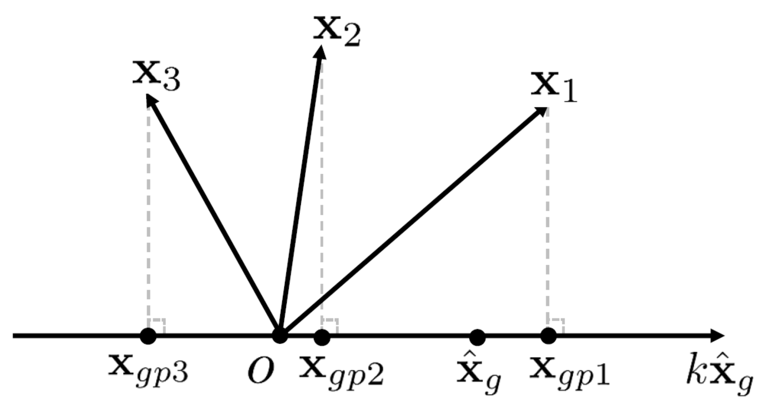


➢ **Vanishing length problem**
- **Reason**: the projection process will make the length of raw output decrease and further lead to unstable training.
- **Solution**: add a regularization term, $\lambda(\mathbf{x}_{gp} - \hat{\mathbf{x}}_g)$, which leads to our final **Regularized Projective Manifold Gradient (RPMG)**:

$$\mathbf{g}_{RPM} = \mathbf{x} - \mathbf{x}_{gp} + \lambda(\mathbf{x}_{gp} - \hat{\mathbf{x}}_g)$$



➢ **Reflection problem**
- **Reason**: the analytical solution of the inverse image assumes $\hat{\mathbf{x}}_g$ close to x
- **Solution**: use a small hyperparameter τ in Riemannian optimization.



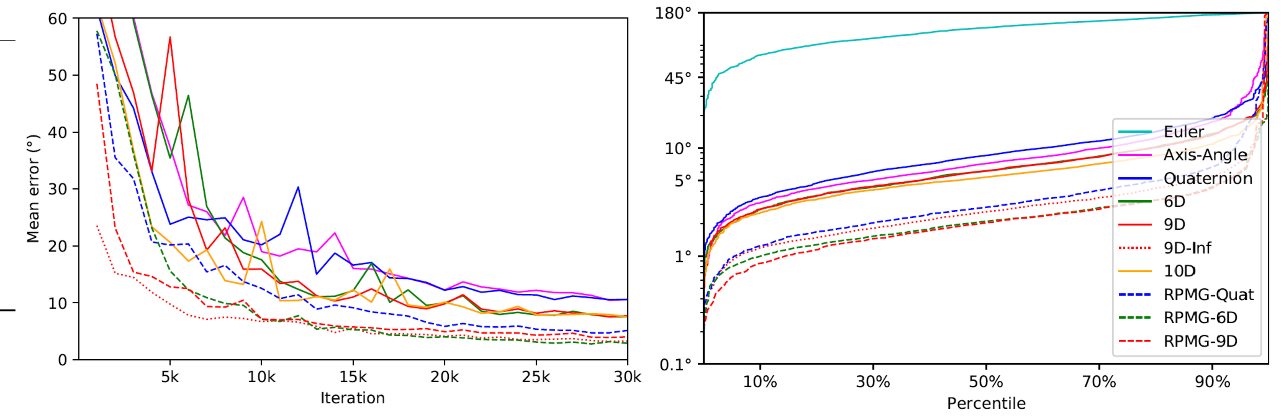
Reference

[1] Yi Zhou *et al.* On the continuity of rotation representations in neural networks. CVPR 2019.
[2] Jake Levinson *et al.* An analysis of svd for deep rotation estimation. NeurIPS 2020.

Experiment

Rotation regression w/ GT supervision from point clouds

	Mean (°)	Med (°)	5° Acc (%)
Euler	124.88	130.37	0.0
Axis-Angle	10.57	7.20	29.1
Quaternion	10.55	8.57	21.7
6D	7.25	6.15	38.2
6D-Inf (MG-6D)	3.37	2.77	85.7
9D	7.70	6.16	38.0
9D-Inf (MG-9D)	3.25	2.57	87.1
10D	7.42	5.34	44.5
RPMG-Quat	5.12	2.83	80.4
RPMG-6D	**2.85**	2.09	**92.7**
RPMG-9D	3.97	**2.02**	92.5



Rotation regression w/o GT supervision from point clouds

	Instance-Level Self-Supervise			Category-Level Self-Supervise		
	Mean (°)	Med (°)	5° Acc (%)	Mean (°)	Med (°)	5° Acc (%)
Euler	129.3	132.9	0	12.14	6.91	33.6
Axis-Angle	36.31	6.98	37	35.49	20.80	4.7
Quaternion	4.04	3.30	74	11.54	7.67	29.8
6D	43.9	6.49	44	14.13	9.41	23.4
9D	2.47	2.02	92.5	11.44	8.01	23.8
9D-Inf	101.5	96.61	0	4.07	3.28	76.7
10D	2.18	1.91	96.5	9.28	7.05	32.6
RPMG-Quat	2.88	2.38	91.5	4.86	3.25	75.8
RPMG-6D	3.08	2.92	89.5	**2.71**	**2.04**	**92.1**
RPMG-9D	**1.40**	**1.17**	**100**	3.75	2.10	91.1

Rotation regression w/ GT supervision from images

	Chair			Sofa		
	Mean (°)	Med (°)	5° Acc (%)	Mean (°)	Med (°)	5° Acc (%)
Euler	21.46	10.95	10.4	27.46	12.00	9.4
Axis-Angle	25.71	14.27	7.2	30.25	14.55	6.2
Quaternion	25.75	14.99	6.3	30.00	15.73	5.7
6D	19.60	9.09	19.1	17.51	7.33	27.3
9D	17.46	8.30	23.1	19.75	7.58	24.9
9D-Inf	12.10	5.09	49.2	12.48	3.45	69.7
10D	18.40	9.02	19.6	20.89	8.73	19.8
RPMG-Quat	13.03	5.90	39.9	13.02	3.60	66.6
RPMG-6D	12.94	4.74	53.1	11.52	2.79	77.1
RPMG-9D	**11.93**	**4.36**	**58.1**	**10.49**	**2.41**	**81.7**

Ablation Study

			Complete		
			Mean (°)	Med (°)	5° Acc (%)
L2 w/ 6D	-	-	7.25	6.15	38.2
MG-6D	$\lambda = 1$	$\tau_{convergence}$	3.27	2.68	86.1
		τ_{gt}	3.37	2.77	85.7
PMG-6D	$\lambda = 0$	$\tau_{convergence}$	64.41	40.99	2.8
		τ_{gt}	103.2	100.4	0.0
RPMG-6D	$\lambda = 0.01$	τ_{init}	3.60	2.30	91.1
		$\tau_{convergence}$	3.41	**2.04**	87.2
		τ_{gt}	4.12	2.22	87.1
		$\tau_{init} \to \tau_{convergence}$	**2.85**	2.09	**92.7**
	$\lambda = 0.005$	$\tau_{init} \to \tau_{convergence}$	3.07	2.11	89.6
	$\lambda = 0.1$	$\tau_{init} \to \tau_{convergence}$	**2.85**	2.19	90.9

Camera pose estimation from images

	King's College		Old Hospital		Shop Facade		St Mary's Church		Average	
	T(m)	R(°)	T(m)	R(°)	T(m)	R(°)	T(m)	R(°)	T(m)	R(°)
Euler	1.16	2.85	2.54	2.95	1.25	6.48	1.98	6.97	1.73	4.81
Axis-Angle	1.12	2.63	2.41	3.38	0.84	5.05	2.16	7.58	1.63	4.66
Quaternion	0.98	2.50	2.39	3.44	1.06	6.01	2.59	8.81	1.76	5.19
6D	1.10	2.56	2.21	3.43	1.01	5.43	**1.73**	5.82	1.51	4.31
9D	1.14	3.03	2.11	3.50	**0.88**	6.39	1.95	5.95	1.52	4.72
9D-Inf	**0.98**	2.32	**1.89**	3.32	1.15	6.36	1.96	6.25	**1.50**	4.56
10D	1.54	2.62	2.32	3.39	1.20	5.76	1.85	6.69	1.73	4.62
RPMG-Quat	1.04	1.91	2.42	**2.72**	0.98	4.28	1.82	4.89	1.57	**3.45**
RPMG-6D	1.55	**1.70**	2.62	3.09	0.95	5.01	2.44	5.18	1.89	3.75
RPMG-9D	1.57	1.82	4.37	3.12	0.93	**4.17**	1.92	**4.69**	2.20	**3.45**