IMPERIAL

Meta

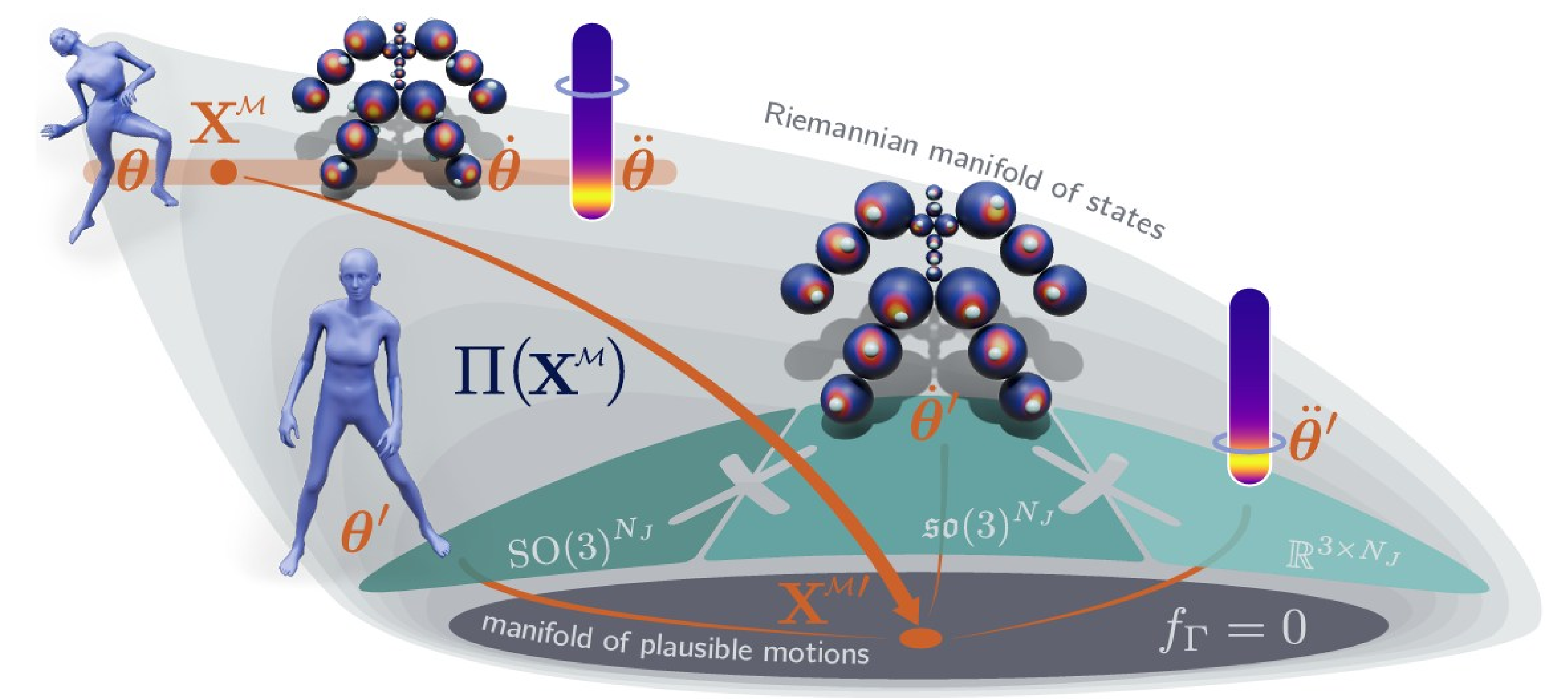
Geometric Neural Distance Fields for Learning Human Motion Priors

Zhengdi Yu[1] Simone Foti[1] Linguang Zhang[2] Amy Zhao[2] Cem Keskin[2] Stefanos Zafeiriou[1] Tolga Birdal[1]

[1]Imperial College London [2]Meta Reality Labs

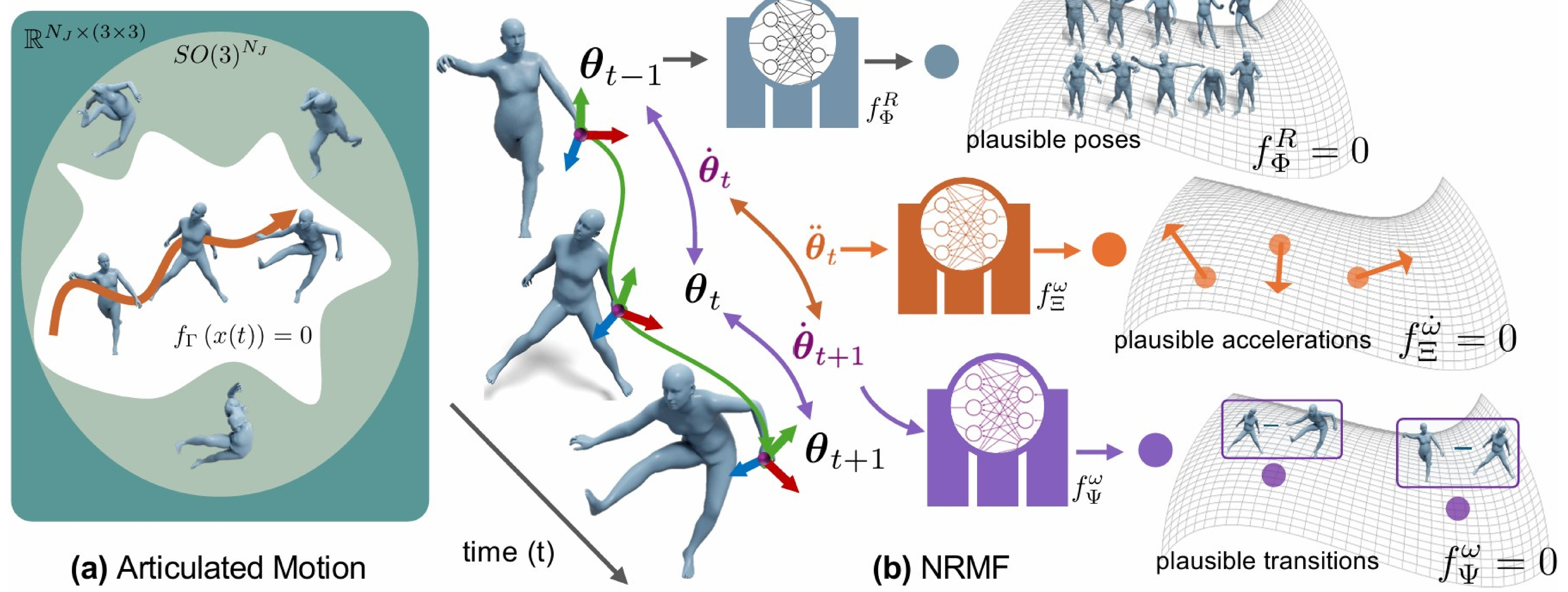
CVPR Nashville JUNE 11-15, 2025

NRMF = Neural Riemannian Motion Fields



NRMF is a general-purpose, expressive and robust unconditional motion prior. It models the space of plausible **poses (θ), transitions ($\dot{\theta}$)** and **accelerations ($\ddot{\theta}$)** on the zero-level set of a **geometric neural distance field**. Poses are depicted along side their transitions and accelerations, which are visualized as blue dots onto the per-joint distributions of learned transitions and as blue rings around the magnitude distribution of all accelerations.

Learning NRMF



$$\mathbf{X} = [t_r, \theta, \dot{\theta}, \ddot{\theta}] \in (\mathbb{R}^3 \times \mathcal{M}) \qquad \text{(Motion state)}$$

$$f_\Gamma = \begin{bmatrix} f_\Phi^R(\theta) \\ f_\Psi^\omega(\dot{\theta} \mid \theta) \\ f_\Xi^{\dot{\omega}}(\ddot{\theta} \mid \theta, \dot{\theta}) \end{bmatrix} \begin{matrix} \text{(pose field)} \\ \text{(transition field)} \quad \text{(RMF)} \\ \text{(acceleration field)} \end{matrix}$$

$$\theta_{t+1} \coloneqq \Pi^R(\theta_t) = \mathrm{Exp}_{\theta_t}\left(-\alpha_\theta f_\Phi^R(\theta_t) \frac{\mathrm{grad} f_\Phi^R(\theta_t)}{\|\mathrm{grad} f_\Phi^R(\theta_t)\|}\right)$$

$$\dot{\theta}_{t+1} \coloneqq \Pi_t^\omega(\dot{\theta}_t) = \dot{\theta}_t - \alpha_{\dot{\theta}} f_\Psi^\omega(\dot{\theta}_t) \frac{\nabla f_\Psi^\omega(\dot{\theta}_t \mid \theta_t)}{\|\nabla f_\Psi^\omega(\dot{\theta}_t \mid \theta_t)\|}$$

$$\ddot{\theta}_{t+1} \coloneqq \Pi_t^{\dot{\omega}}(\ddot{\theta}_t) = \ddot{\theta}_t - \alpha_{\dot{\omega}} f_\Xi^{\dot{\omega}}(\ddot{\theta}_t) \frac{\nabla f_\Xi^{\dot{\omega}}(\ddot{\theta}_t \mid \theta_t, \dot{\theta}_t)}{\|\nabla f_\Xi^{\dot{\omega}}(\ddot{\theta}_t \mid \theta_t, \dot{\theta}_t)\|}$$

$$\Gamma^* = \begin{bmatrix} \Phi^* = \arg\min_\Phi \sum_{i=1}^N \left\| f_\Phi^R(\theta_i) - \min_{\theta' \in \mathcal{D}_\theta} d_{SO(3)}^{N_J}(\theta_i, \theta') \right\| \\ \Psi^* = \arg\min_\Psi \sum_{i=1}^N \left\| f_\Psi^\omega(\dot{\theta}_i) - \min_{\dot{\theta}' \in \mathcal{D}_{\dot{\theta}}} d_{\mathfrak{so}(3)}^{N_J}(\dot{\theta}_i, \dot{\theta}') \right\| \\ \Xi^* = \arg\min_\Xi \sum_{i=1}^N \left\| f_\Xi^{\dot{\omega}}(\ddot{\theta}_i) - \min_{\ddot{\theta}' \in \mathcal{D}_\theta} d_{\mathbb{R}^3}^{N_J}(\ddot{\theta}_i, \ddot{\theta}') \right\| \end{bmatrix}$$

Learning

$f_\Gamma^* \rightarrow$ Neural Riemannian Motion Field

α: learning rate, ω: angular velocity, t_r: root translation, $\mathcal{D}$: dataset

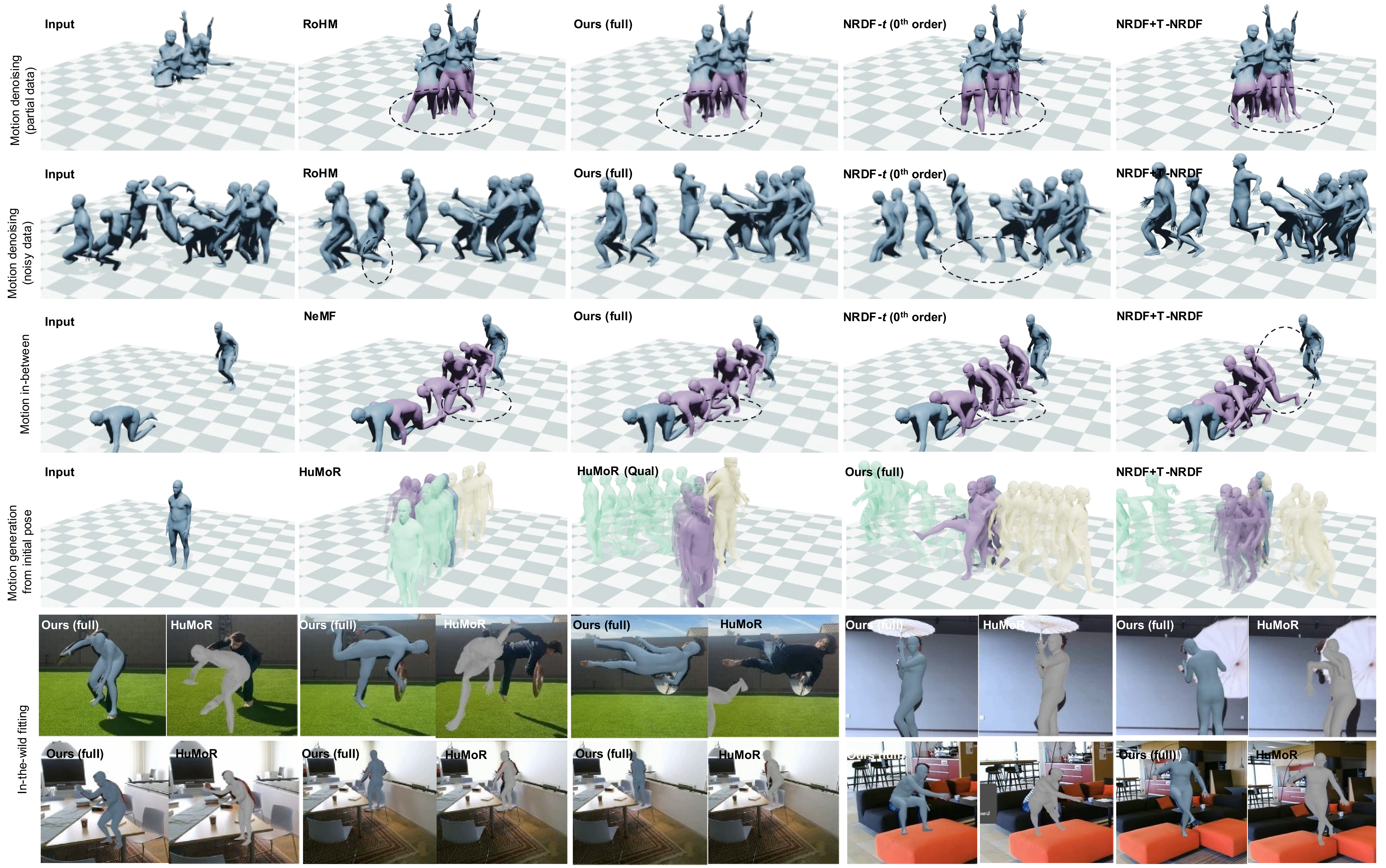


Figure: Qualitative comparisons. The mesh and body parts are the observation inputs and we show the output or optimization results. For in-the-wild fitting, we conduct experiments on both in-door datasets (bottom row) and online videos (upper row). For motion generation, we show samples generated from the common standing pose.

Evaluation

Method	MPJPE (mm)	MPVPE (mm)	Acc Err (m/s^2)	Trans Err ($\times 10^{-3}$)
SMPLer-X [5]	82.65	94.23	23.71	31.63
+ No prior	84.67	96.82	26.75	34.54
+ VPoser [24]	79.98	91.53	25.82	30.58
+ DPoser [21]	75.45	87.04	27.54	28.69
+ PoseNDF [33]	73.49	84.61	25.12	29.77
+ NRDF [14]	71.88	83.23	24.31	26.38
+ T-NRDF	66.98	76.94	9.89	7.98
+ A-NRDF	70.88	81.92	6.73	11.87
+ RoHM [42]	69.78	79.72	9.13	12.37
+ **NRMF** (full)	**66.13**	**75.61**	**6.52**	**5.67**

Tab 1: Quantitative evaluation results of motion refinement on 3DPW.

Method	RGB-D				RGB			
	$\|Acc\|$	Dist	FID_m	Freq	$\|Acc\|$	Dist	FID_m	Freq
HuMoR [27]	1.88	35.67	11.65	6.13%	2.35	41.92	12.47	10.52%
PhaseMP [31]	-	-	-	-	1.80	46.96	-	-
RoHM [42]	1.79	6.67	12.16	3.21%	2.19	**14.31**	13.65	4.65%
VPoser-t	3.45	53.49	21.89	10.66%	3.24	54.96	19.64	13.95%
PoseNDF-t	2.84	48.77	16.45	11.35%	2.93	50.13	16.87	13.46%
NRDF-t	2.67	45.91	14.58	9.97%	2.77	46.78	14.32	12.98%
DPoser-t	3.32	49.45	17.32	10.89%	3.08	51.35	16.98	12.21%
NRMF	**1.48**	**5.78**	**8.95**	2.78%	**1.73**	16.21	**14.13**	5.12%

Tab 2: Motion estimation on PROX dataset.

Method	APD ↑ (cm)	FID_p ↓	FID_m ↓	$\|Acc\|$ ↓
GMM [4]	16.28	0.435	-	-
GAN-S [11]	15.68	0.201	-	-
VPoser [24]	10.75	**0.113**	-	-
Pose-NDF [33]	18.54	1.124	-	-
DPoser [21]	**36.25**	2.344	-	-
NRDF [14]	24.87	0.706	-	-
MVAE [18]	85.26	1.475	10.571	8.15
HuMoR [27]	**98.47**	0.278	9.147	4.96
Motion-NDF	93.46	0.894	8.351	-
NRMF (ours)	96.37	**0.718**	**4.218**	**3.18**

Tab 3: Quantitative evaluation of motion generation.

Method	Pos. Err. (mm)			Motion Prop.			
	All	Legs	Vtx	Contact	Freq	Dist (mm)	Acc Err
MVAE [18]	26.1	32.3	44.6	-	1.72%	1.4	6.74
HuMoR [27]	22.7	26.1	35.5	0.97	1.18%	0.8	4.67
PhaseMP [31]	26.6	28.7	39.1	0.95	1.33%	1.2	6.47
RoHM [42]	20.9	24.1	32.4	0.97	**1.12**%	0.6	2.61
MDM [32]	25.4	26.7	40.1	0.95	1.32%	**0.5**	6.24
VPoser-t [24]	36.5	44.2	52.4	-	1.53%	0.7	7.21
PoseNDF-t [33]	24.6	27.5	32.7	-	1.42%	0.9	6.97
DPoser-t [21]	25.1	27.9	32.4	-	1.35%	0.8	7.13
NRDF-t [14]	22.8	22.4	28.1	-	1.31%	0.7	6.89
DPoser + T-PoseNDF	19.1	21.1	25.9	0.97	1.31%	0.7	3.29
DPoser + T-NRDF	17.3	17.9	23.3	0.95	1.22%	0.5	3.13
VPoser + T-PoseNDF	21.4	23.5	31.6	0.97	1.20%	0.6	3.47
VPoser + T-NRDF	19.0	20.8	27.5	0.96	1.29%	0.6	3.18
PoseNDF + T-PoseNDF	18.6	19.2	26.5	0.98	1.28%	0.6	3.24
PoseNDF + T-NRDF	17.0	18.2	25.2	0.96	1.26%	0.7	3.08
NRDF + T-PoseNDF	18.8	20.5	25.4	0.96	1.26%	0.7	3.19
NRDF + T-NRDF	16.7	17.5	22.6	0.98	1.22%	0.5	2.97
Motion-NDF	18.1	19.8	23.5	0.98	1.22%	0.5	2.64
NRMF (ours)	**16.4**	**17.1**	**19.9**	**0.98**	1.17%	0.5	**2.25**

Tab 4: Motion denoising on AMASS dataset.

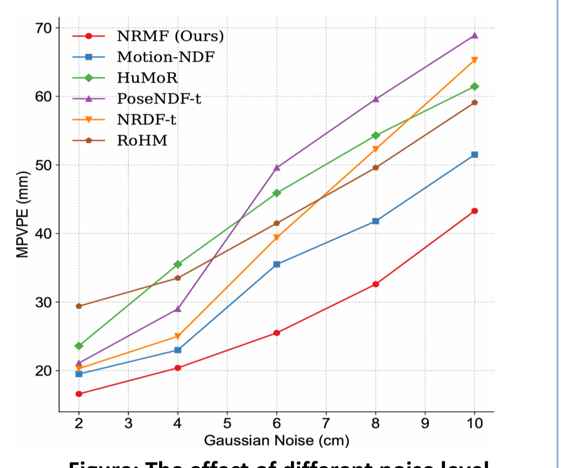


Figure: The effect of different noise level

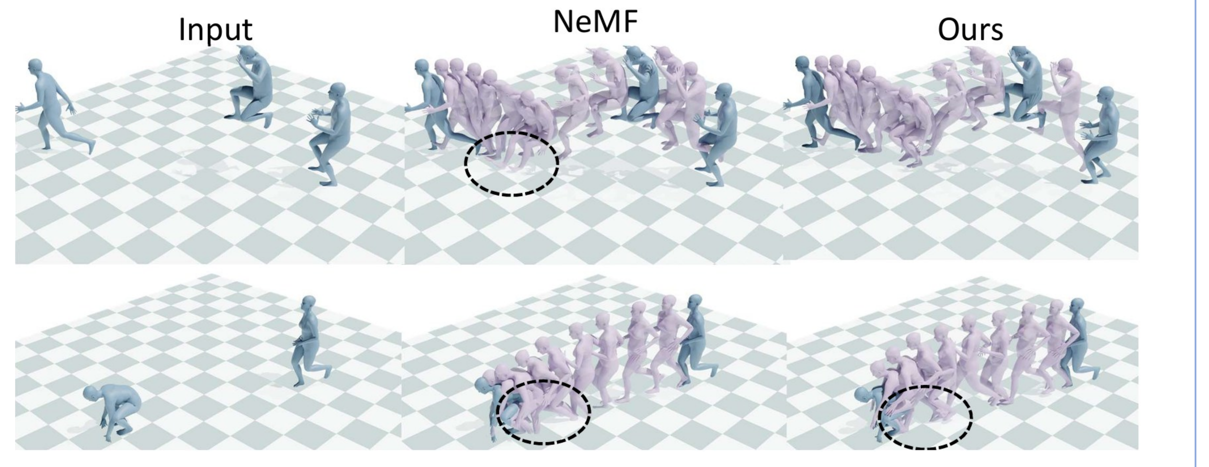


Figure: Comparison on motion in-betweening.

Optimization

$\mathcal{L}_{\dot{\theta}} \coloneqq f_\Psi^\omega(\dot{\theta}_i), \mathcal{L}_{\ddot{\theta}} \coloneqq f_\Xi^{\dot{\omega}}(\ddot{\theta}_i)$

$$E_I'(t_r, \theta, \beta) = \mathcal{L}_{\text{data}} + \lambda_\beta \mathcal{L}_\beta + \lambda_\theta \mathcal{L}_\theta + \lambda_{\text{reg}} \mathcal{L} + \lambda_{\ddot{\theta}} \mathcal{L}_{\dot{\theta}} + \lambda_{\ddot{\theta}} \mathcal{L}_{\ddot{\theta}}$$

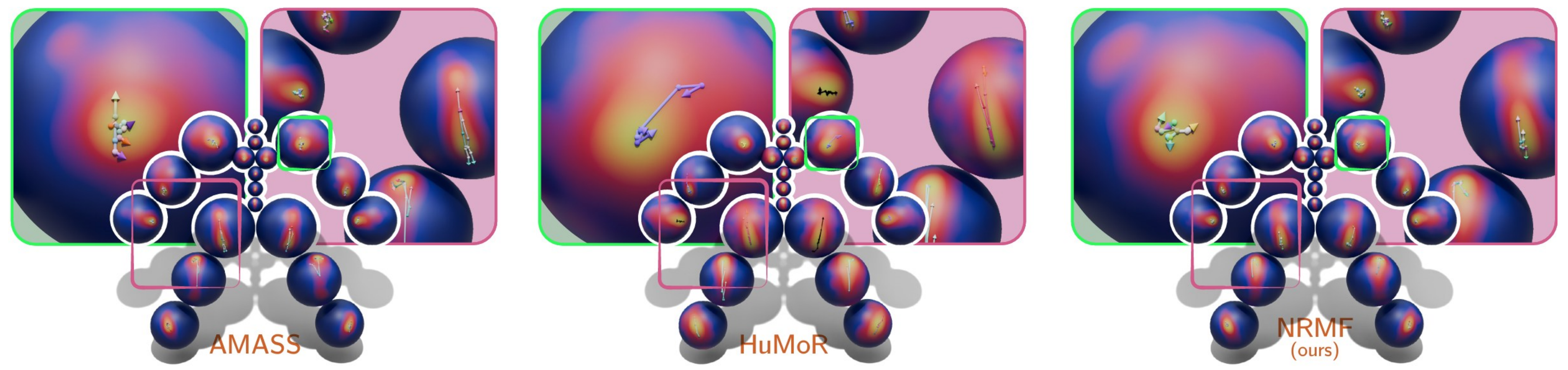
Algorithm 1 RMF Integrator

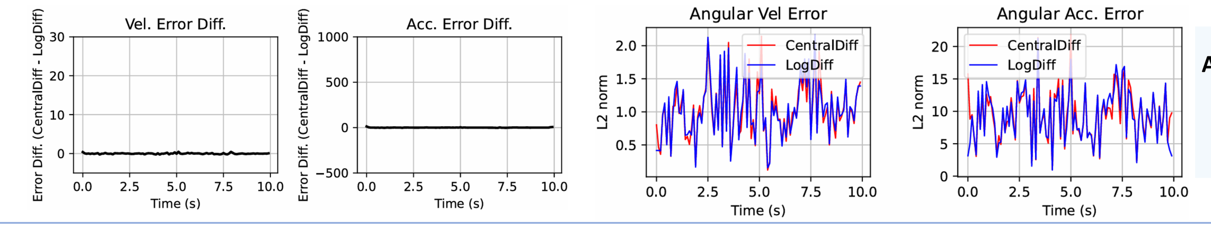
Input: An initial pose $\boldsymbol{\theta}_0$ and velocity $\dot{\boldsymbol{\theta}}_0$, a possibly noisy acceleration sequence $\{\ddot{\boldsymbol{\theta}}_t\}_t$ and all components of the trained network f_Γ (and hence the projectors Π),

Output: Plausible motion $\{\mathbf{X}_t\}_t$

1: **for** $t \leftarrow 1, \ldots, T$ **do** ▷ Euler integrate of velocities
2: $\dot{\boldsymbol{\theta}}_t \leftarrow \dot{\boldsymbol{\theta}}_{t-1} + \lambda_t \ddot{\boldsymbol{\theta}}_{t-1}$
3: $\dot{\boldsymbol{\theta}}_t \leftarrow \Pi^\omega(\dot{\boldsymbol{\theta}}_t)$
4: **end for**
5: **for** $t \leftarrow 1, \ldots, T$ **do** ▷ Update poses
6: $\boldsymbol{\theta}_t \leftarrow \mathrm{Exp}_{\boldsymbol{\theta}_{t-1}}(\alpha_t [\dot{\boldsymbol{\theta}}_{t-1}]_x)$
7: $\boldsymbol{\theta}_t \leftarrow \Pi^R(\boldsymbol{\theta}_t)$
8: **end for**
9: Compose $\{\mathbf{X}_t\}_t$ from individual estimates $\{\ddot{\boldsymbol{\theta}}_t, \dot{\boldsymbol{\theta}}_t, \boldsymbol{\theta}_t\}_t$

Analysis

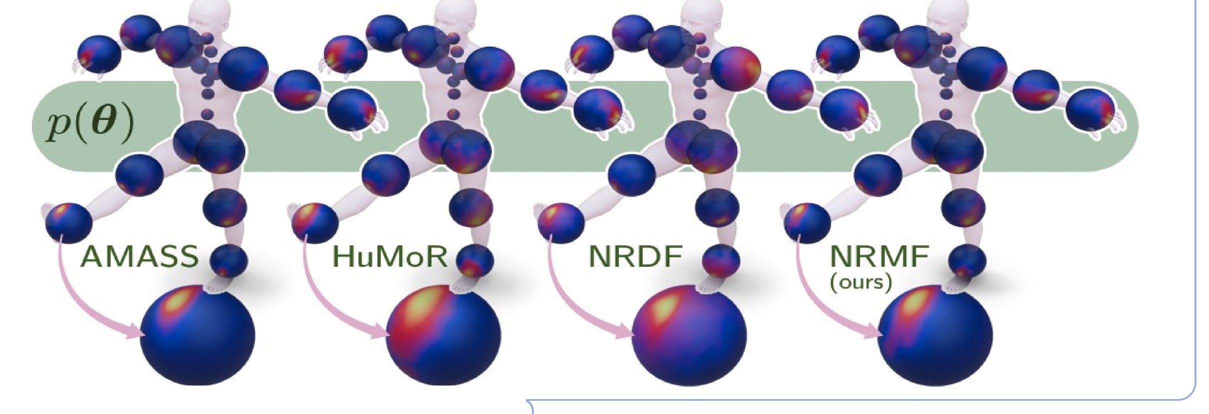


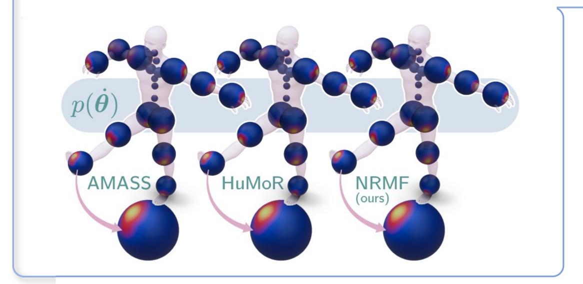


Angular acceleration: $\frac{\mathrm{d}[\boldsymbol{\omega}_t]_x}{\mathrm{d}t} \coloneqq [\dot{\boldsymbol{\omega}}_t]_x \coloneqq \ddot{\mathbf{R}}_t = \mathbf{R}_t(\boldsymbol{\Omega}_t^2 + \dot{\boldsymbol{\Omega}}_t)$ *where $\boldsymbol{\Omega}_t^2 + \dot{\boldsymbol{\Omega}}_t$ is also skew-symmetric*

Angular velocity: $\Omega(t) = R^\top(t)\dot{R}(t) \in \mathfrak{so}(3)$ *where $[\omega]_x = \Omega, \omega \in \mathbb{R}^3$*

Distribution





References

[1] D. Rempe, et al., HuMoR: 3DHuman Motion Model for Robust Pose Estimation. **ICCV 2021.**
[2] S. Zhang, et al., Rohm: Robust human motion reconstruction via diffusion. **CVPR 2024.**
[3] Y. He et al., Nrdf: Neural Riemannian distance fields for learning articulated pose priors. **CVPR 2024.**
[4] Z. Cai et al., SMPLer-X: Scaling up expressive human poseand shape estimation. **NeurIPS 2023.**
[5] C. He et al., Nemf: Neural motion fields for kinematic animation. **NeurIPS 2022.**
[6] G. Pavlakos et al., Expressive body capture: 3d hands, face, and body from a single image. **CVPR 2019**
[7] G. Tiwari et al., PoseNDF: Modeling human pose manifolds with neural distance fields. **ECCV 2022**