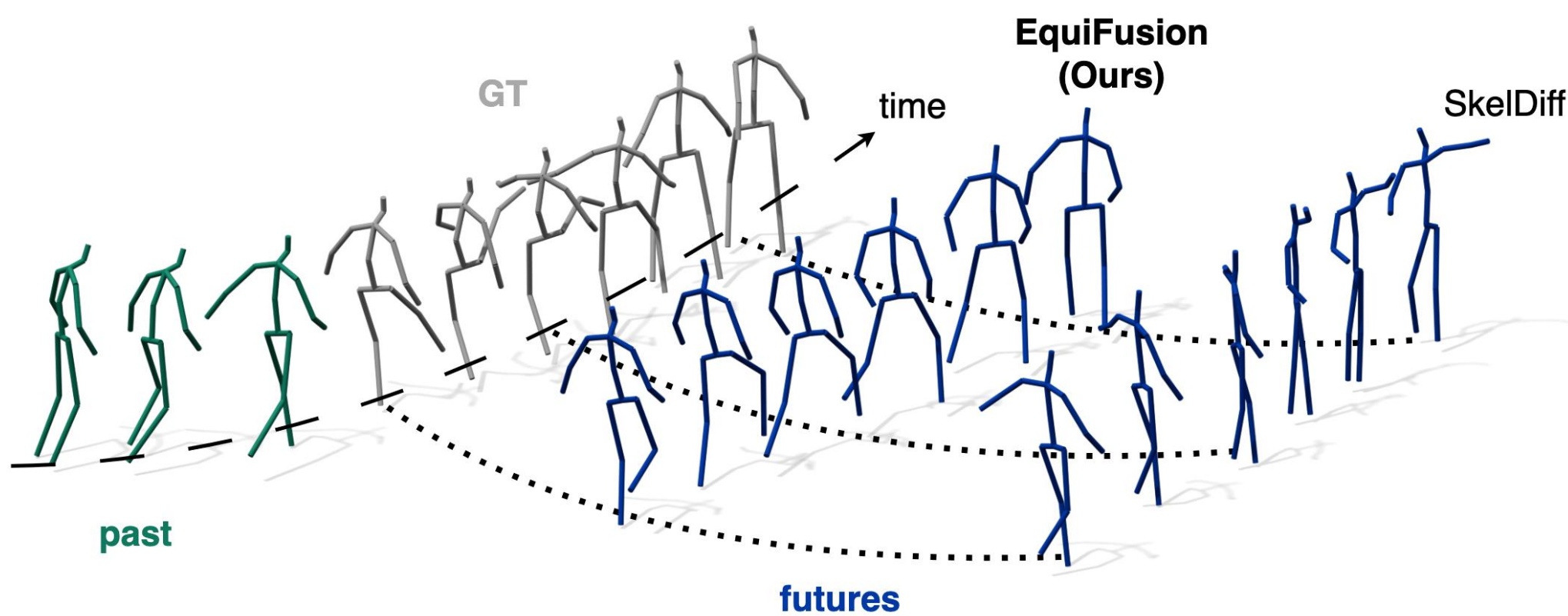


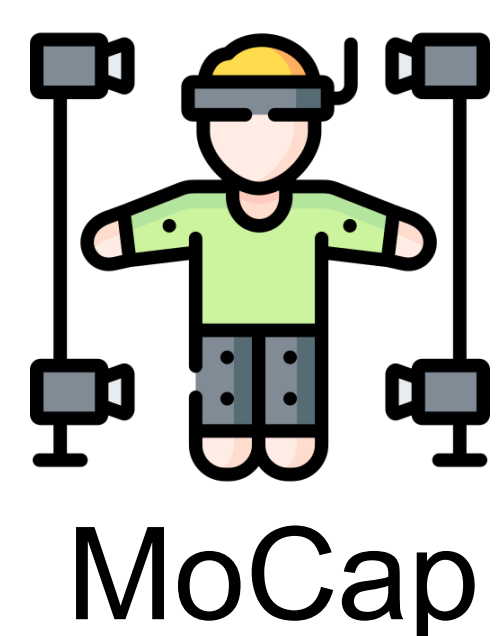
Human Motion Prediction



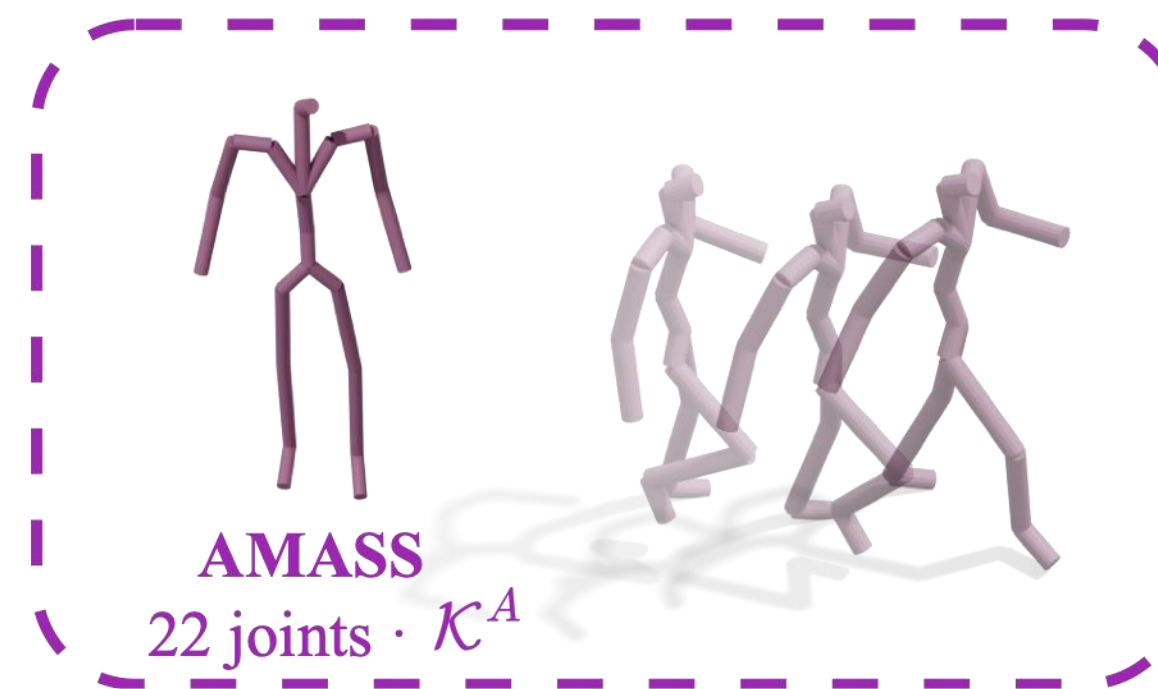
- Given a **past** motion, generate multiple **futures**
- Generated motions should be **meaningful, realistic, and diverse**

Challenges

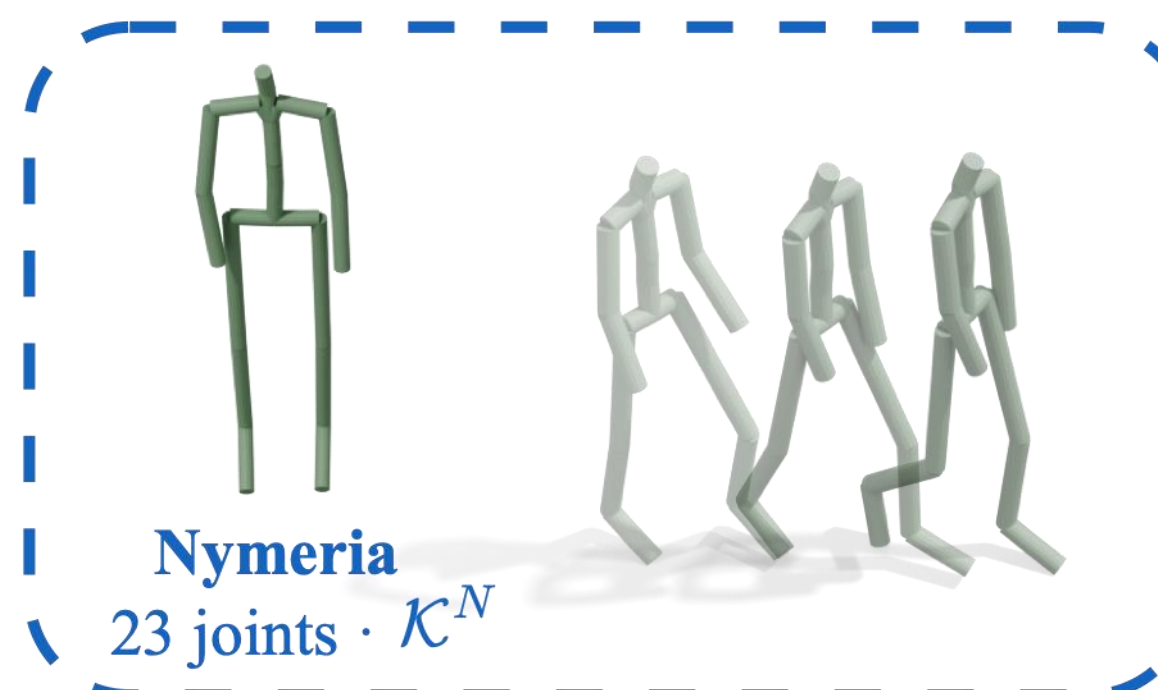
- Due to different capturing devices datasets have **different skeleton i.e. Kinematics $\mathcal{K}$**
- Baselines** can only run inference on their training kinematics or require retargeting



MoCap



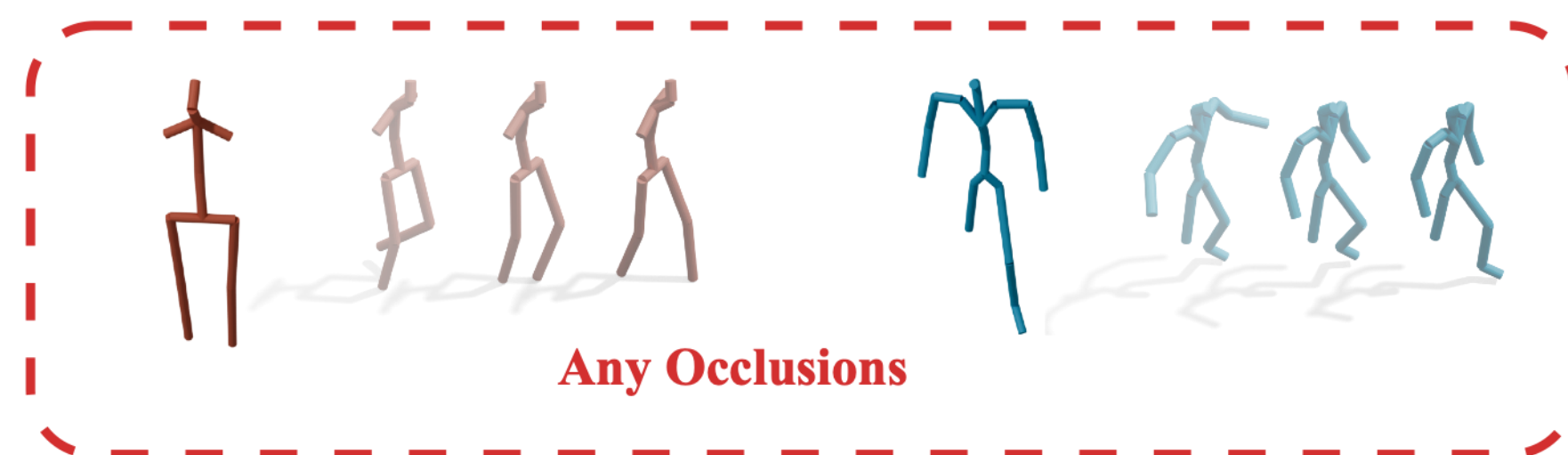
AMASS
22 joints · $\mathcal{K}^A$



Nymeria
23 joints · $\mathcal{K}^N$

Project Aria

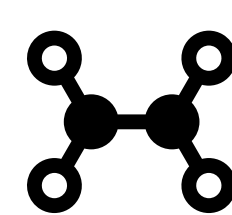
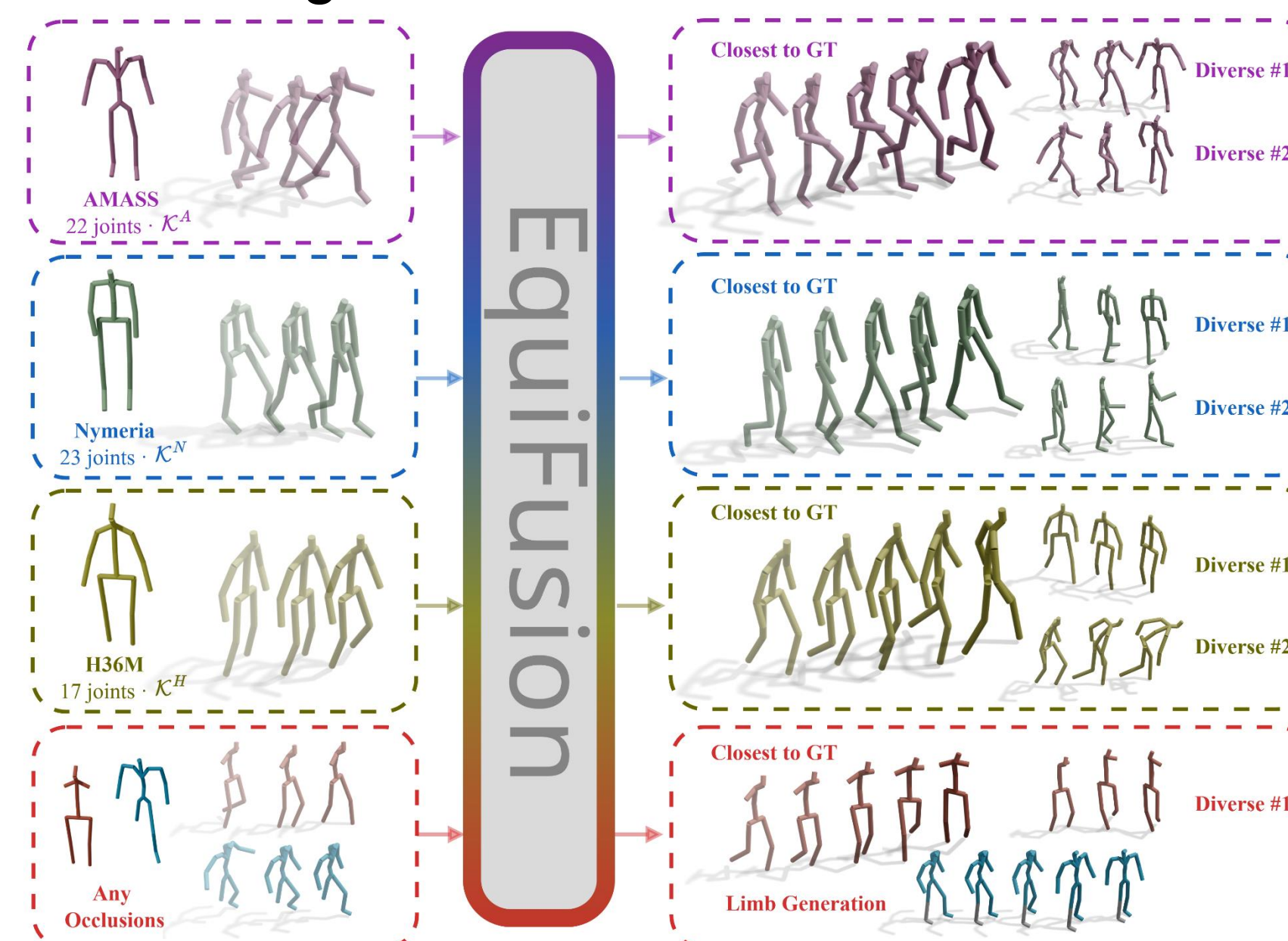
- Baselines** always need **full input data** and can't handle **occluded or missing limbs**



Any Occlusions

EquiFusion

- EquiFusion** is the first HMP model to allow **cross-dataset** training and inference



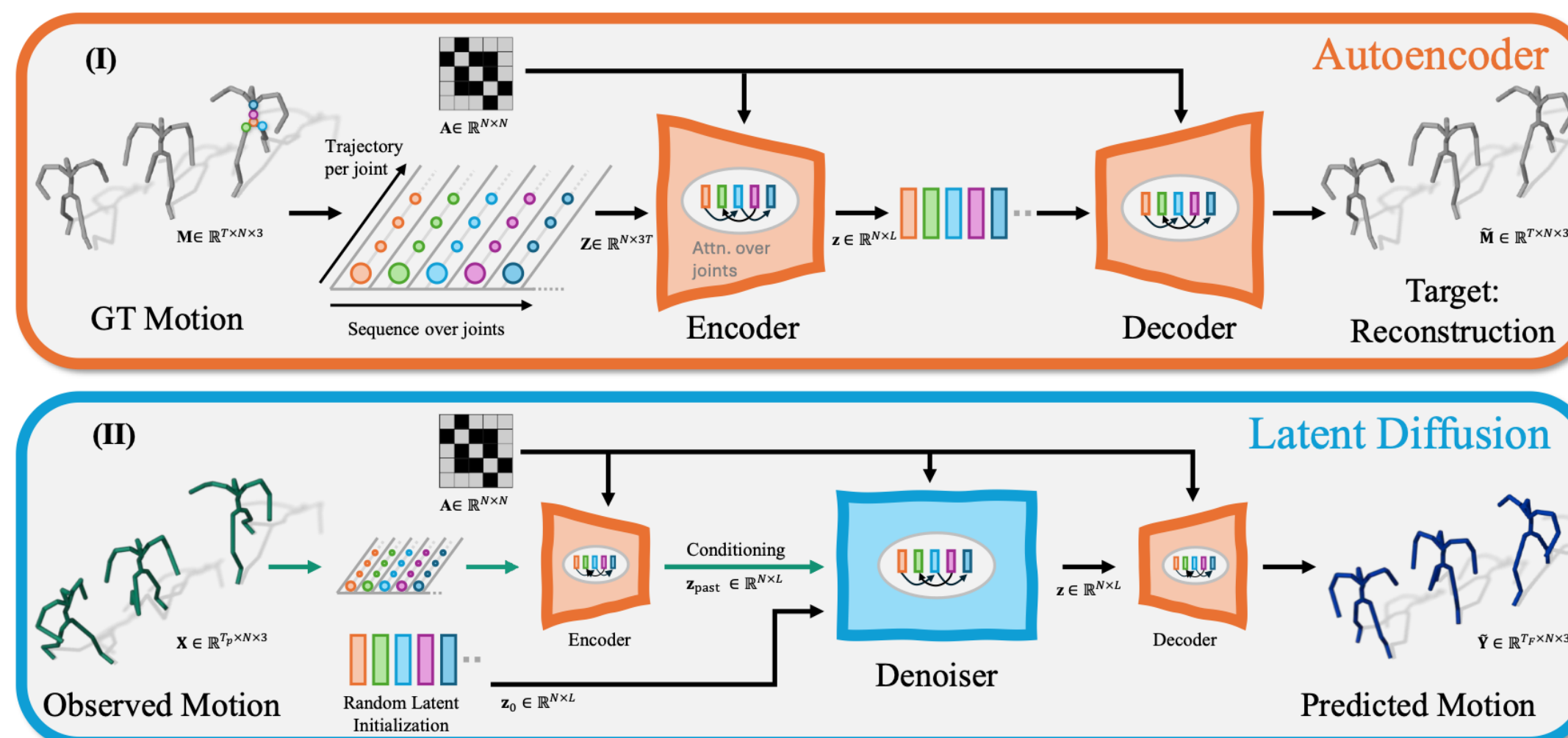
The first fully **kinematics-agnostic** model for **Stochastic Human Motion Prediction**



Runs training and inference **independent** of joint number and type

	cross-dataset	baselines	EquiFusion
training		✗	✓
inference		✗	✓

Equivariant Latent Diffusion

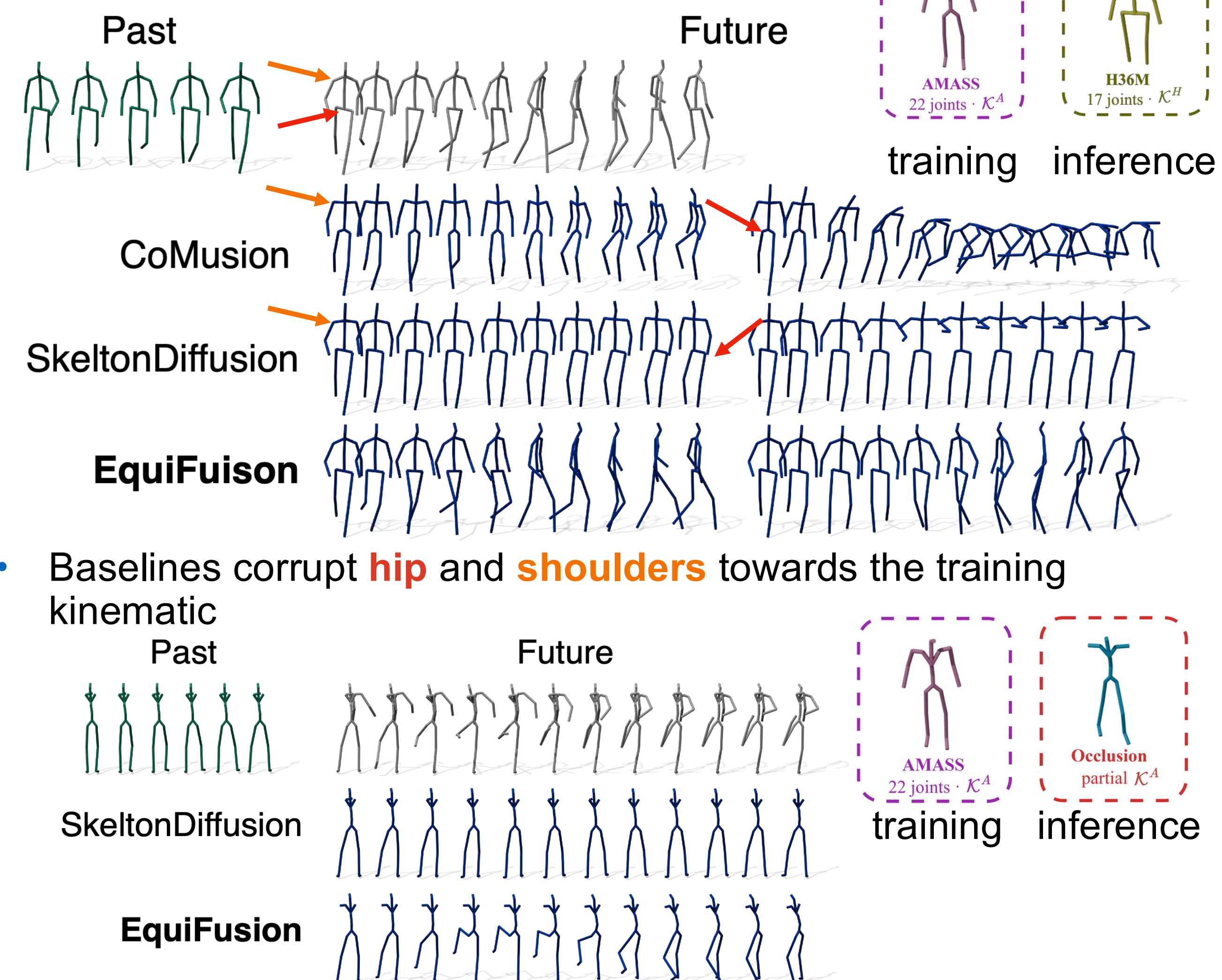


- We make the **kinematics** an explicit input and design a fully **permutation equivariant** architecture for our **latent diffusion model**

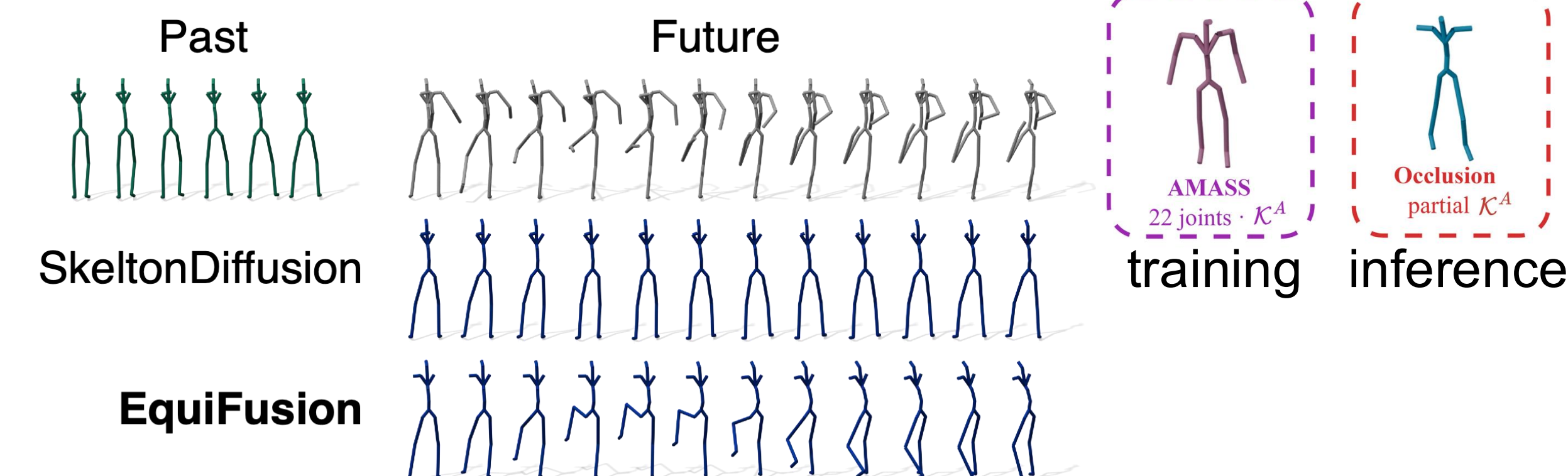
$$f(\mathbf{P}\mathbf{X}, \mathbf{P}\mathbf{A}\mathbf{P}^T) = \mathbf{P}f(\mathbf{X})$$

Zero-Shot Kinematics

- Baselines **need** kinematic-conversion models to convert their input and output leading to **motion corruptions**
- EquiFusion** runs zero-shot inference on arbitrary skeleton kinematics, in full body and occluded limbs

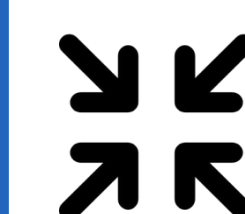


- Baselines corrupt **hip** and **shoulders** towards the training kinematic



Evaluation

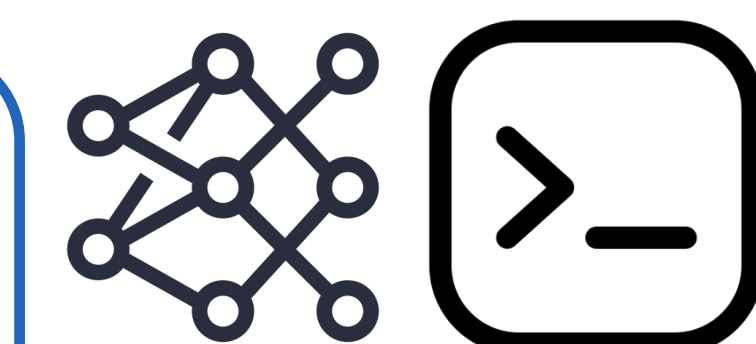
Method	new	Precision ↓		MM GT ↓		Div ↑	Realism ↓	
		uADE	uFDE	uMMA	uMMF	uAPD	CMD	FID
DivSamp	✗	9.27	12.61	11.42	13.27	4.210	47.783	5.629
BeLFusion	✗	9.24	11.62	10.93	12.16	1.305	8.031	1.195
CoMusion	✗	10.07	12.15	12.49	12.96	<u>2.070</u>	8.587	1.426
SkelDiff	✗	10.81	14.99	12.75	15.42	0.992	7.616	5.252
EquiFusion(A)	✓	<u>7.86</u>	<u>10.47</u>	<u>10.66</u>	<u>11.61</u>	1.973	7.061	<u>0.691</u>
EquiFusion(A+N)	✓	7.71	10.21	10.58	11.36	1.797	<u>7.349</u>	0.504



We achieve up to **75%** parameter reduction in network size allowing for **fast inference**



A new 3D parametrization: **limb directions**. Predicted motions keep the skeleton size **without limb stretching** or **jittering**



model code



paper page

