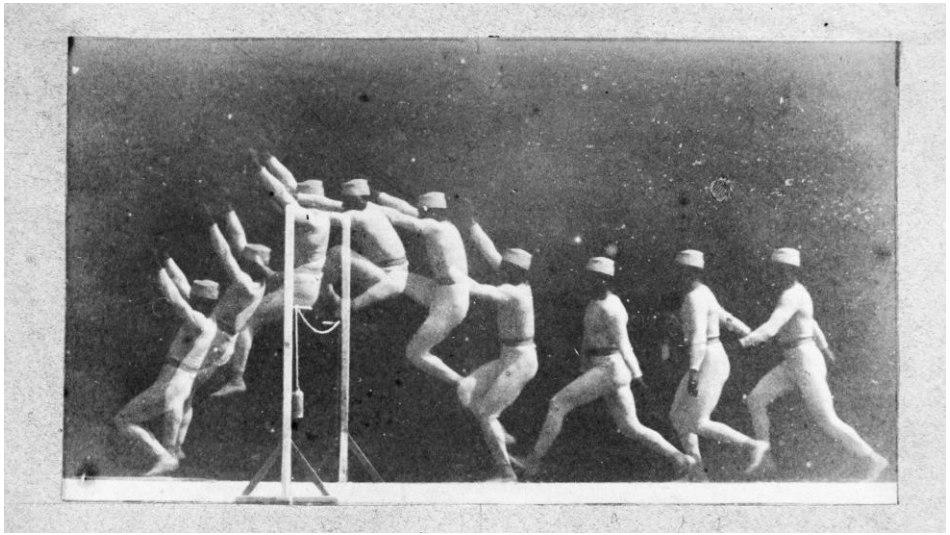
AI-Based Synthesis for Complex Human Poses, Motions, and Interactions

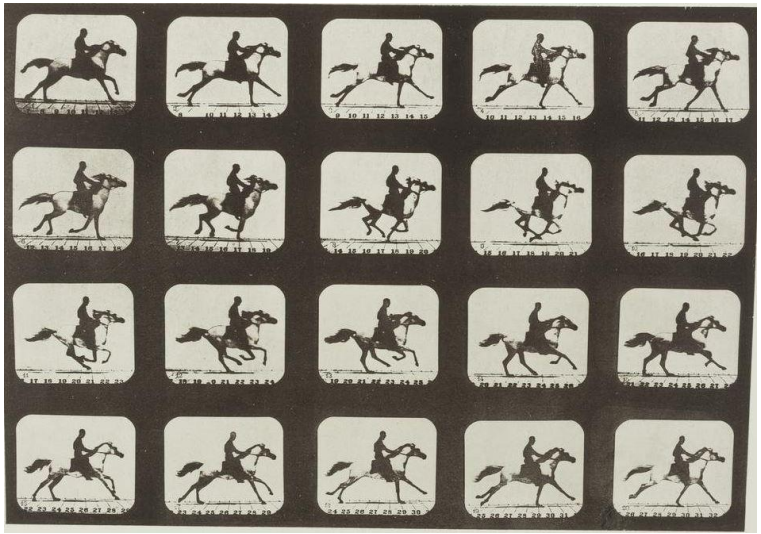
Candidate: Georgios Albanis, Supervisor: Dr. Konstantinos Kolomvatsos

Quote

Motion Capture Origins



"*The movement of a man jumping hurdle*" by Étienne-Jules Marey



"*The Horse in Motion*" by Eadweard Muybridge

Motion Capture Applications

Existing Solutions

Optical MoCap

High-End Cameras track reflective markers placed on the subject.

IMU MoCap

Inertial Sensors placed on the human body.

Markerless MoCap

No need for sensors based on AI keypoints detections

Existing Solutions

Feature	High-end optical	Low-end optical	Markerless
Cost	•••••	•••	••
Setup effort	•••••	•••	•
Space needs	•••••	•••	•
Cleanup effort	••	••••	•••••
Data fidelity	•••••	•••	••

The Shift Toward Accessibility

- 💰 Lower CapEx
- 🛠️ Lower complexity and expertise needed to operate
- 🧹 Remove the need for manual cleanup
- 🚫📍 Move to a markerless solution

On what Price?

💰 Lower CapEx → More Noise

🛠️ Lower complexity and expertise needed to operate

🧹 Remove the need for manual cleanup

Automation using AI

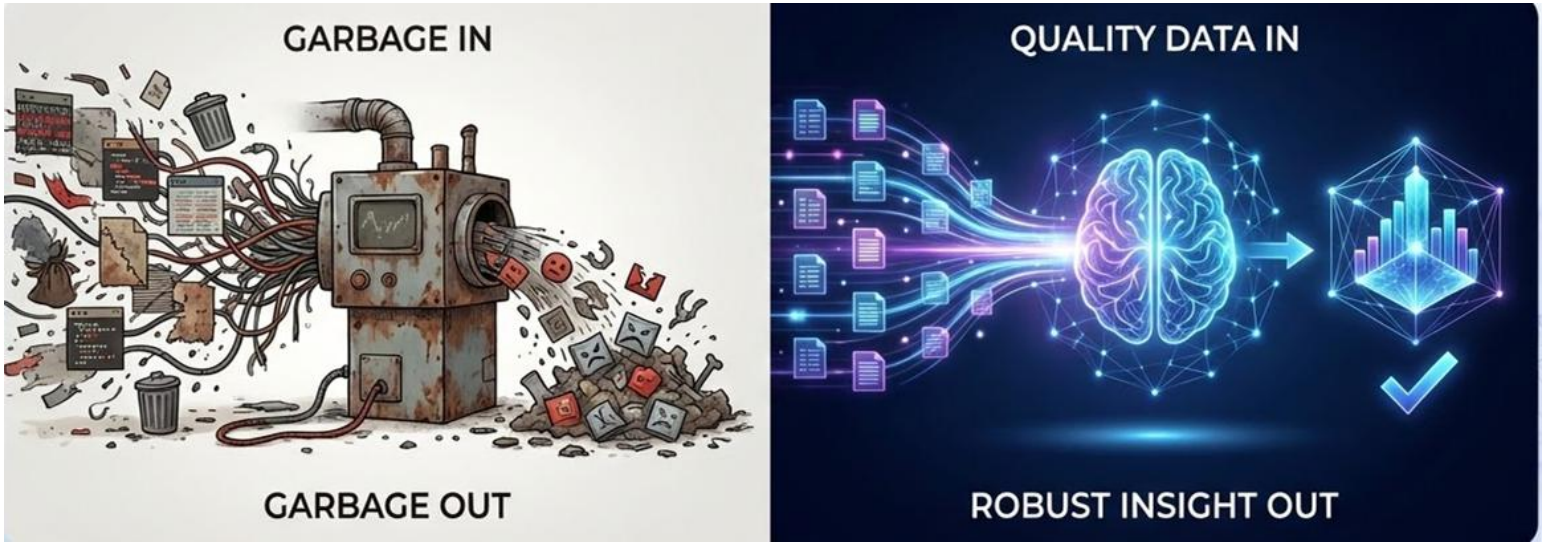
🚫📍 Move to a markerless solution

🚫📍

→ Improved quality & Maturation

Adversary I – Data Bias

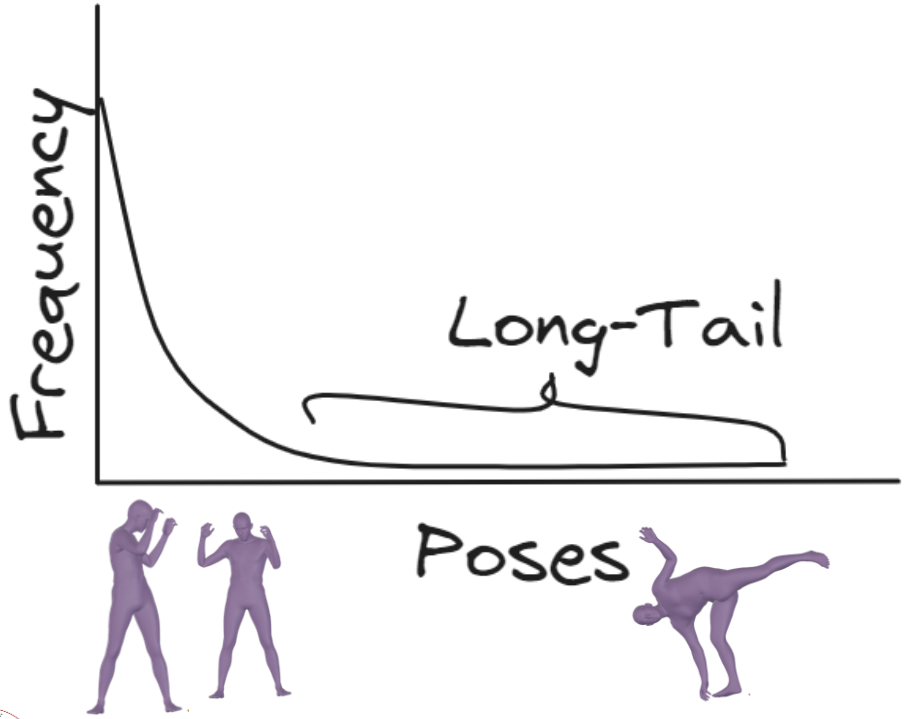
AI Models vs Data

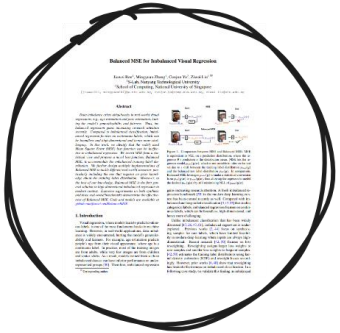


Challenges: Data Redundancy

Motion cycle contains similar parts

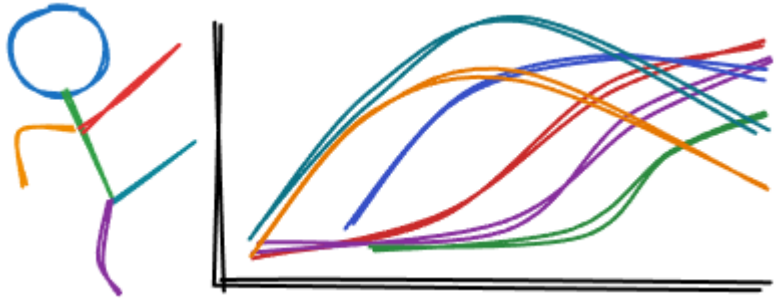
Challenges: Long-tail distribution

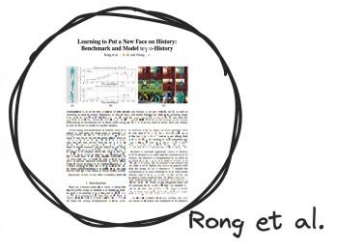




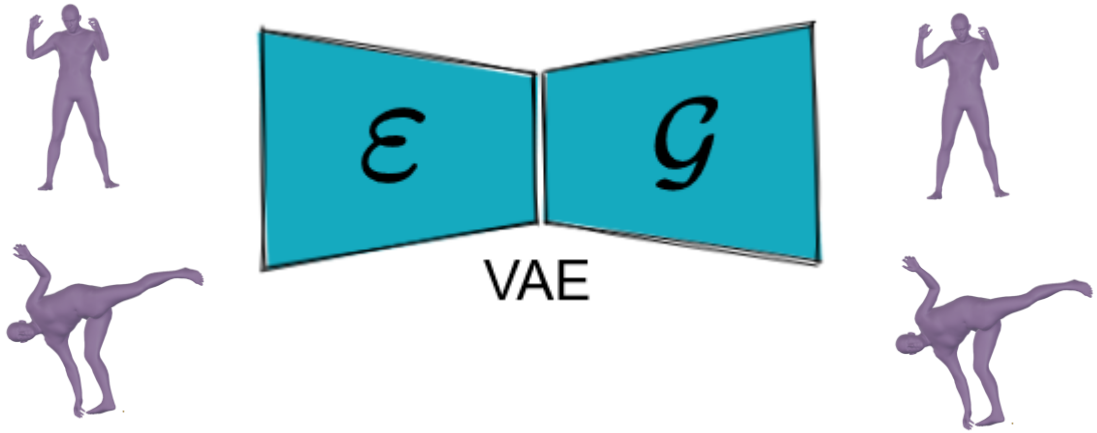
Ren et al.

Challenges: High-dimensionality



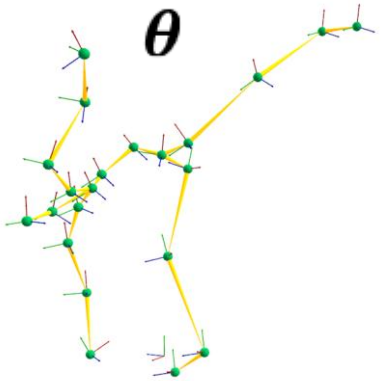


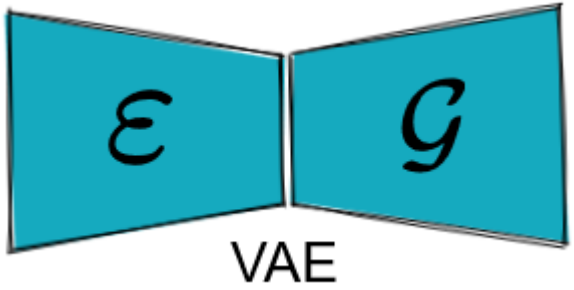
Representation Learning

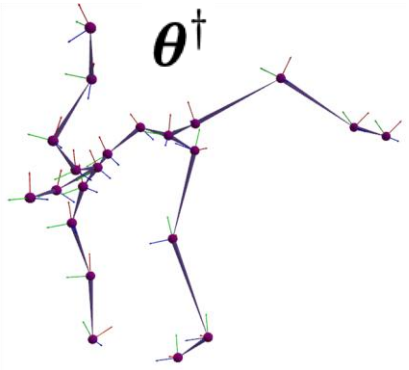


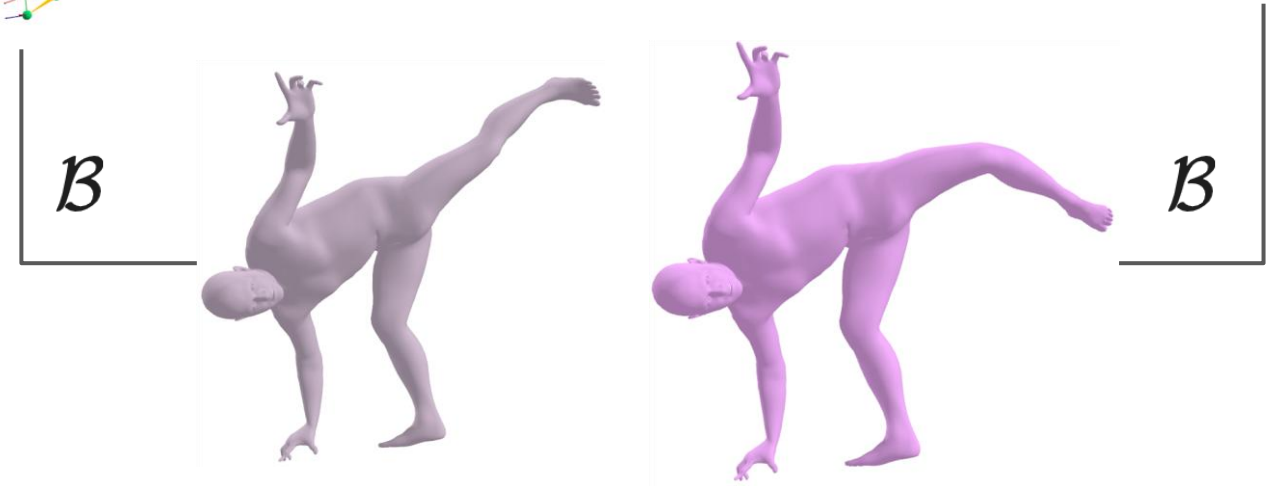
Reflect the bias of the data

Relevance via Reconstructability

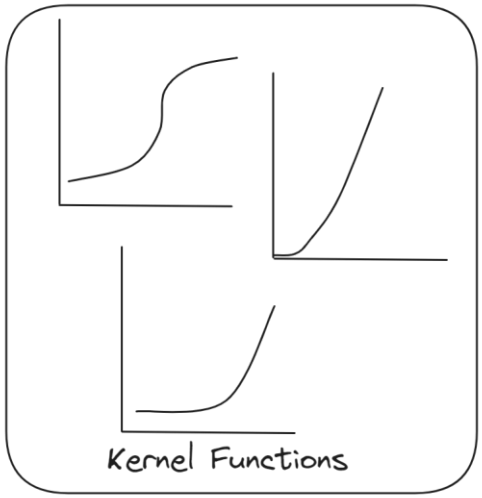






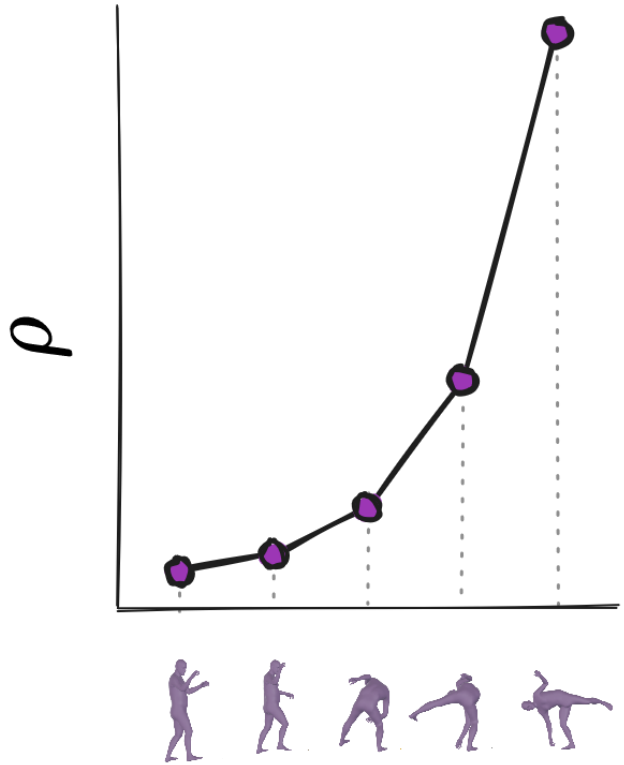


Relevance via Reconstructability



Relevance / Importance

Relevance via Reconstructability

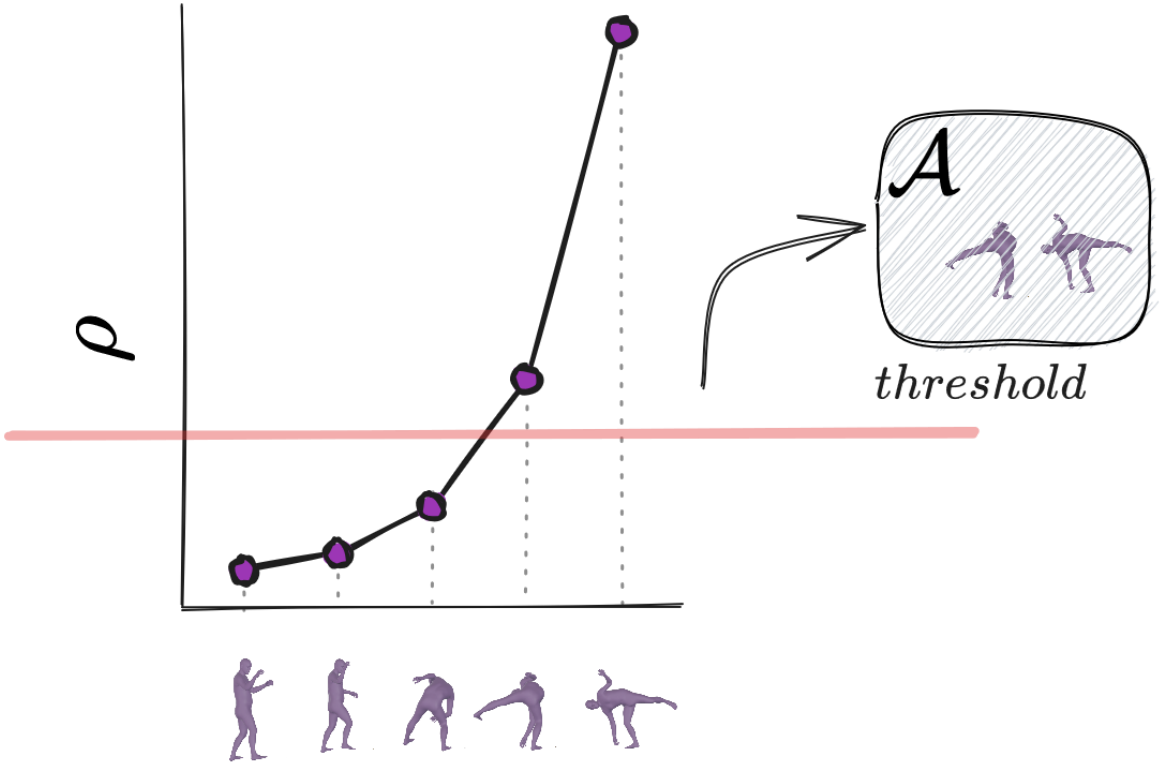


$$\rho * (\mathcal{L}_{JS} + \mathcal{L}_{w}^{\nu}) = \mathcal{L}$$

Overweight the tail samples while training

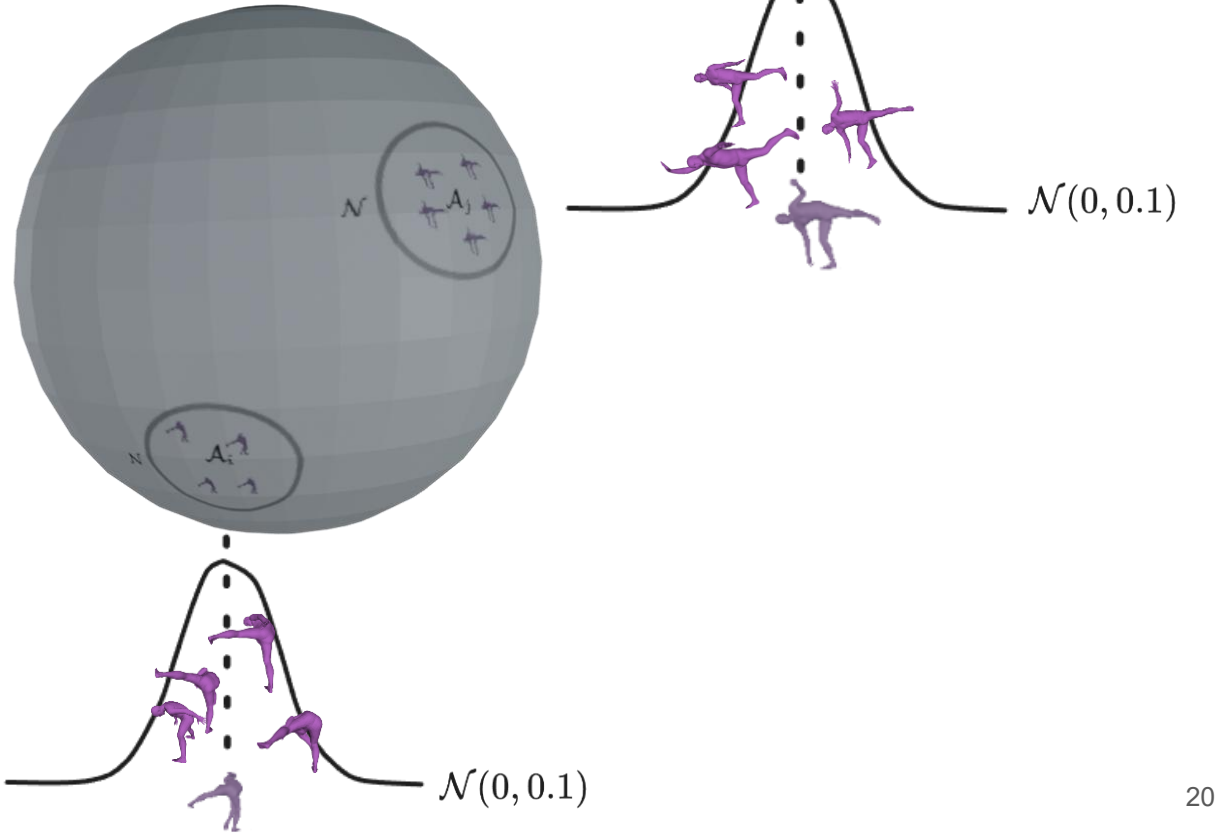
Oversampling using Synthesis

Oversampling using Synthesis

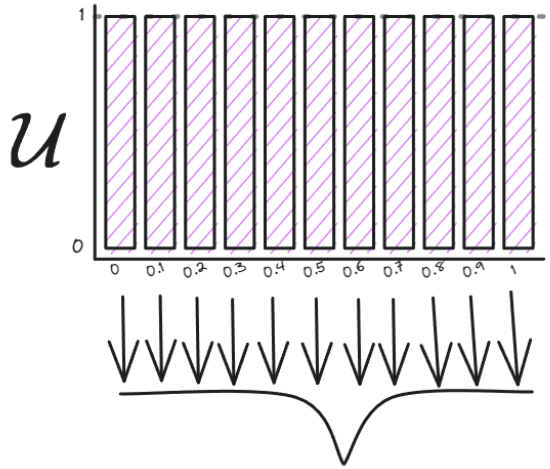


Oversampling using Synthesis

Sample the latent
space
around anchors

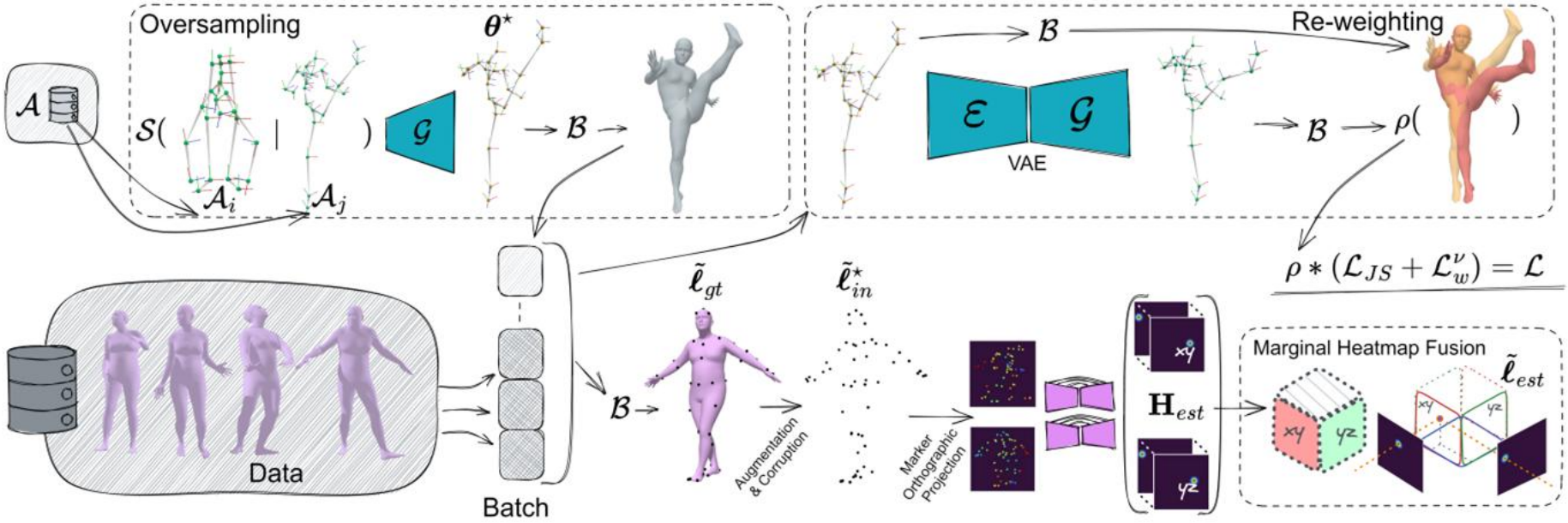


Oversampling using Synthesis



$$S(\quad,\quad, b) = \quad$$

Training Framework



Results - TAIL dataset

		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	[125]	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	[13]	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.10%	94.92%	93.10%	95.10%
sampling	RANDOM	35.80 *mm*	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.20%	95.70%	93.01%	95.30%
sampling	LERP	33.50 *mm*	28.19 *mm*	25.02%	12.30%	23.22%	79.82%	76.31%	81.10%	95.22%	93.05%	95.10%
sampling	SLERP	33.25 *mm*	25.82 *mm*	24.83%	14.23%	30.29%	80.85%	77.08%	82.16%	95.48%	93.18%	**95.75%**
sampling	SQUAD	**32.68** *mm*	25.53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.78%**	**84.15%**	95.55%	93.72%	94.63%
sampling	BEZIER	34.01 *mm*	26.26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	81.34%	**95.81%**	93.69%	95.66%
sampling	B-spline	34.82 *mm*	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *mm*	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.02%	94.95%	93.12%	95.01%
relevance	σ(R)	33.9 *mm*	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.10%	95.21%	93.02%	94.33%
relevance	e(D)	33.81 *mm*	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
relevance	σ(D)	**33.79** *mm*	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.40%	**95.21%**	92.60%	**95.45%**

Results THuman

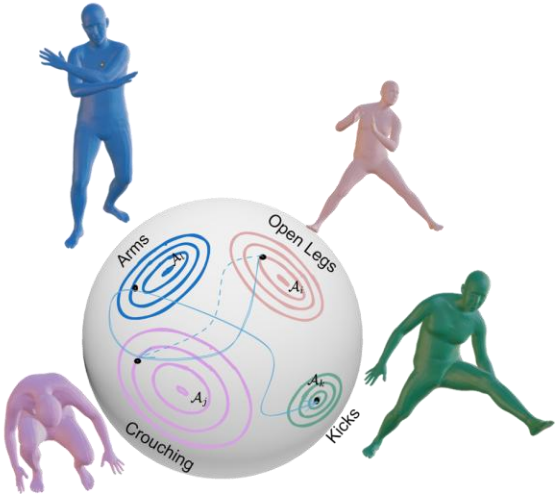
		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	[125]	22.01 *mm*	17.97 *mm*	27.15%	19.01%	26.10%	91.45%	87.43%	95.05%	98.01%	97.13%	98.01%
	[13]	21.81 *mm*	17.68 *mm*	28.10%	19.89%	26.55%	91.89%	87.12%	95.11%	98.55%	97.10%	98.49%
	Ours	21.41 *mm*	17.57 *mm*	28.69%	21.21%	27.33%	92.08%	87.50%	95.01%	98.59%	97.30%	98.89%
sampling	BMSE	22.21 *mm*	18.00 *mm*	25.51%	19.55%	25.02%	91.90%	86.10%	95.05%	98.62%	97.05%	98.15%
sampling	RANDOM	21.52 *mm*	16.78 *mm*	31.60%	24.11%	30.11%	92.49%	88.10%	95.11%	98.11%	97.10%	98.95%
sampling	LERP	21.59 *mm*	16.80 *mm*	29.48%	21.11%	29.07%	92.68%	88.25%	95.15%	98.58%	97.89%	99.01%
sampling	SLERP	20.43 *mm*	16.29 *mm*	30.41%	22.62%	29.02%	93.67%	88.40%	95.32%	98.92%	98.04%	99.05%
sampling	SQUAD	**18.80** *mm*	**15.33** *mm*	32.94%	25.03%	31.20%	**94.81%**	89.10%	95.41%	**99.19%**	**98.20%**	**99.31%**
sampling	BEZIER	18.94 *mm*	15.51 *mm*	32.90%	**25.03%**	31.23%	94.56%	**89.92%**	**96.15%**	99.04%	98.16%	99.22%
sampling	B-spline	20.20 *mm*	16.06 *mm*	**33.88%**	24.78%	**36.31%**	93.49%	88.47%	94.99%	98.82%	98.07%	98.81%
relevance	e(R)	20.65 *mm*	16.99 *mm*	31.11%	22.1%	28.15%	93.01%	88.5%	95.34%	98.55%	97.45%	98.89%
relevance	σ(R)	20.60 *mm*	16.67 *mm*	30.99%	22.4%	28.32%	92.79%	**88.65%**	**95.5%**	98.61%	97.65%	98.91%
relevance	e(D)	20.83 *mm*	16.59 *mm*	31.5%	**25.54%**	29.18%	**93.39%**	88.56%	94.75%	98.69%	**97.72%**	98.76%
relevance	σ(D)	**20.49** *mm*	**16.51** *mm*	**31.57%**	23.26%	**29.35%**	92.72%	86.63%	94.84%	**98.72%**	97.62%	**98.95%**

Best Sampling

		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	[125]	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	[13]	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.10%	94.92%	93.10%	95.10%
	RANDOM	35.80 *mm*	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.20%	95.70%	93.01%	95.30%
	LERP	33.50 *mm*	28.19 *mm*	25.02%	12.30%	23.22%	79.82%	76.31%	81.10%	95.22%	93.05%	95.10%
	SLERP	33.25 *mm*	25.82 *mm*	24.83%	14.23%	30.29%	80.85%	77.08%	82.16%	95.48%	93.18%	**95.75%**
	SQUAD	**32.68** *mm*	25.53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.78%**	**84.15%**	95.55%	93.72%	94.63%
	BEZIER	34.01 *mm*	26.26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	81.34%	**95.81%**	93.69%	95.66%
	B-spline	34.82 *mm*	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *mm*	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.02%	94.95%	93.12%	95.01%
	σ(R)	33.9 *mm*	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.10%	95.21%	93.02%	94.33%
	e(D)	33.81 *mm*	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
	σ(D)	**33.79** *mm*	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.40%	**95.21%**	92.60%	**95.45%**

Best Sampling

$SLERP(\mathcal{A}_i, \mathcal{A}_j)$

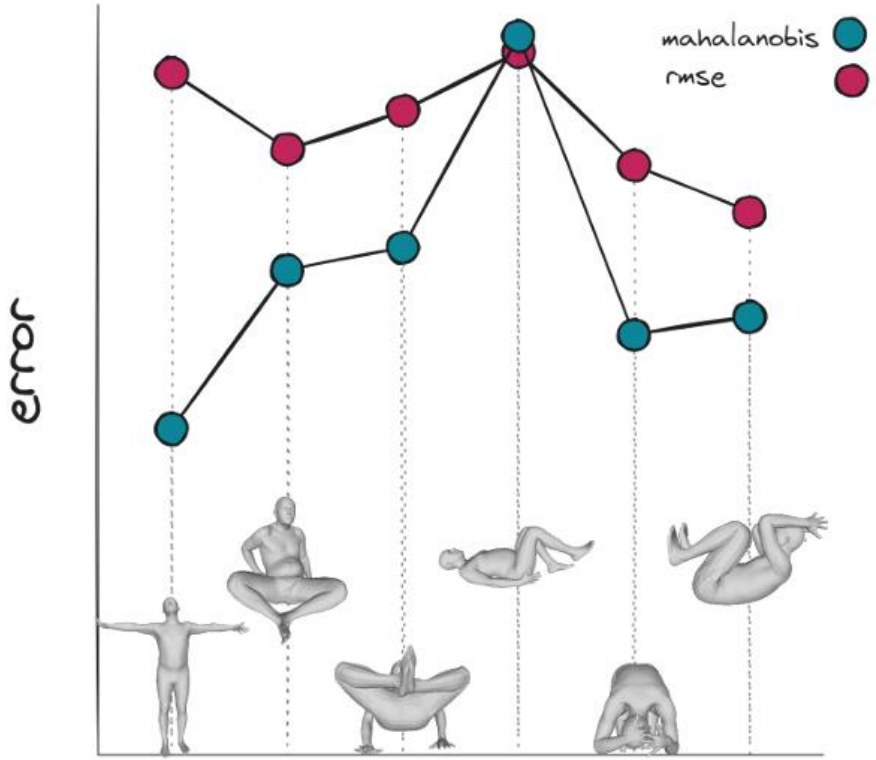


$SQUAD(\mathcal{A}_i, \mathcal{A}_j, \mathcal{A}_k, \mathcal{A}_l)$

Best Relevance

		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	[125]	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	[13]	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.10%	94.92%	93.10%	95.10%
sampling	RANDOM	35.80 *mm*	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.20%	95.70%	93.01%	95.30%
sampling	LERP	33.50 *mm*	28.19 *mm*	25.02%	12.30%	23.22%	79.82%	76.31%	81.10%	95.22%	93.05%	95.10%
sampling	SLERP	33.25 *mm*	25.82 *mm*	24.83%	14.23%	30.29%	80.85%	77.08%	82.16%	95.48%	93.18%	**95.75%**
sampling	SQUAD	**32.68** *mm*	25.53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.78%**	**84.15%**	95.55%	93.72%	94.63%
sampling	BEZIER	34.01 *mm*	26.26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	81.34%	**95.81%**	93.69%	95.66%
sampling	B-spline	34.82 *mm*	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *mm*	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.02%	94.95%	93.12%	95.01%
relevance	σ(R)	33.9 *mm*	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.10%	95.21%	93.02%	94.33%
relevance	e(D)	33.81 *mm*	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
relevance	σ(D)	**33.79** *mm*	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.40%	**95.21%**	92.60%	**95.45%**

Best Relevance



Synergistically

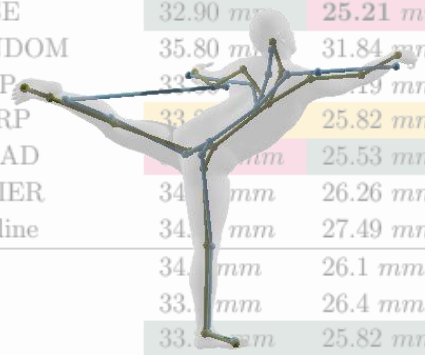
		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	[125]	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	[13]	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.10%	94.92%	93.10%	95.10%
	RANDOM	35.80 *mm*	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.20%	95.70%	93.01%	95.30%
	LERP	33.50 *mm*	28.19 *mm*	25.02%	12.30%	23.22%	79.82%	76.31%	81.10%	95.22%	93.05%	95.10%
	SLERP	33.25 *mm*	25.82 *mm*	24.83%	14.23%	30.29%	80.85%	77.08%	82.16%	95.48%	93.18%	**95.75%**
	SQUAD	**32.68** *mm*	25.53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.78%**	**84.15%**	95.55%	93.72%	94.63%
	BEZIER	34.01 *mm*	26.26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	81.34%	**95.81%**	93.69%	95.66%
	B-spline	34.82 *mm*	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *mm*	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.02%	94.95%	93.12%	95.01%
	σ(R)	33.9 *mm*	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.10%	95.21%	93.02%	94.33%
	e(D)	33.81 *mm*	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
	σ(D)	**33.79** *mm*	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.40%	**95.21%**	92.60%	**95.45%**
ortho	e(D) + SQUAD	34.06 *mm*	26.59 *mm*	25.21%	14.28%	30.16%	82.78%	78.21%	83.33%	95.80%	**95.94%**	95.35%
	σ(D)+ SQUAD	**32.54** *mm*	**24.48** *mm*	**28.15%**	**18.90%**	29.68%	**83.89%**	**79.85%**	82.90%	95.54%	95.60%	94.33%

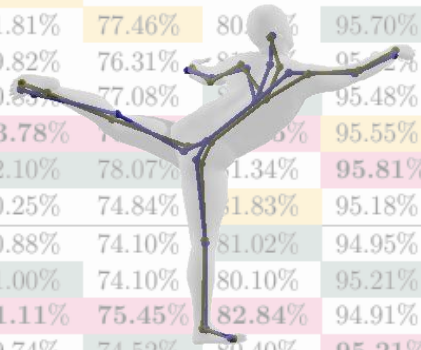
Results - TAIL dataset

		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	125	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	13	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.[illegible] *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.3[illegible]	[illegible].10%	94.92%	93.10%	95.10%
	RANDOM	35.8[illegible]	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46[illegible]	[illegible].20%	95.70%	93.01%	95.30%
	LERP	[illegible]	[illegible].19 *mm*	25.02%	12.30%	23.22%	79.82%	[illegible]	[illegible]%	95.22%	93.05%	95.10%
	SLERP	33.[illegible]	[illegible].82 *mm*	24.83%	14.23%	30.29%	[illegible]%	[illegible]	[illegible]%	95.48%	93.18%	**95.75%**
	SQUAD	**32.6**[illegible]	[illegible].53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.7**[illegible]	[illegible]	95.55%	93.72%	94.63%
	BEZIER	34.01	[illegible].26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	[illegible]%	**95.81%**	93.69%	95.66%
	B-spline	34.82	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	[illegible].83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *m*[illegible]	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	[illegible].02%	94.95%	93.12%	95.01%
	σ(R)	33.9 *m*[illegible]	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	[illegible].10%	95.21%	93.02%	94.33%
	e(D)	33.81 *m*[illegible]	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
	σ(D)	**33.79** *m*[illegible]	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	[illegible].40%	**95.21%**	92.60%	**95.45%**
ortho	e(D) + SQUAD	34.06 *mm*	26.59 *mm*	25.21%	14.28%	30.16%	82.78%	78.21%	83.33%	95.80%	**95.94%**	95.35%
	σ(D)+ SQUAD	**32.54** *mm*	**24.48** *mm*	**28.15%**	**18.90%**	29.68%	**83.89%**	**79.85%**	82.90%	95.54%	95.60%	94.33%

Results - TAIL dataset

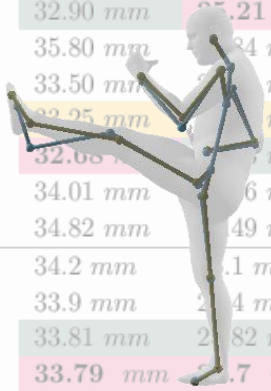
		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	125	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	13	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	**25.21** *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.10%	94.92%	93.10%	95.10%
	RANDOM	35.80 *mm*	31.84 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.[illegible]%	95.70%	93.01%	95.30%
	LERP	33.[illegible] *mm*	[illegible] *mm*	25.02%	12.30%	23.22%	79.82%	76.31%	[illegible]%	[illegible]%	93.05%	95.10%
	SLERP	33.[illegible] *mm*	25.82 *mm*	24.83%	14.23%	30.29%	80.[illegible]%	77.08%	[illegible]%	95.48%	93.18%	**95.75%**
	SQUAD	[illegible] *mm*	25.53 *mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	[illegible]%	[illegible]%	95.55%	93.72%	94.63%
	BEZIER	34.[illegible] *mm*	26.26 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	81.34%	**95.81%**	93.69%	95.66%
	B-spline	34.[illegible] *mm*	27.49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.83%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.[illegible] *mm*	26.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.02%	94.95%	93.12%	95.01%
	σ(R)	33.[illegible] *mm*	26.4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.10%	95.21%	93.02%	94.33%
	e(D)	33.[illegible] *mm*	25.82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.84%**	94.91%	**93.62%**	94.93%
	σ(D)	**33.79** *mm*	**25.7** *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.40%	**95.21%**	92.60%	**95.45%**
ortho	e(D) + SQUAD	34.06 *mm*	26.59 *mm*	25.21%	14.28%	30.16%	82.78%	78.21%	83.33%	95.80%	**95.94%**	95.35%
	σ(D)+ SQUAD	**32.54** *mm*	**24.48** *mm*	**28.15%**	**18.90%**	29.68%	**83.89%**	**79.85%**	82.90%	95.54%	95.60%	94.33%

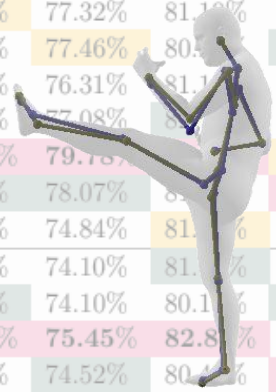




Results - TAIL dataset

		RMSE ↓	MPJPE ↓	PCK1 ↑			PCK3 ↑			PCK7 ↑		
				all	arms	legs	all	arms	legs	all	arms	legs
	125	36.10 *mm*	32.60 *mm*	20.10%	12.13%	23.23%	80.23%	77.10%	80.33%	93.12%	93.43%	94.92%
	13	36.05 *mm*	32.11 *mm*	21.01%	13.41%	24.53%	80.12%	77.43%	80.73%	93.23%	93.1%	95.12%
	Ours	35.80 *mm*	31.15 *mm*	22.04%	14.99%	25.68%	80.27%	77.94%	81.00%	94.31%	93.%	95.21%
sampling	BMSE	32.90 *mm*	25.21 *mm*	**27.66%**	16.21%	26.13%	81.98%	77.32%	81.1%	94.92%	93.10%	95.10%
	RANDOM	35.80 *mm*	34 *mm*	23.00%	14.56%	26.01%	81.81%	77.46%	80.%	95.70%	93.01%	95.30%
	LERP	33.50 *mm*	*mm*	25.02%	12.30%	23.22%	79.82%	76.31%	81.1%	95.22%	93.05%	95.10%
	SLERP	33.25 *mm*	*mm*	24.83%	14.23%	30.29%	80.85%	77.08%	8%	95.48%	93.18%	**95.75%**
	SQUAD	32.68 *mm*	*mm*	25.11%	**16.97%**	**30.35%**	**83.78%**	**79.78%**		95.55%	93.72%	94.63%
	BEZIER	34.01 *mm*	6 *mm*	24.38%	15.37%	26.14%	82.10%	78.07%	8%	**95.81%**	93.69%	95.66%
	B-spline	34.82 *mm*	49 *mm*	23.30%	14.53%	24.81%	80.25%	74.84%	81.%	95.18%	**94.02%**	95.23%
relevance	e(R)	34.2 *mm*	.1 *mm*	23.10%	12.01%	25.10%	80.88%	74.10%	81.%	94.95%	93.12%	95.01%
	σ(R)	33.9 *mm*	4 *mm*	23.61%	12.11%	26.55%	81.00%	74.10%	80.1%	95.21%	93.02%	94.33%
	e(D)	33.81 *mm*	82 *mm*	**25.53%**	**18.32%**	**28.67%**	**81.11%**	**75.45%**	**82.8%**	94.91%	**93.62%**	94.93%
	σ(D)	**33.79** *mm*	.7 *mm*	22.63%	13.18%	27.27%	79.74%	74.52%	80.%	**95.21%**	92.60%	**95.45%**
ortho	e(D) + SQUAD	34.06 *mm*	26.59 *mm*	25.21%	14.28%	30.16%	82.78%	78.21%	83.33%	95.80%	**95.94%**	95.35%
	σ(D)+ SQUAD	**32.54** *mm*	**24.48** *mm*	**28.15%**	**18.90%**	29.68%	**83.89%**	**79.85%**	82.90%	95.54%	95.60%	94.33%





Results - Different Model Architecture

		RMSE ↓	MPJPE ↓	PCK1 ↑	PCK3 ↑	PCK7 ↑
TH2	[13]	21.81 *mm*	17.68 *mm*	28.10%	91.89%	98.55%
	[13] + BMSE	22.01 *mm*	18.05 *mm*	25.05%	91.79%	98.65%
	[13] + Ours	**19.95** *mm*	**16.45** *mm*	**32.30%**	**92.45%**	**98.70%**
TAIL	[13]	36.05 *mm*	32.11 *mm*	21.01%	80.12%	93.23%
	[13] + BMSE	33.34 *mm*	26.15 *mm*	**25.35%**	81.03%	93.52%
	[13] + Ours	**33.01** *mm*	**25.55** *mm*	24.99%	**82.75%**	**93.90%**
GENE	[13]	190.32 *mm*	105.33 *mm*	28.01%	59.92%	73.45%
	[13] + BMSE	150.15 *mm*	101.01 *mm*	28.50%	60.01%	74.65%
	[13] + Ours	**110.01** *mm*	**70.55** *mm*	**30.15%**	**62.75%**	**80.05%**

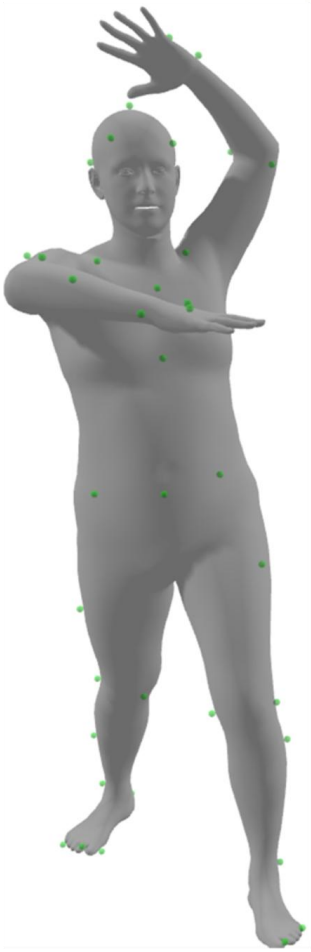
Results - Different VAE

		RMSE ↓	MPJPE ↓	PCK1 ↑	PCK3 ↑	PCK7 ↑
TH2	Benchmark	21.41 *mm*	17.57 *mm*	28.69%	92.08%	98.59%
	Benchmark + [4]	19.01 *mm*	16.55 *mm*	31.11%	93.33%	98.85%
	Benchmark + Ours	**18.69** *mm*	**15.20** *mm*	**35.59%**	**94.53%**	**99.09%**
TAIL	Benchmark	35.80 *mm*	31.15 *mm*	22.04%	80.27%	94.31%
	Benchmark + [4]	33.45 *mm*	26.01 *mm*	24.35%	82.03%	94.25%
	Benchmark + Ours	**32.54** *mm*	**24.48** *mm*	**25.15%**	**83.29%**	**95.54%**
GENE	Benchmark	170.31 *mm*	103.83 *mm*	29.04%	60.91%	73.85%
	Benchmark + [4]	125.15 *mm*	75.55 *mm*	30.01%	61.01%	75.65%
	Benchmark + Ours	**108.74** *mm*	**65.33** *mm*	**31.52%**	**62.85%**	**81.48%**

Adversary II – Noise

Model Noise

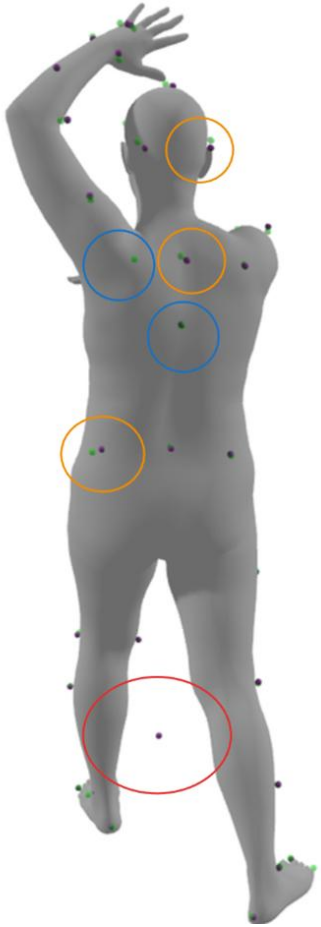
AI model -> information noise
(certain -> uncertain predictions)



Signal Noise

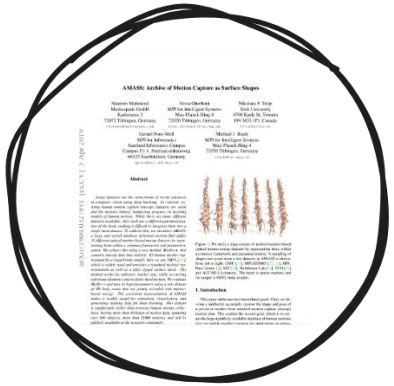
Low-cost sensors ->
noisy inputs & intense ghosting

Sparse views -> missing information



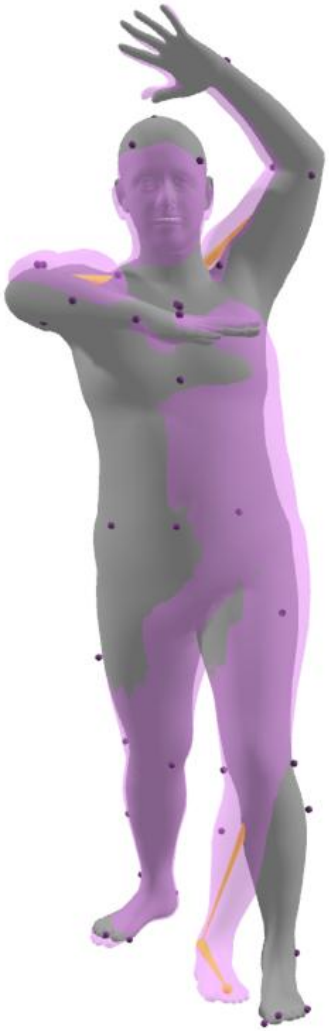
Non-linear optimization

Result is affected by the different sources of noise



Mahmood et al.

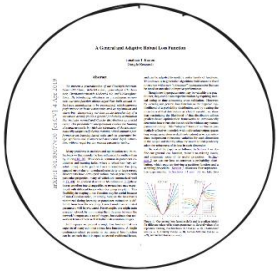
Loper et al.



Robust Estimators

Huber et al.

Hanson et al.



Barron et al.

- Known distribution
- Intensive tuning
- Require confidence knowledge

Noise-aware solver

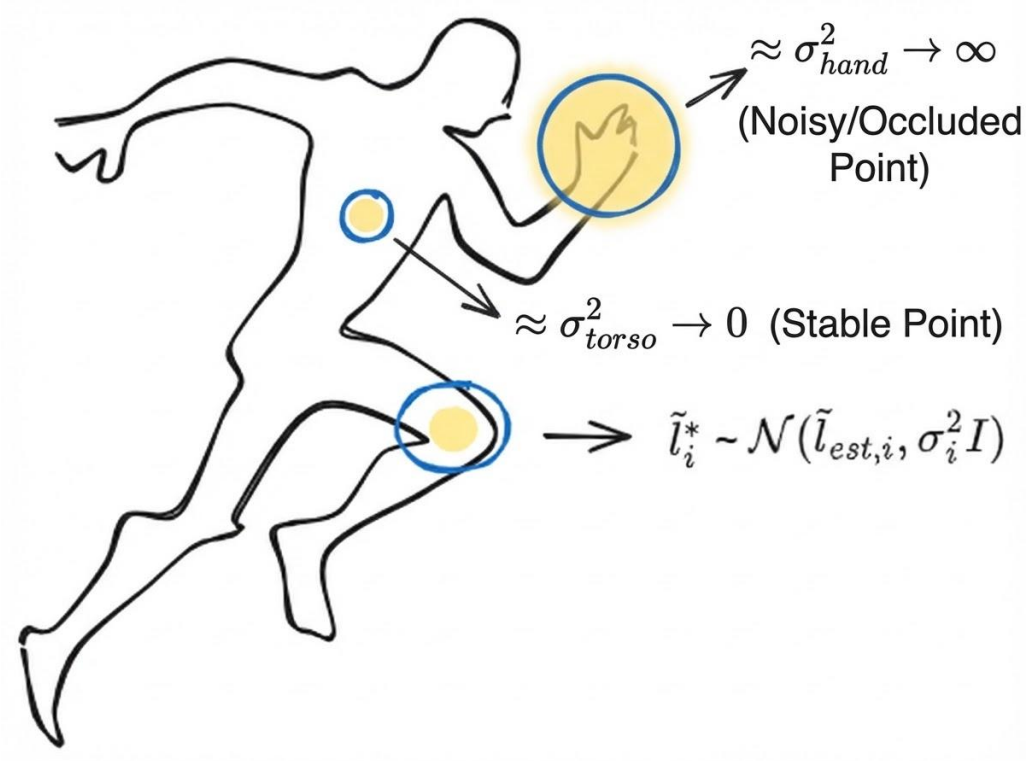
Adaptive fitting

- optimize for the observations' uncertainty region

$$E_{data} = \sum_{i}^{L} \frac{1}{2\sigma_i^2} \|\tilde{\ell}_{est,i} - \tilde{\ell}_i^*\|_2 + \log \sigma_i$$

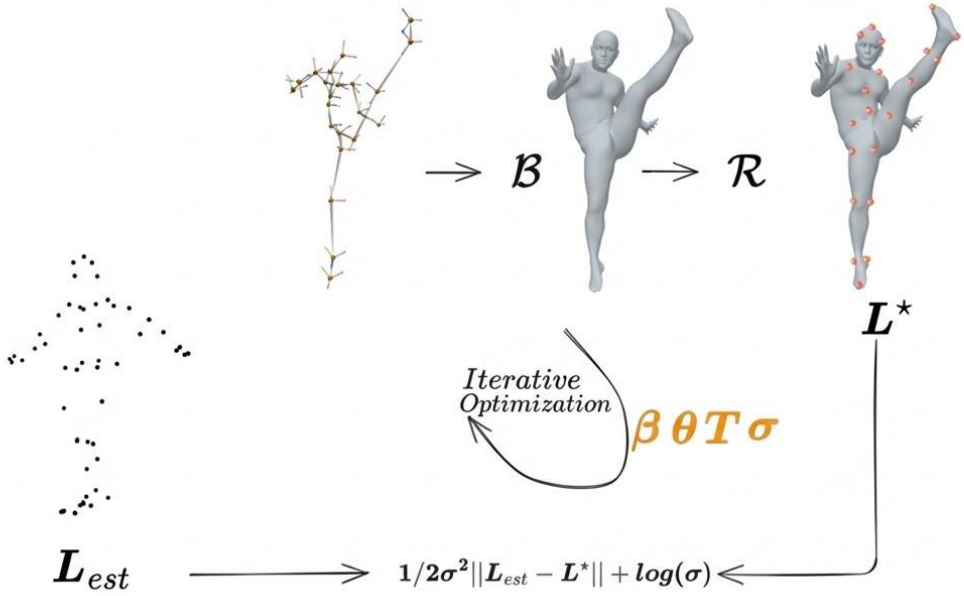
Different weights for each landmark

Noise Modelling

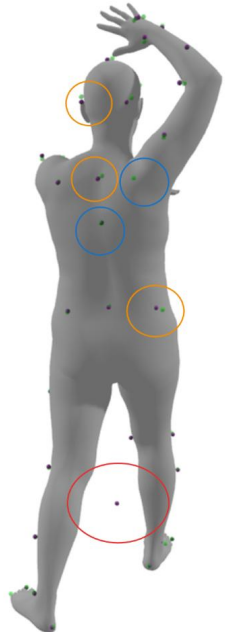


- Separate noise distribution
- Each landmark has its own variance

Noise-aware solver



Experiments

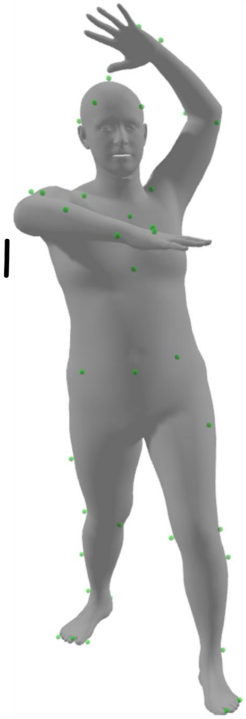


Data Noise n_d

- Ghost markers
- Occluded markers
- Measurement Noise as random offset to the markers

Inference Noise n_m

- Automatic Labeling from previous model

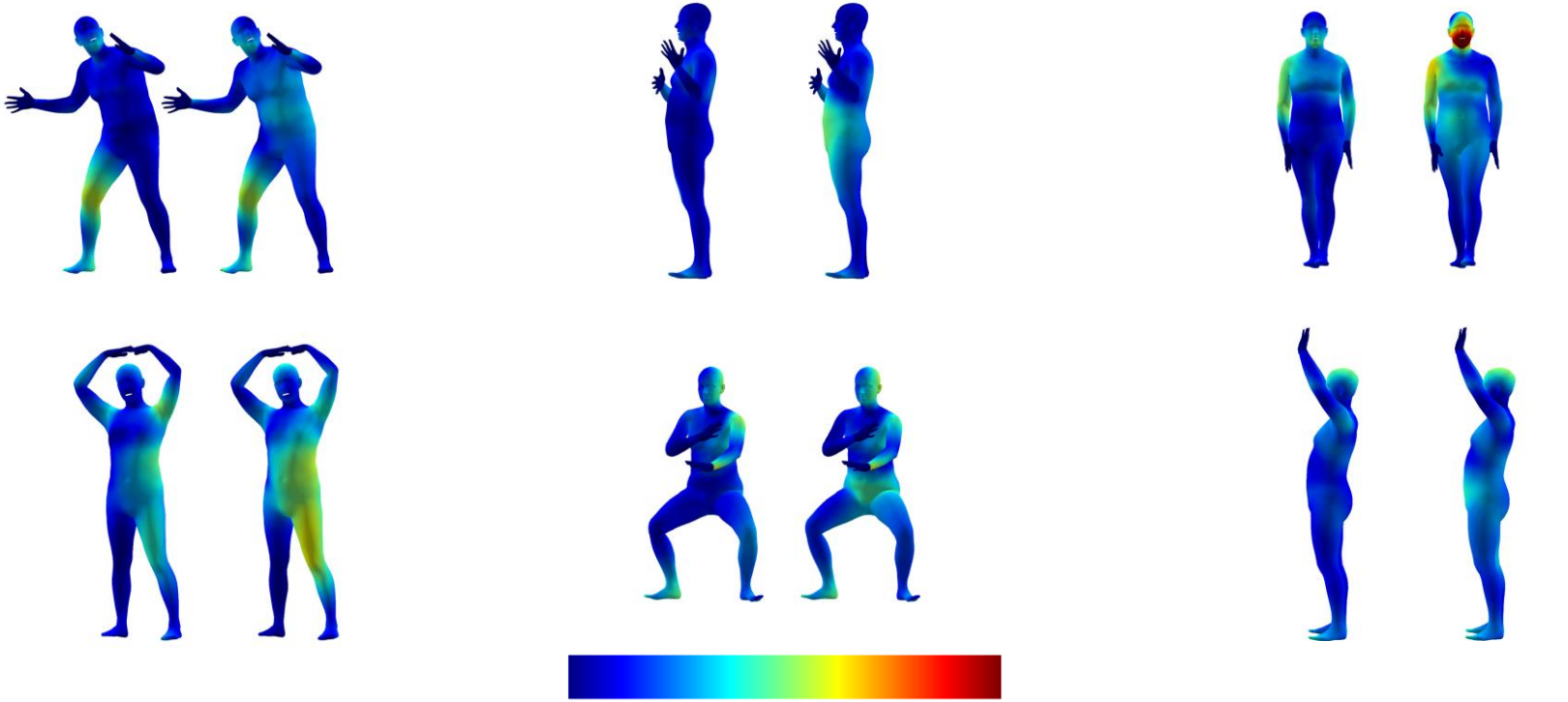


Results THuman

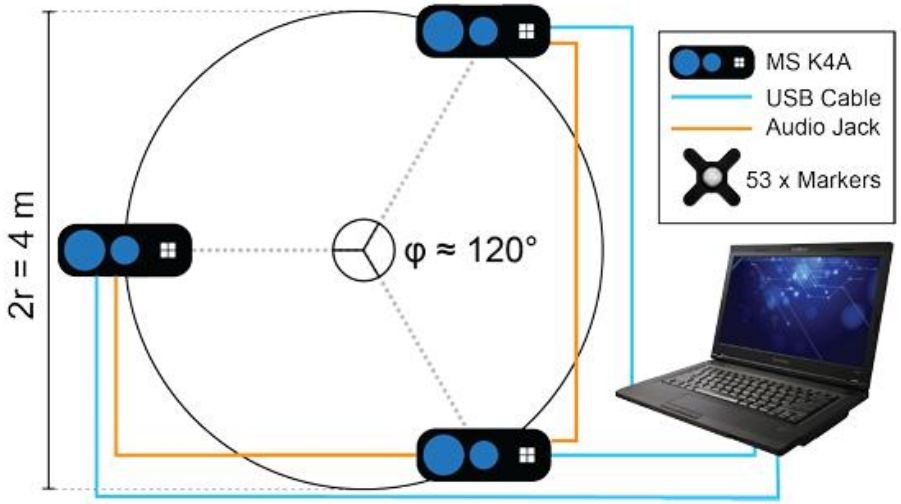
	n_d	n_m	RMSE ↓	MAE ↓	PCK1 ↑	PCK3 ↑	PCK7 ↑
[2, 5]	✓	✗	30.10 *mm*	3.49°	11.79%	66.85%	98.34%
[20]			30.80 *mm*	3.10°	12.71%	67.06%	97.71%
Ours (ℓ^m)			28.90 *mm*	2.98°	14.71%	69.86%	98.18%
Ours ($\ell^m \vert \ell^j$)			**23.40** *mm*	**2.29°**	**19.66%**	**81.06%**	**99.11%**
[2, 5]	✗	✓	20.60 *mm*	1.93°	28.71%	89.03%	**99.05%**
[20]			21.71 *mm*	1.91°	36.38%	87.75%	98.22%
Ours (ℓ^m)			18.70 *mm*	1.85°	41.99%	90.95%	98.81%
Ours ($\ell^m \vert \ell^j$)			**18.50** *mm*	**1.49°**	**42.18%**	**91.44%**	98.56%
[2, 5]	✓	✓	23.80 *mm*	2.03°	24.26%	85.63%	**98.22%**
[20]			24.87 *mm*	1.94°	31.99%	84.05%	97.00%
Ours (ℓ^m)			22.40 *mm*	1.79°	36.01%	87.14%	97.53%
Ours ($\ell^m \vert \ell^j$)			**21.90** *mm*	**1.52°**	**36.67%**	**88.09%**	97.69%

Table 5.2: Noisy landmark fitting on THuman 2.0. Comparison of our noise-aware fitting approach vs. a variant of the fitting method from [2, 5]. RMSE is in meters and PCK in (%). Subscripts j and m denote "joints" and "markers", while n_d and n_m indicate data and marker noise, respectively.

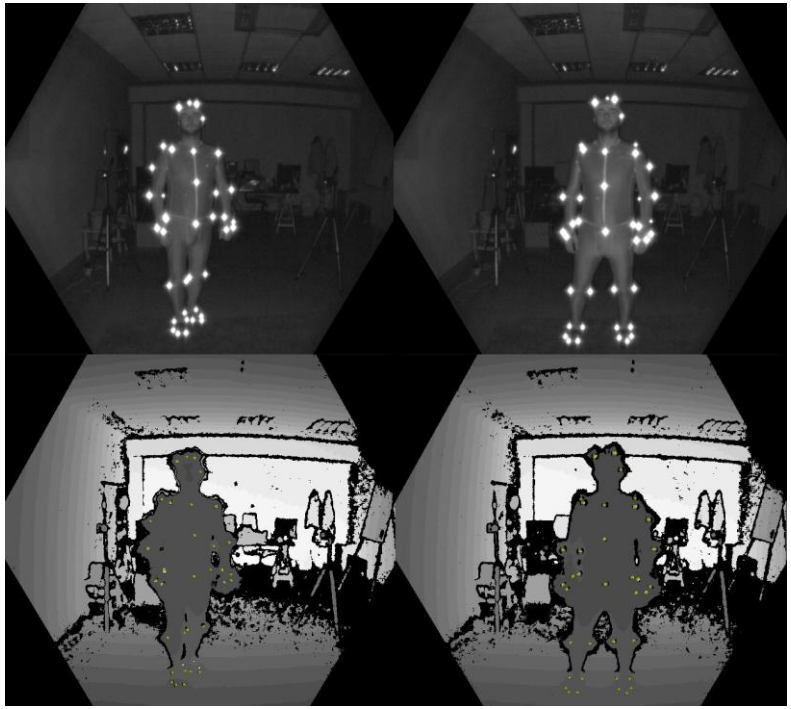
Results vs Mosh++



Real-World Deployment



Marker Detection



- Detect Markers in IR

- Ring Around Marker -> Centroid-> Output: (x, y, depth) for each marker ->Deproject to Sensor Space m_n^s

Marker Fusion

~133

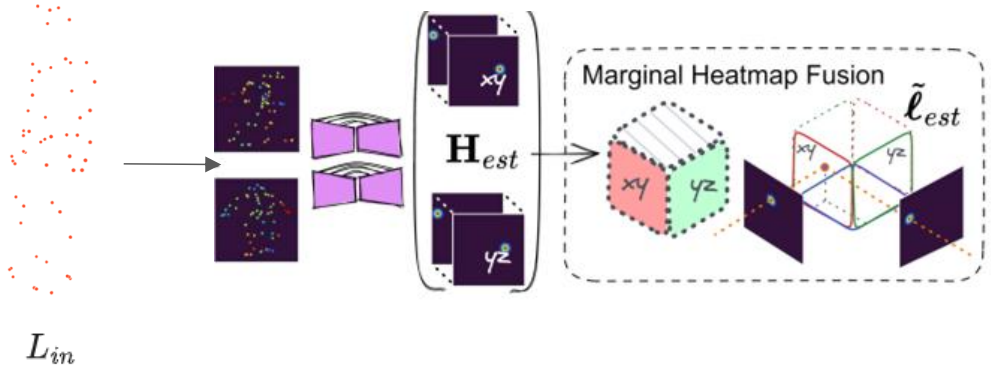
Fused in Global Space

m_n^G

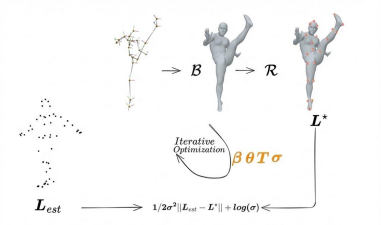
-> Cluster ->

Input Markers~57

Labelling & Solving

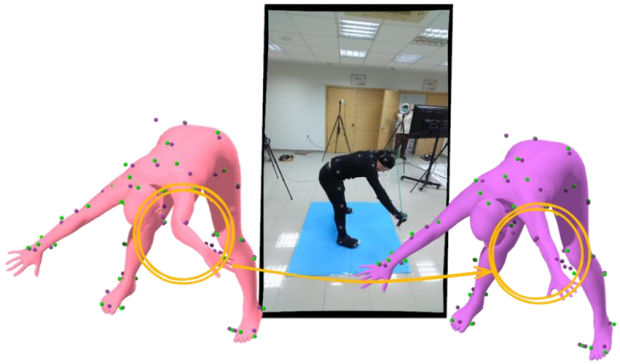


Labeled Markers & Joints

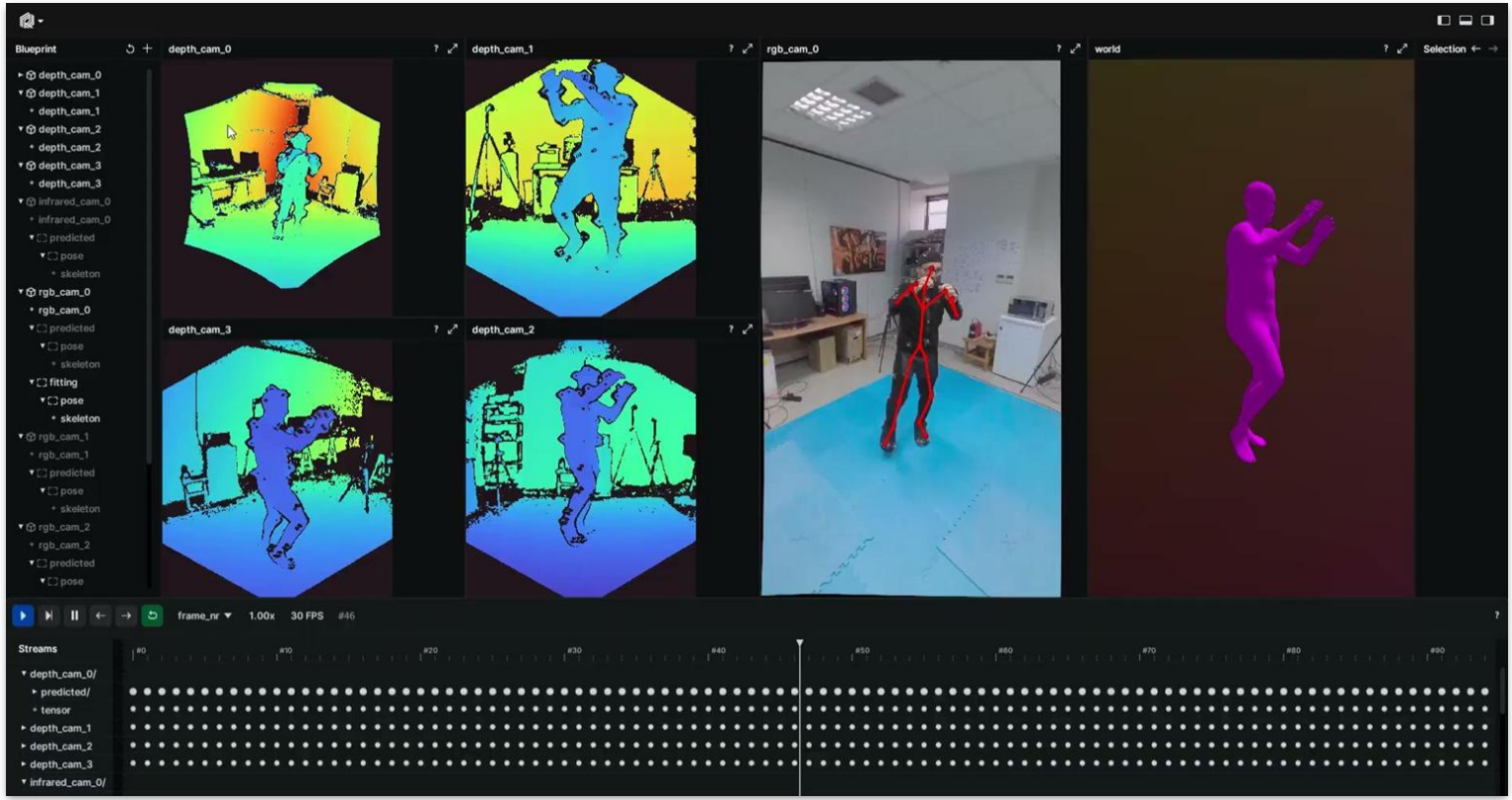


Noise Aware Solver

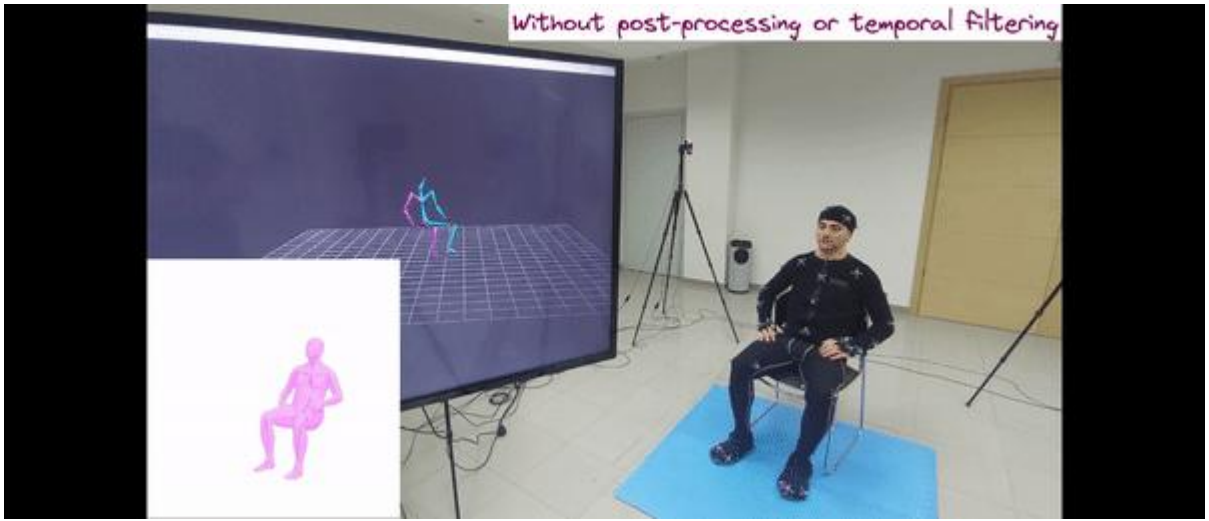
Ours vs Mosh++



In the wild results



In the wild results



ICCV Demo

ICCV23
PARIS

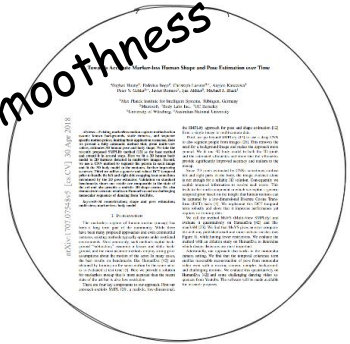
Adversary III - Markerless MoCap

Markerless Challenges

- ⚡ High-frequency jitter and sudden changes
- 🚫 Missing key points due to body part occlusions
- 🔀 Flipping body parts
- 🎞️ Inconsistent predictions across adjacent frames

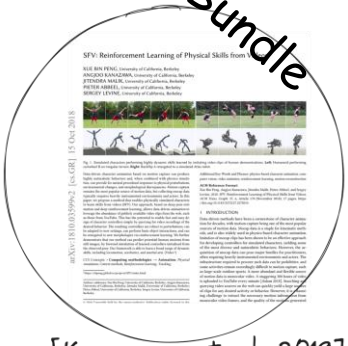
Related Work

DCT smoothness



[Huang et al. 2017]

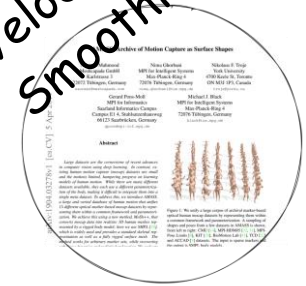
Bundle



[Kanazawa et al. 2019]

Velocity

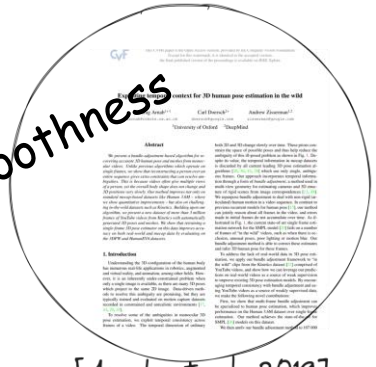
Smoothness



[Mahmood et al. 2019]

Joints smoothness

[Ye et al. 2022]



[Arnab et al. 2019]

- Temporal
- Prior constraints
- To enforce temporal consistency

Related Work

[Huang et al. 2017]

[Kanazawa et al. 2019]

[Mahmood et al. 2019]

Time consuming

[Ye et al. 2022]

[Arnab et al. 2019]

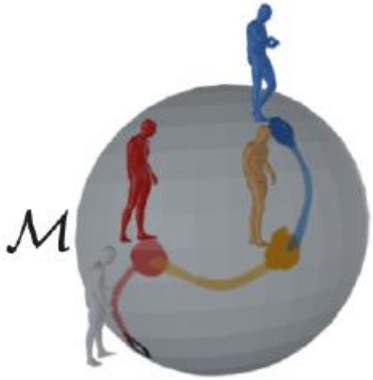
Require several Resources

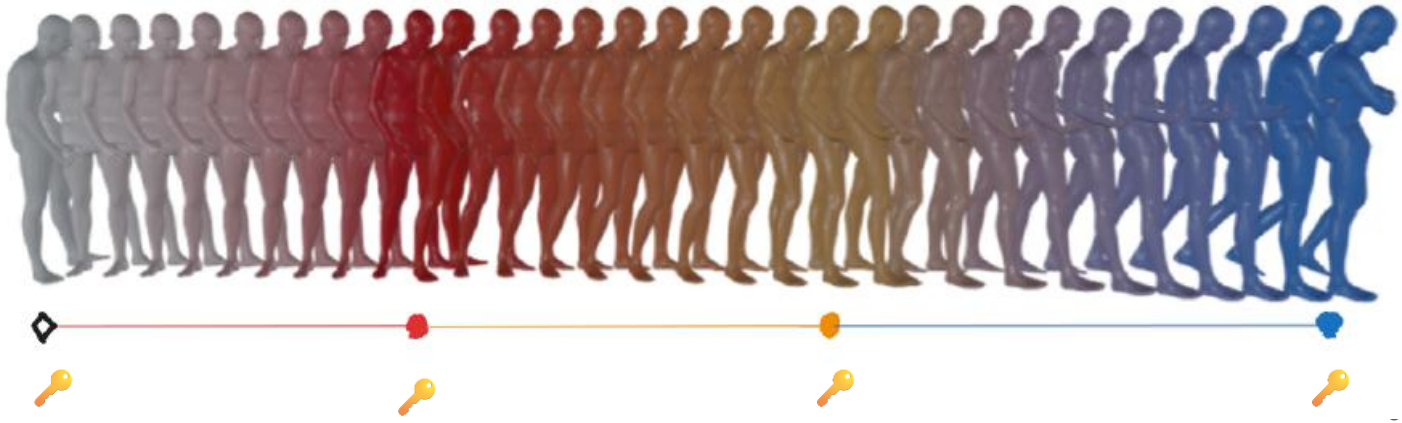
Temporal

Prior constraints

To enforce temporal consistency

BundleMoCap





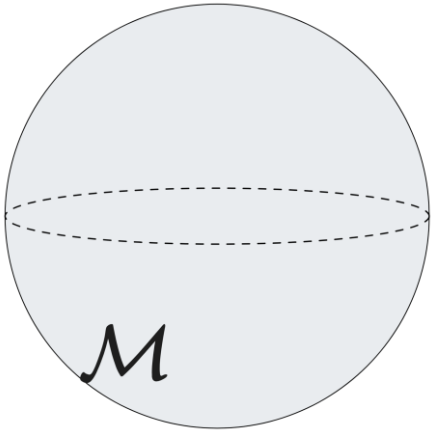
🔄 Gains robustness to outlier estimates

🕒 Produces smooth motions, even without smoothness terms

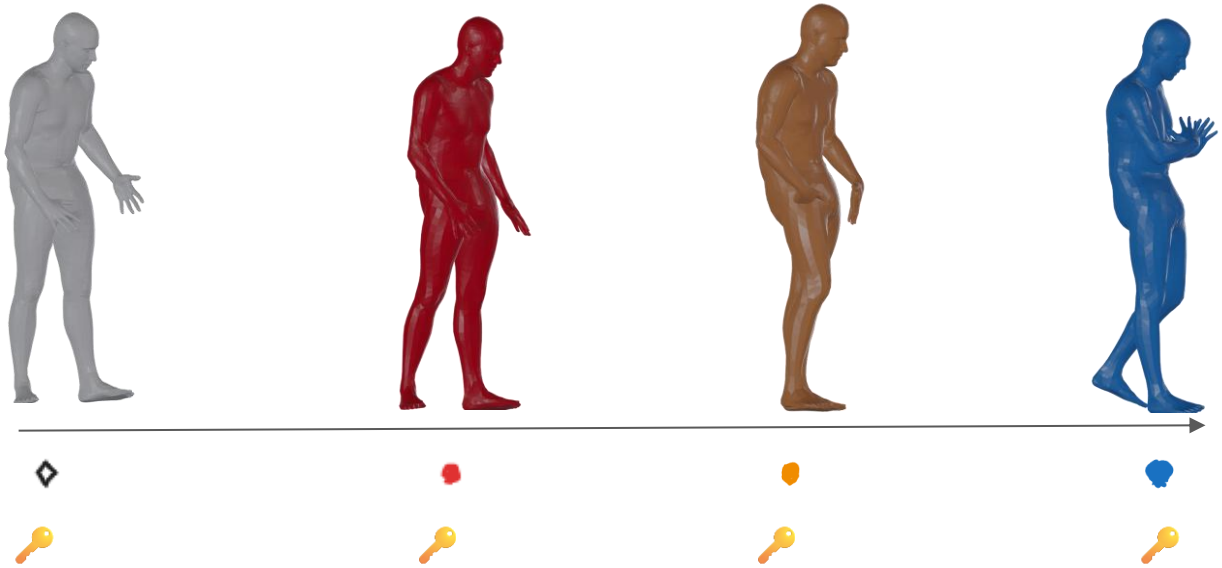
⚡ Maximizes runtime efficiency

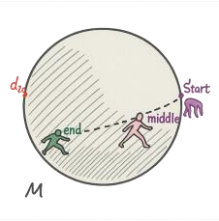
Spherical Human Pose Prior

📊 prior distribution over valid human poses

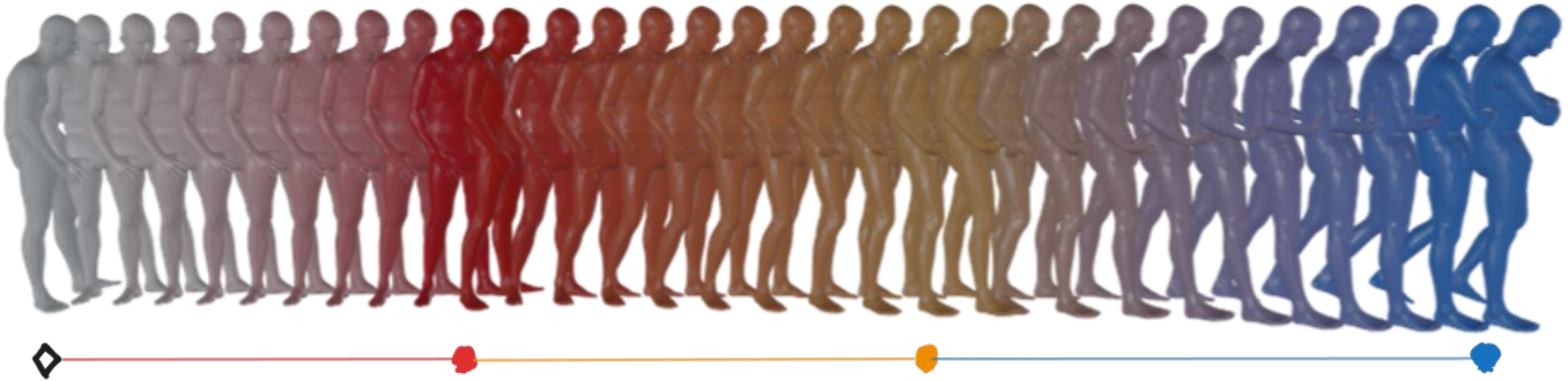


Pose Interpolation

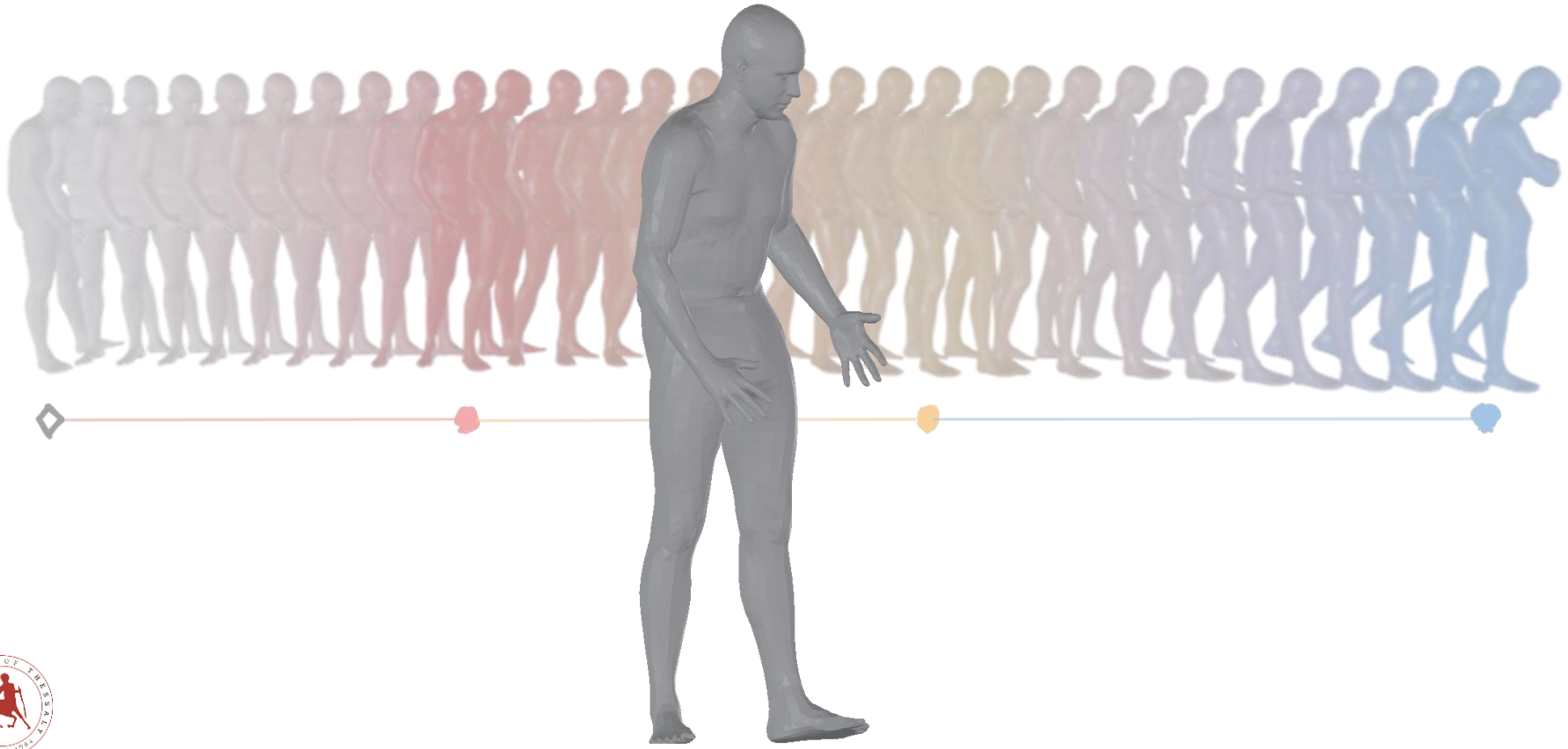




Pose Interpolation



Pose Interpolation



Approach

$t^1 = 0$

$t^1 = T$

$\mathcal{E}_{prior}^{T^1}$

Approach

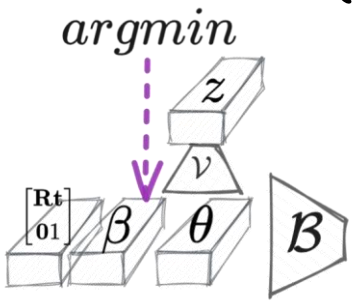
$t^1 = 0$

$t^1 = T$

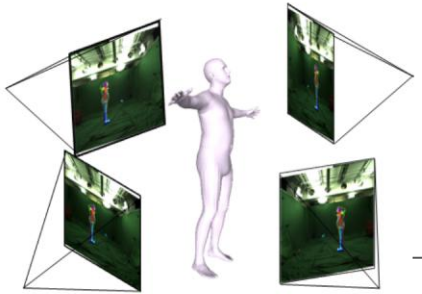
$$\theta^t = \mathcal{G}(\mathcal{S}^t(z^0, z^{T^1}))$$
$$\mathcal{R}^t = \mathcal{S}^t(\mathcal{R}^0, \mathcal{R}^{T^1})$$
$$t^t = \mathcal{L}^t(t^0, t^{T^1})$$

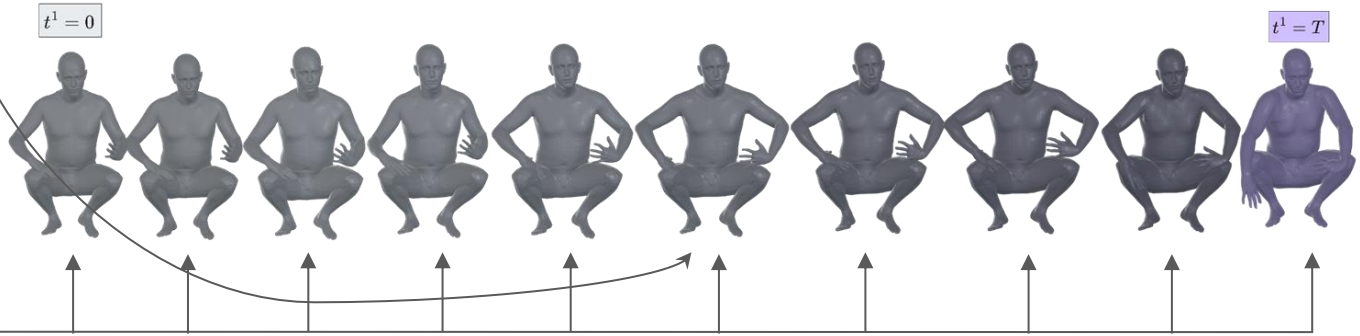
Approach

Constraint-Solved over the entire bundle



multi-view $2D$ keypoint constraints





$\mathcal{E}_{data}^{T}$

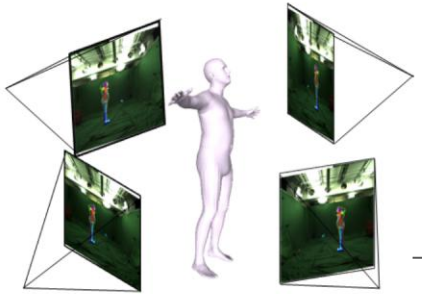
Sliding Windows

$t^1 = T$

$$\theta^t = \mathcal{G}(\mathcal{S}^t(z^0, z^{T^1}))$$
$$\mathcal{R}^t = \mathcal{S}^t(\mathcal{R}^0, \mathcal{R}^{T^1})$$
$$t^t = \mathcal{L}^t(t^0, t^{T^1})$$

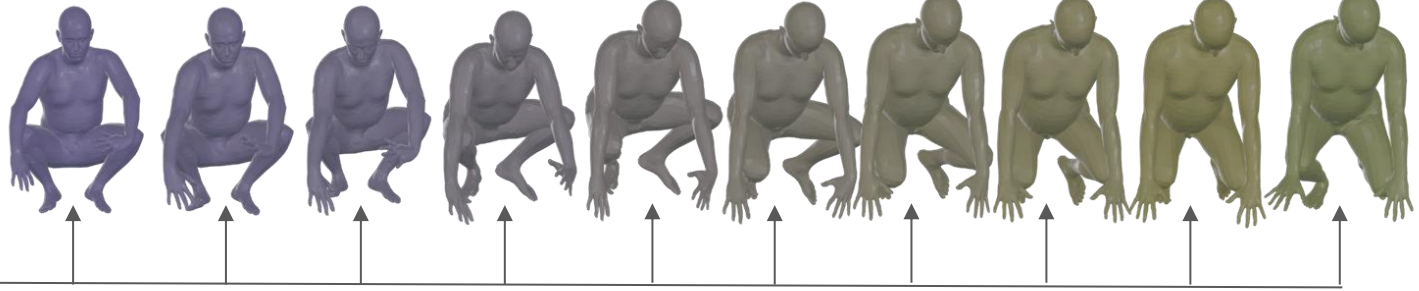
$\mathcal{E}^{T^2}_{prior}$

multi-view $2D$
keypoint constraints



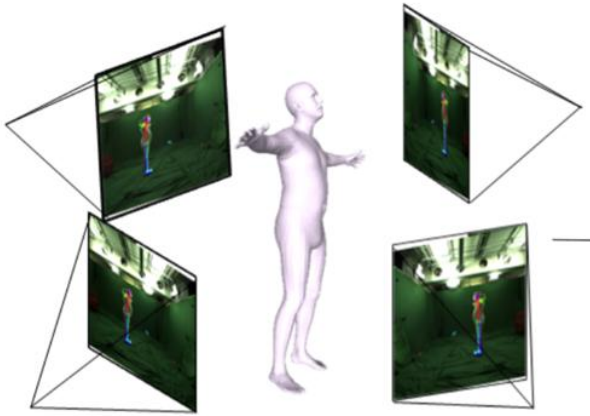
$t^1 = T$

$t^2 = T$

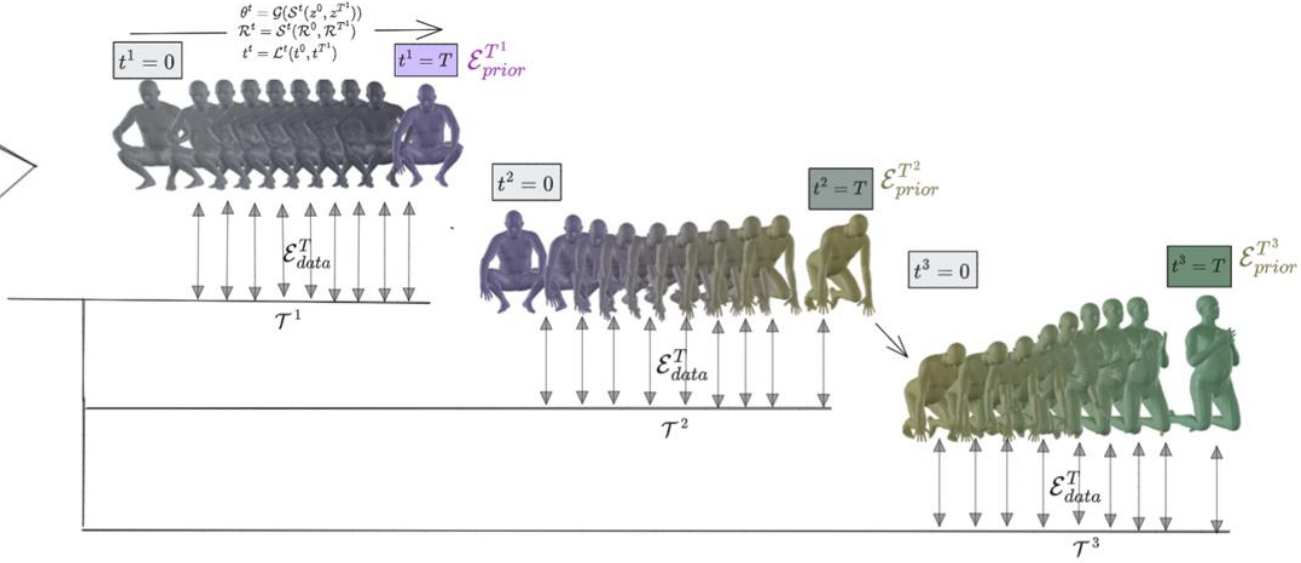


$\mathcal{E}^{T}_{data}$

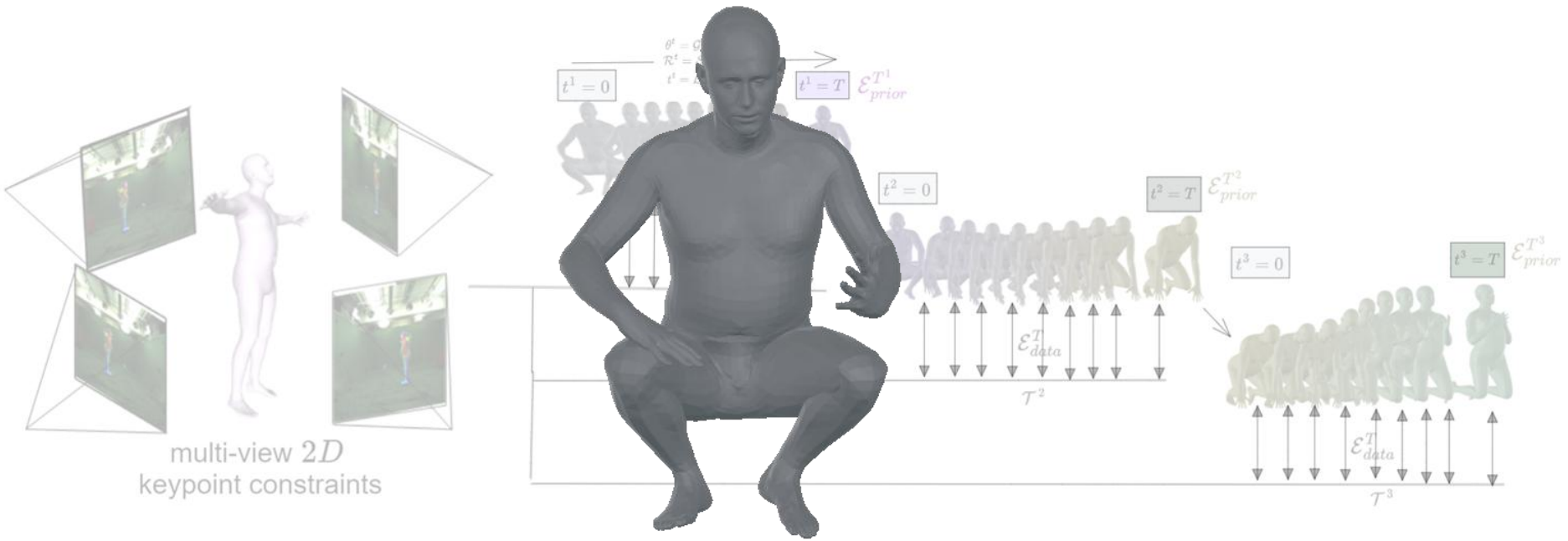
Overview



multi-view $2D$
keypoint constraints



Overview


multi-view 2D
keypoint constraints
$t^1 = 0$
$t^1 = T$
$\mathcal{E}^{T^1}_{prior}$
$t^2 = 0$
$t^2 = T$
$\mathcal{E}^{T^2}_{prior}$
$\mathcal{E}^{T}_{data}$
T^2
$t^3 = 0$
$t^3 = T$
$\mathcal{E}^{T^3}_{prior}$
$\mathcal{E}^{T}_{data}$
T^3

Results on standard benchmarks

	MPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
MuVS	53.36 *mm*	58.54 *mm*	10.23 °	28.02%	79.57%	10.99 mm/s^2
DCT [6]	50.88 *mm*	56.72 *mm*	12.32 °	29.21%	80.41%	09.19 mm/s^2
DMMR [9]	60.69 *mm*	65.16 *mm*	11.48 °	20.93%	69.48%	09.57 mm/s^2
SLAHMR [10]	50.49*mm*	54.52 *mm*	08.57 °	28.92%	79.20%	09.02 mm/s^2
ETC [7]	72.74 *mm*	77.83 *mm*	05.73°	32.63%	84.42%	07.92 mm/s^2
BundleMoCap	38.36 *mm*	43.10 *mm*	04.31°	33.70%	86.24%	06.18mm/s^2
BundleMoCap++	**36.32***mm*	**40.93***mm*	**04.11°**	**46.27%**	**93.77%**	**02.52**mm/s^2

	MPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
MuVS	64.99 *mm*	76.12 *mm*	6.28°	28.20%	73.75%	14.45 mm/s^2
DCT [6]	62.43 *mm*	68.13 *mm*	6.18°	35.84%	83.77%	12.01 mm/s^2
DMMR [9]	57.51 *mm*	67.66 *mm*	6.06°	37.52%	81.11%	11.52 mm/s^2
SLAHMR [10]	61.8 *mm*	62.55 *mm*	5.74°	40.86%	83.97%	11.34 mm/s^2
ETC [7]	59.51 *mm*	61.32 *mm*	5.64°	39.30%	84.50%	10.01 mm/s^2
BundleMoCap	56.41 *mm*	59.12 *mm*	5.43°	44.51%	85.71%	07.39 mm/s^2
BundleMoCap ++	**48.27** *mm*	**56.44** *mm*	**4.33°**	**48.75%**	**89.34%**	**05.66** mm/s^2

Results on standard benchmarks

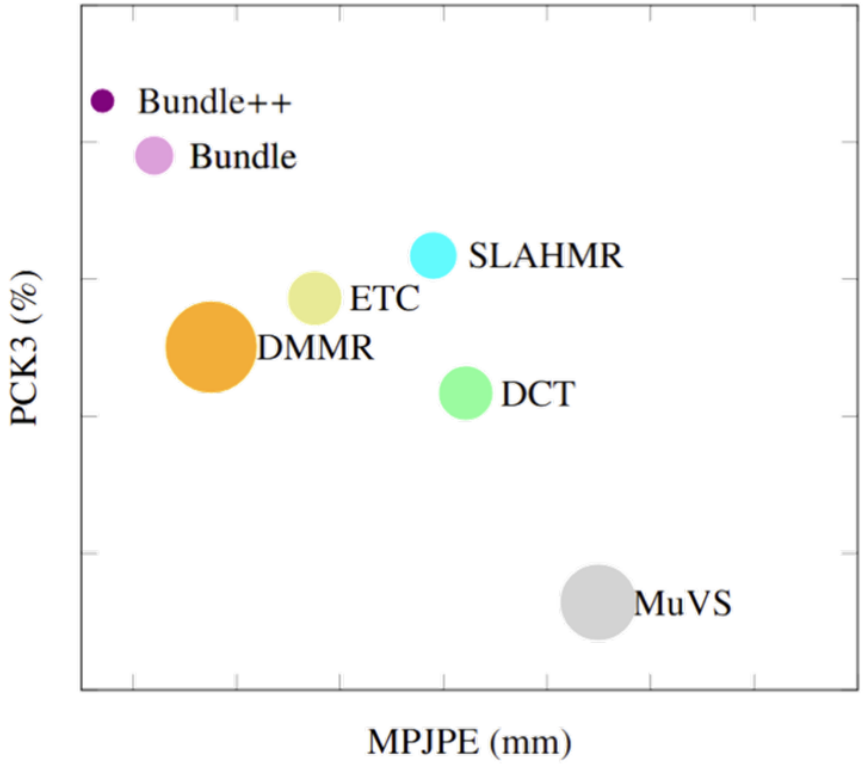
	MPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
MuVS	53.36 *mm*	58.54 *mm*	10.23 °	28.02%	79.57%	10.99 mm/s^2
DCT [6]	50.88 *mm*	56.72 *mm*	12.32 °	29.21%	80.41%	09.19 mm/s^2
DMMR [9]	60.69 *mm*	65.16 *mm*	11.48 °	20.93%	69.48%	09.57 mm/s^2
SLAHMR [10]	50.49*mm*	54.52 *mm*	08.57 °	28.92%	79.20%	09.02 mm/s^2
ETC [7]	72.74 *mm*	77.83 *mm*	05.73°	32.63%	84.42%	07.92 mm/s^2
BundleMoCap	38.36 *mm*	43.10 *mm*	04.31°	33.70%	86.24%	06.18mm/s^2
BundleMoCap++	**36.32***mm*	**40.93***mm*	**04.11°**	**46.27%**	**93.77%**	**02.52**mm/s^2

	MPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
MuVS	64.99 *mm*	76.12 *mm*	6.28°	28.20%	73.75%	14.45 mm/s^2
DCT [6]	62.43 *mm*	68.13 *mm*	6.18°	35.84%	83.77%	12.01 mm/s^2
DMMR [9]	57.51 *mm*	67.66 *mm*	6.06°	37.52%	81.11%	11.52 mm/s^2
SLAHMR [10]	61.8 *mm*	62.55 *mm*	5.74°	40.86%	83.97%	11.34 mm/s^2
ETC [7]	59.51 *mm*	61.32 *mm*	5.64°	39.30%	84.50%	10.01 mm/s^2
BundleMoCap	56.41 *mm*	59.12 *mm*	5.43°	44.51%	85.71%	07.39 mm/s^2
BundleMoCap ++	**48.27** *mm*	**56.44** *mm*	**4.33°**	**48.75%**	**89.34%**	**05.66** mm/s^2

Realism of the provided motion

	FS ↓
MuVS	0.09 *cm/f*
DCT	0.09 *cm/f*
DMMR	0.06 *cm/f*
SLAHMR	0.07 *cm/f*
ETC	0.06 *cm/f*
BundleMoCap	0.07 *cm/f*
BundleMoCap++	**0.03** *cm/f*

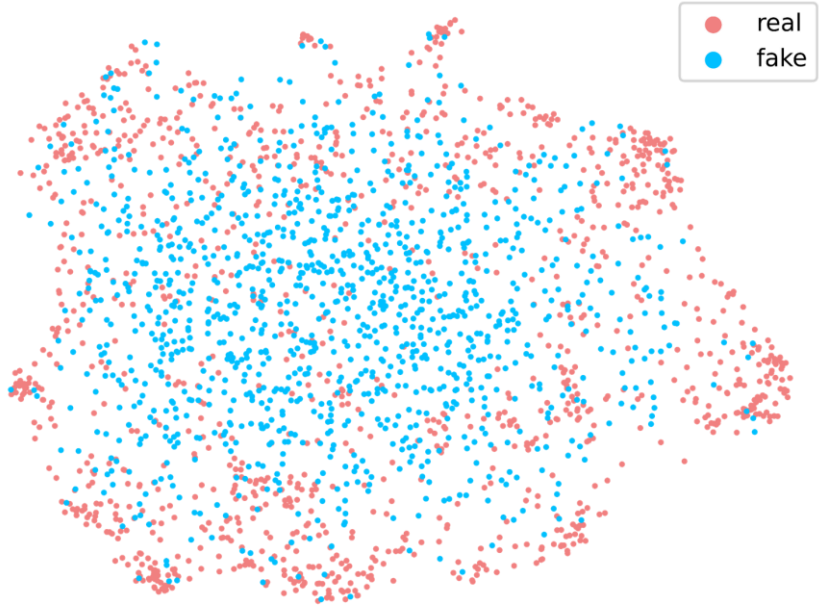
Performance vs Efficiency



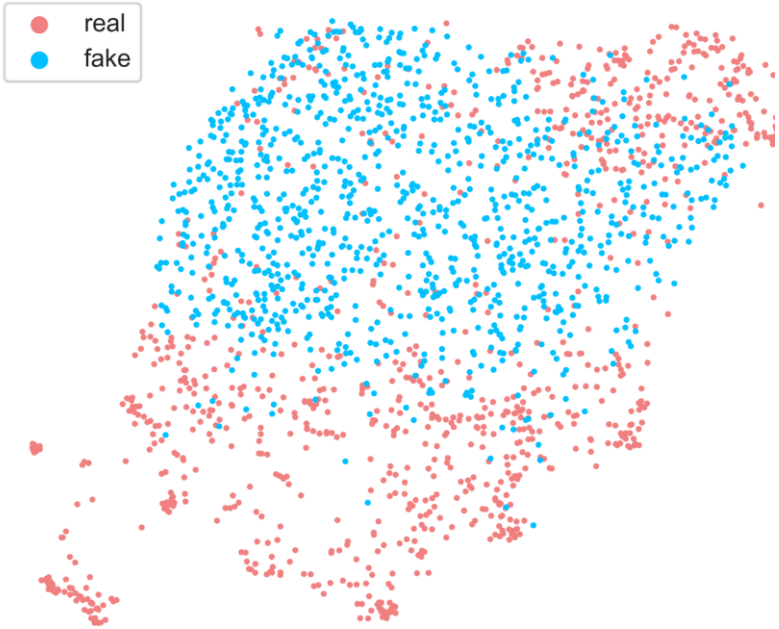
VAE Quality

	Synthesis		Reconstruction		
	FID ↓	DIV ↑	MPJPE ↓	PCK3 ↑	PCK7 ↑
VPoser	9.94	12.11	26.21mm	62.33%	89.82%
RVPoser	8.57	13.24	24.91mm	**70.01%**	**94.05%**
LieVAE	14.55	13.20	32.45 mm	55.32%	85.65%
SVAE	12.22	**19.51**	27.81 mm	66.14%	91.07%
SPoser(ours)	**3.45**	18.12	**24.88** mm	69.67%	93.99%

VAE Quality



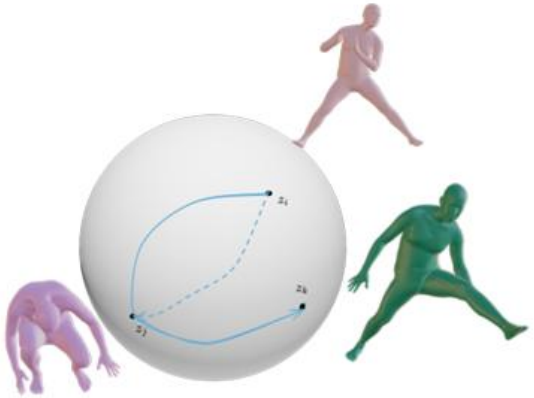
SPoser



VPoser

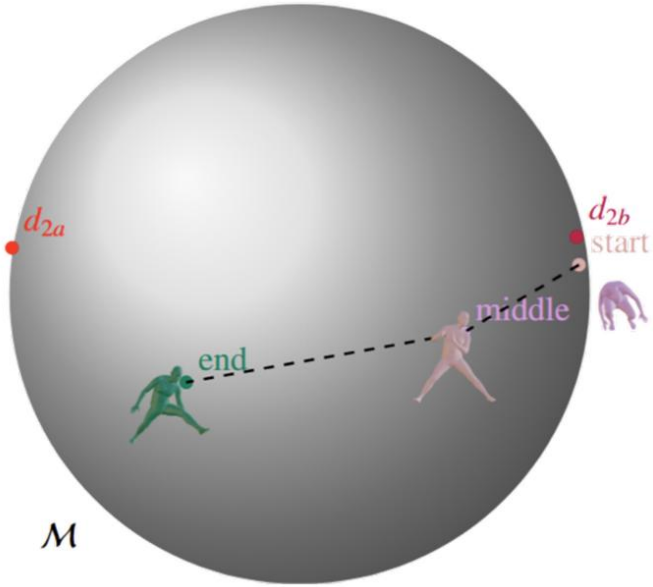
VAE Quality

$SLERP(z_i, z_j)$



$SIDER2(z_i, z_j, z_k)$

SIDER Interpolation



then

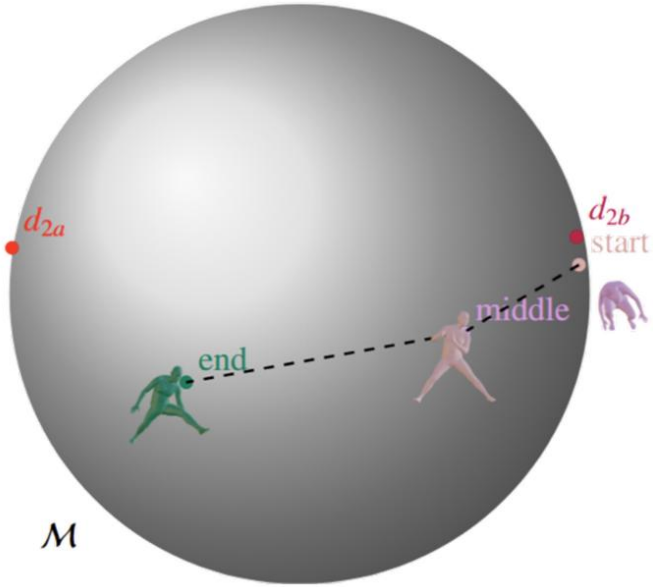
$$\mathbf{d}_{2a} = \mathrm{SLERP}(\mathbf{q}_3, \mathbf{q}_2, 2), \qquad \mathbf{d}_{2b} = \mathrm{SLERP}(\mathbf{q}_1, \mathbf{q}_2, 2),$$

$$\mathbf{c}_{inner} = \mathrm{SLERP}(\mathbf{q}_1, \mathbf{d}_{2a}, t), \qquad \mathbf{c}_{outer} = \mathrm{SLERP}(\mathbf{d}_{2b}, \mathbf{q}_3, t),$$

and finally

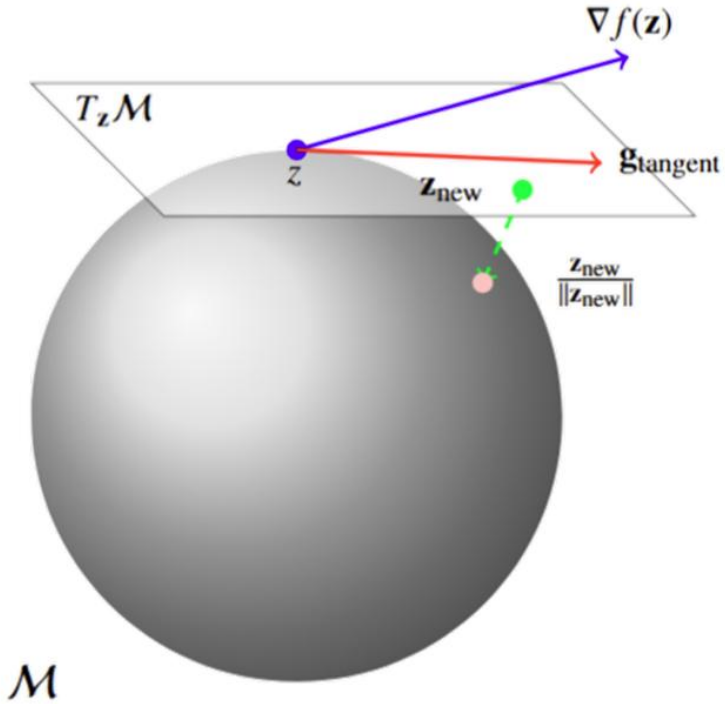
$$\mathrm{SIDER2}(\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, t) = \mathrm{SLERP}(\mathbf{c}_{inner}, \mathbf{c}_{outer}, t).$$

SIDER vs SLEPR

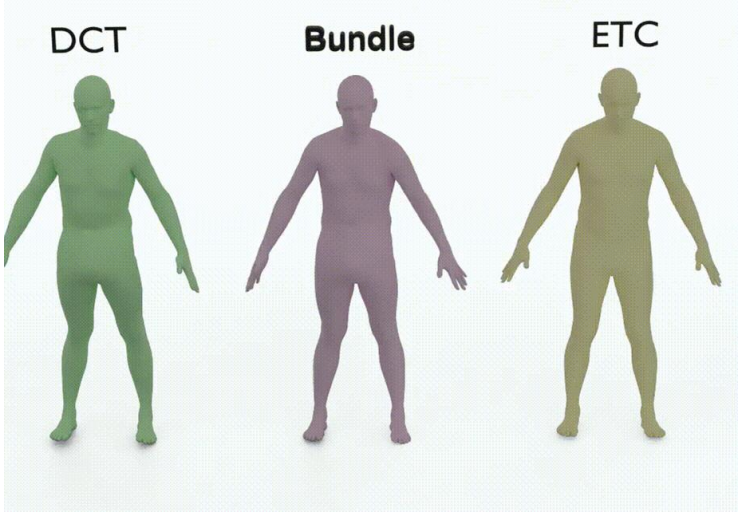


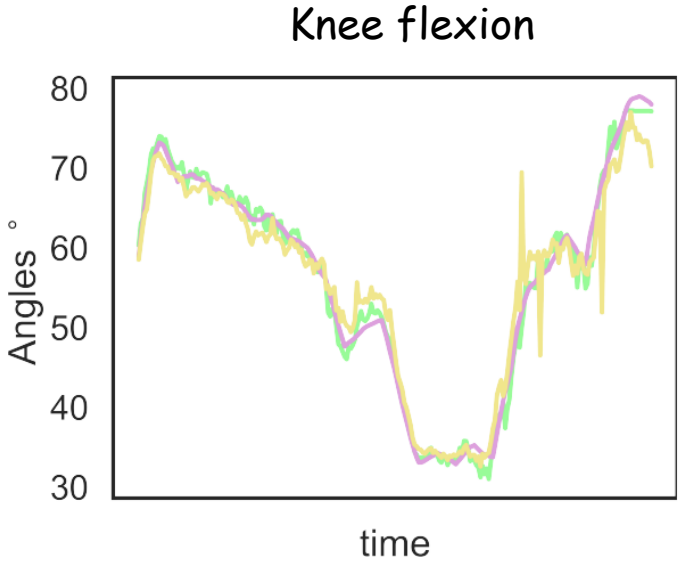
	MPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
SLERP	37.82 *mm*	42.62 *mm*	04.15°	35.86%	86.94%	3.66 mm/s^2
SIDER2	**36.32** *mm*	**40.93** *mm*	**04.09°**	**46.27%**	**93.97%**	**02.52** mm/s^2

Optimisation upon the manifold

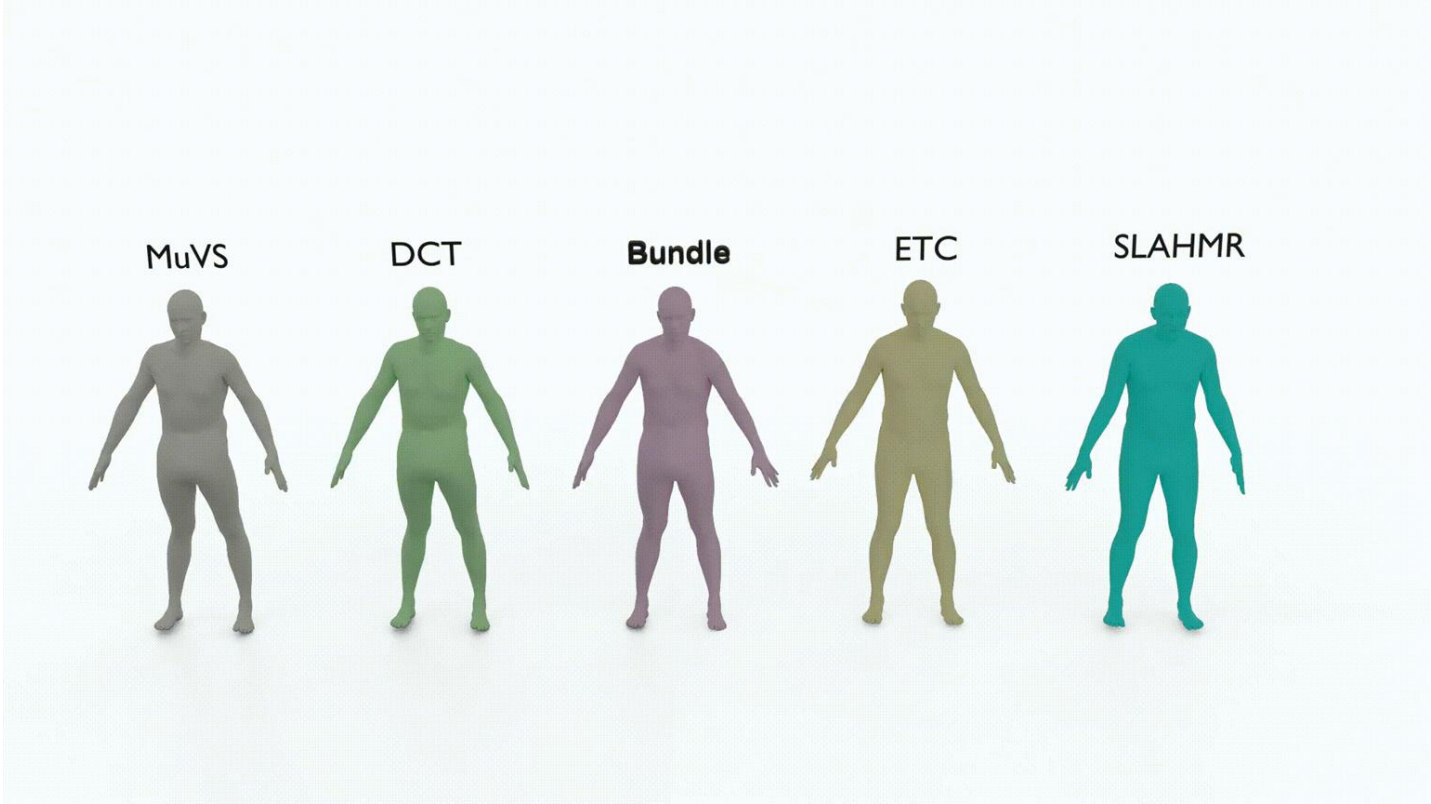


Smoothness





Comparisons



Comparisons vs Pose Estimators

	MPJPE ↓	RMSE ↓	PCK3 ↑	PCK7 ↑	accel ↓
MVN [14]	85.34 *mm*	114.03 *mm*	10.76%	62.48%	07.45 mm/s^2
BundleMoCap++	**48.27** *mm*	**56.44** *mm*	**48.75%**	**89.34%**	**05.34** mm/s^2

MPI-INF-3DHP

Comparisons vs Monocular

	MPJPE ↓	PAMPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
H4D		49.05 *mm*	56.70 *mm*	7.56°	40.34%	82.35%	18.10 mm/s^2
TCMR		110.4 *mm*	135.4 *mm*	15.56°	10.94%	55.46%	10.70 mm/s^2
BundleMoCap++	**48.27** *mm*		**56.44** *mm*	**4.33°**	**48.75%**	**89.34%**	**05.66** mm/s^2

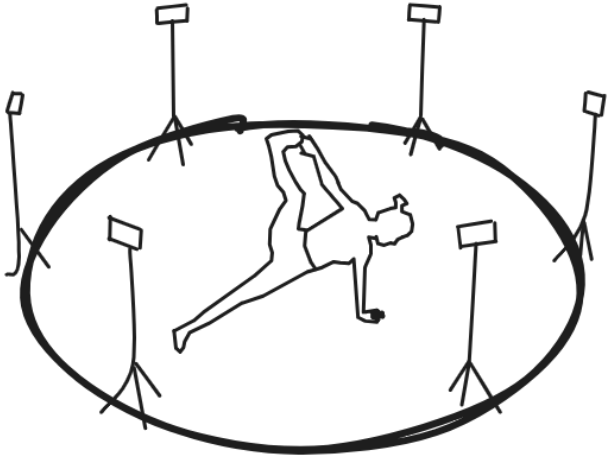
MPI-INF-3DHP

Comparisons vs Monocular

	MPJPE ↓	PAMPJPE ↓	RMSE ↓	MAE ↓	PCK3 ↑	PCK7 ↑	accel ↓
H4D		49.05 *mm*	56.70 *mm*	7.56°	40.34%	82.35%	18.10 mm/s^2
TCMR		110.4 *mm*	135.4 *mm*	15.56°	10.94%	55.46%	10.70 mm/s^2
BundleMoCap++	**48.27** *mm*		**56.44** *mm*	**4.33°**	**48.75%**	**89.34%**	**05.66** mm/s^2

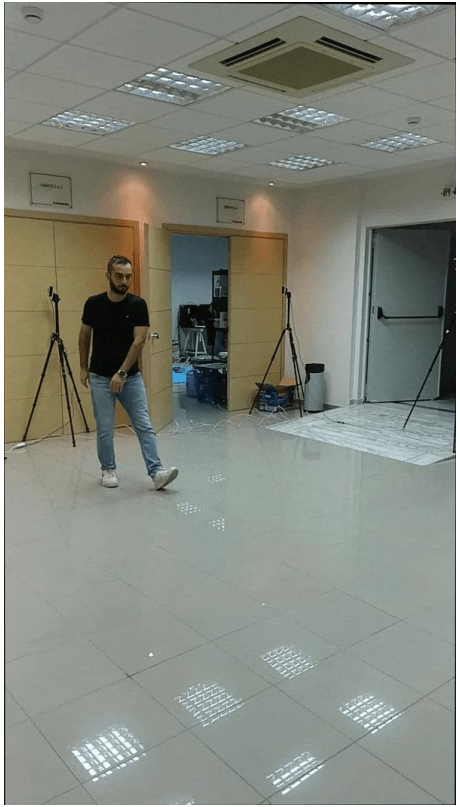
MPI-INF-3DHP

Low-cost setup results

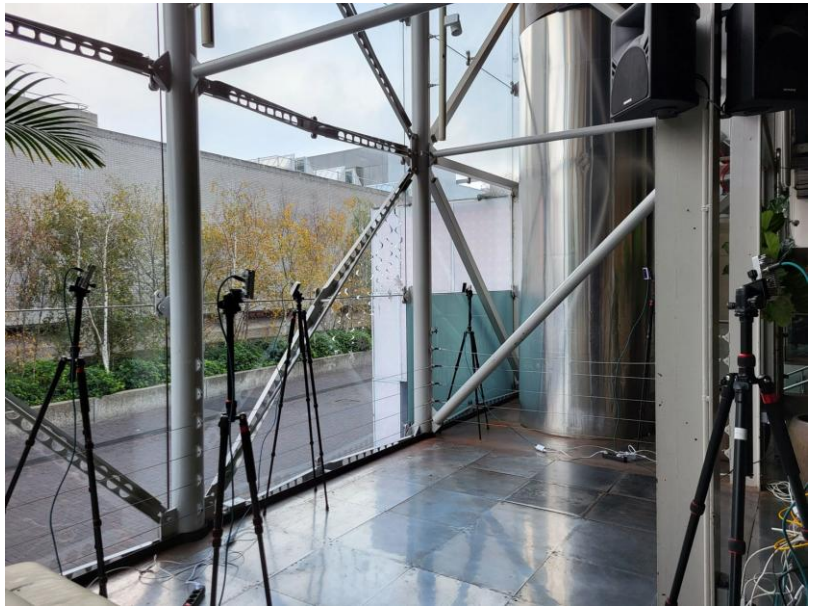


- 8x (calibrated & synchronised sensors)
- 60 fps
- 1920x1080

Low-cost setup results

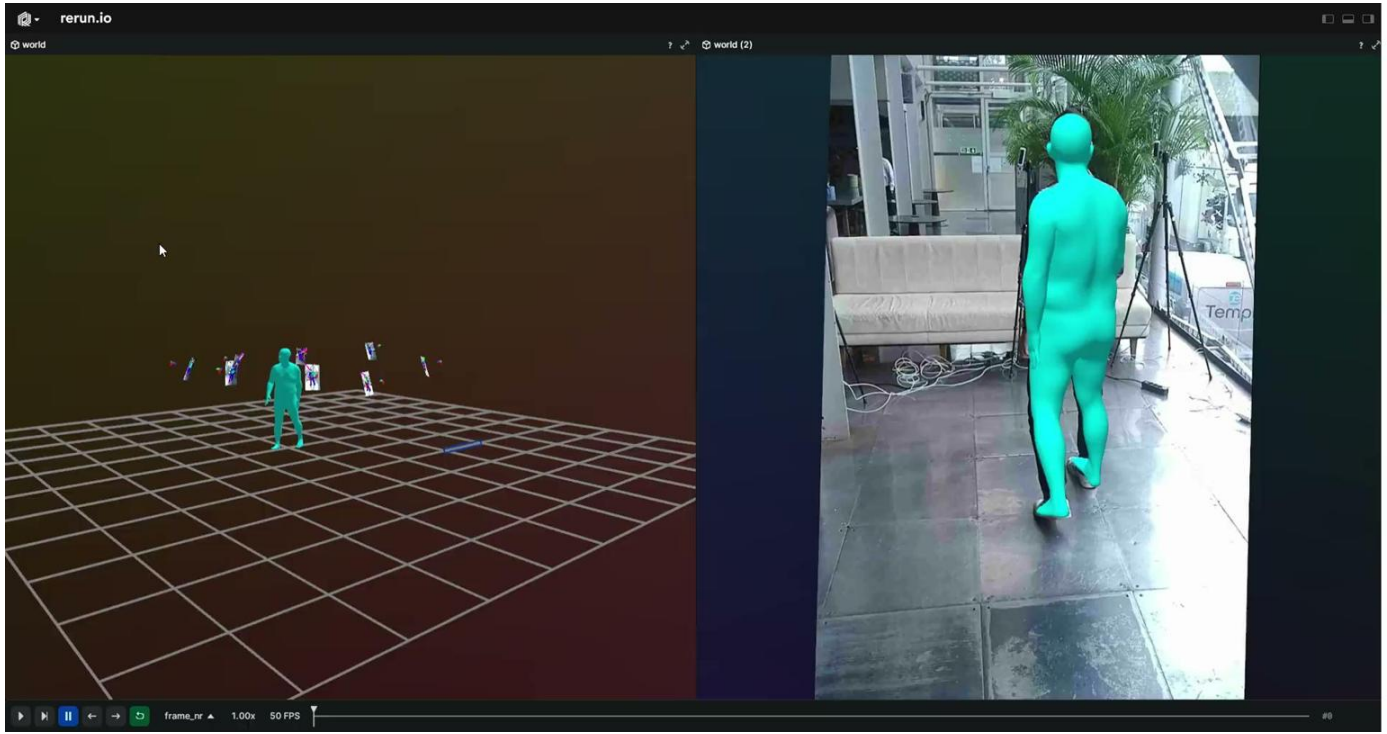


CVMP Demo



CVMP2023

CVMP Demo



Scientific Publications

Towards Scalable and Real-time Markerless Motion Capture.

Noise-in, Bias-out: Balanced and Real-time MoCap Solving

BundleMoCap: Efficient, Robust and Smooth Motion Capture from Sparse Multiview Videos.

BundleMoCap++: Efficient, Robust and Smooth Motion Capture from Sparse Multi view Videos.

From bias to balance: Leverage representation learning for bias-free MoCap solving

Robust and Efficient AI Motion Capture

Scientific Demos

MoCatalyst: Accelerating and Automating MoCap

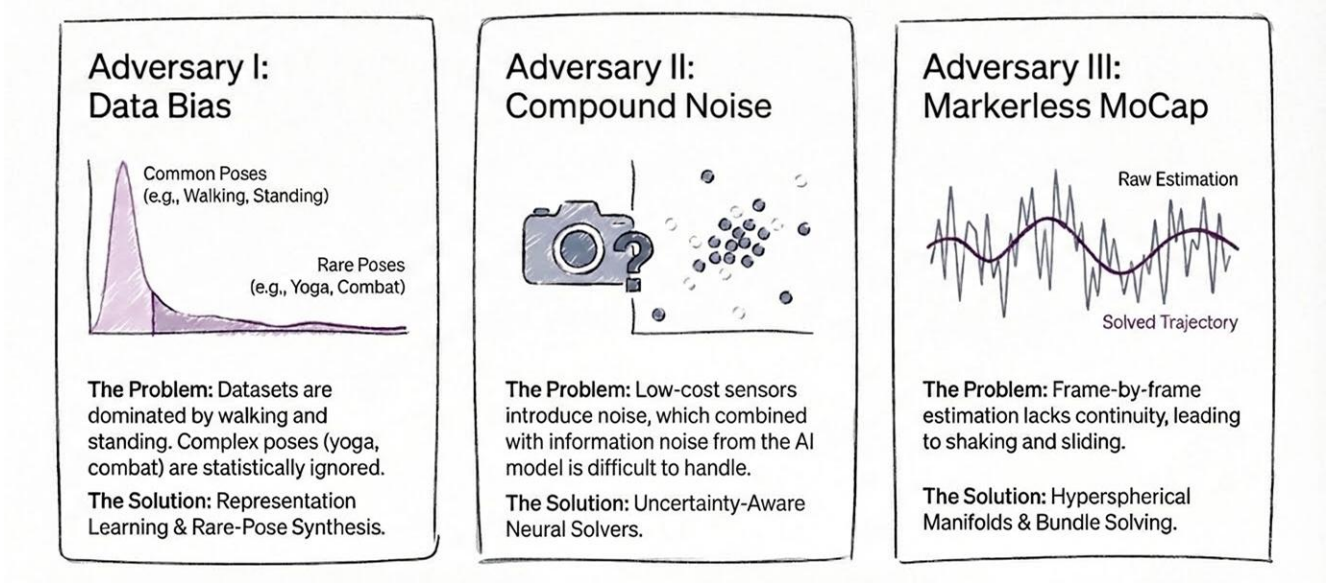
ICCV23 PARIS

LightMoCap: Light-weight, Real-time and Scalable Markerless Motion Capture

CVMP2023

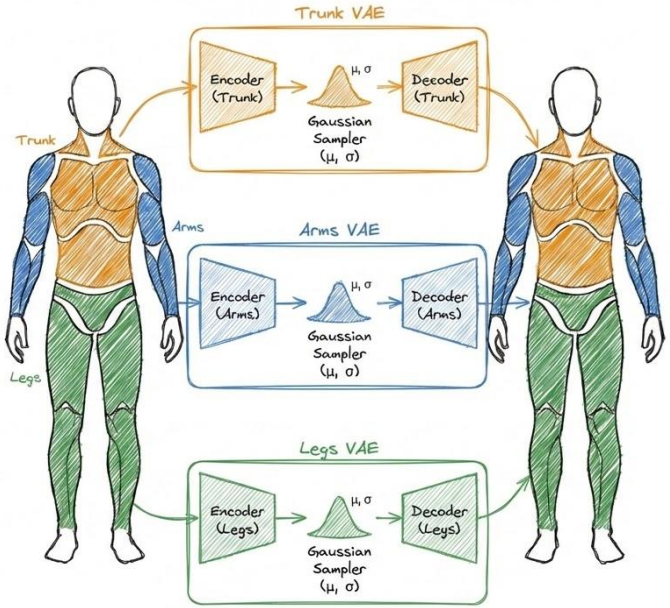
Summary

Representation learning to solve real-world problems around MoCap



Future Investigation

- Disentangled VAE → Better Representations (SO3)

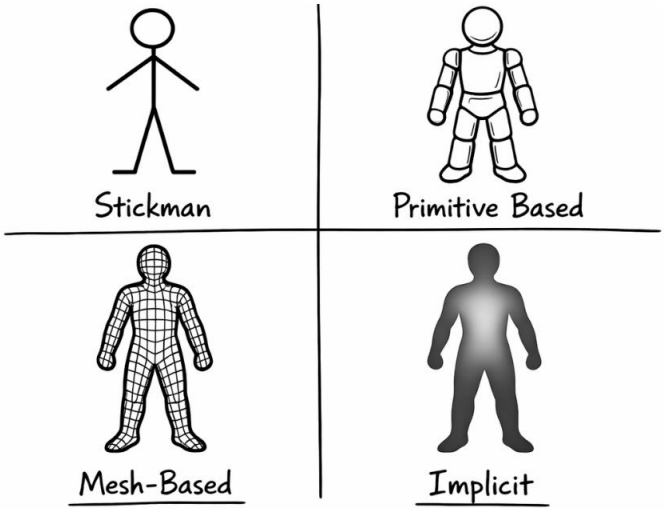


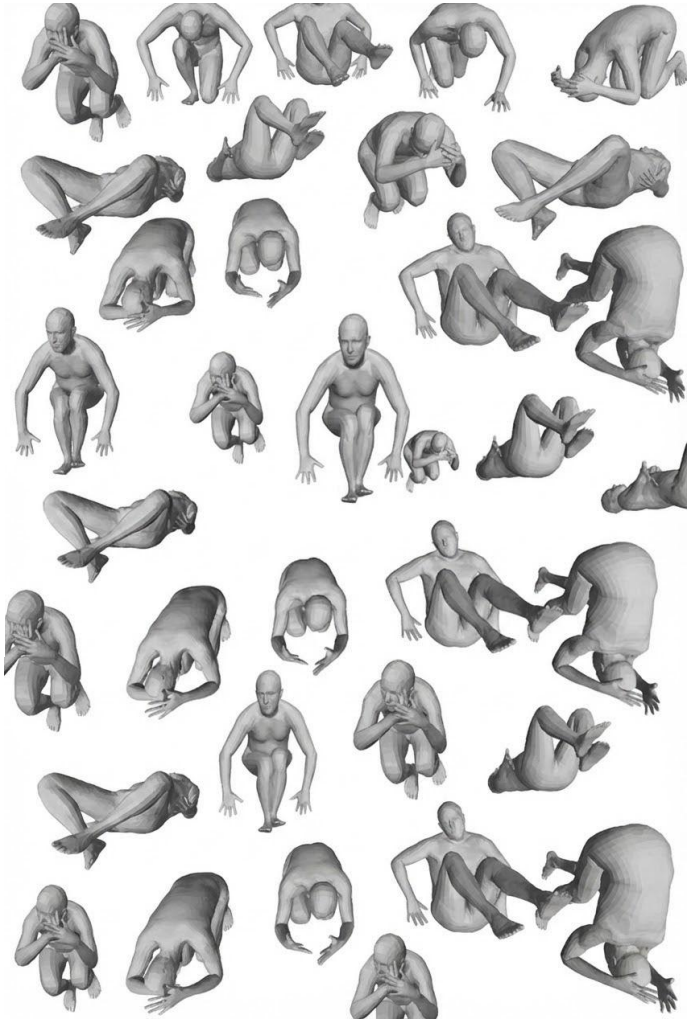
Future Investigation

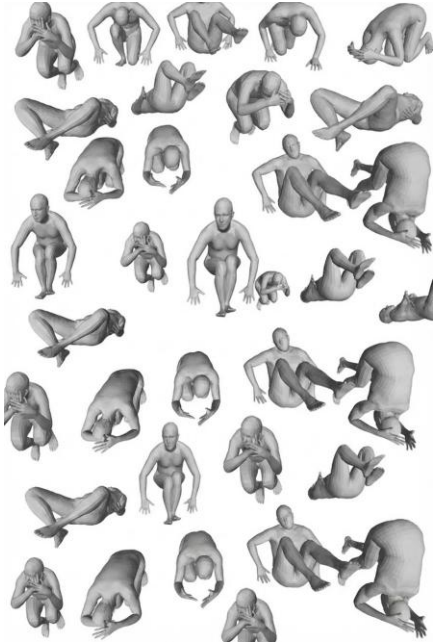
- Bundle Fingers

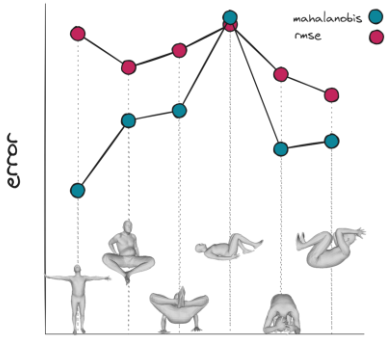
Future Investigation

- Towards implicit representations


Stickman
Primitive Based
Mesh-Based
Implicit






mahalanobis
rmse
error

Thank you

"Nothing happens until something moves" – Albert Einstein

Candidate: Georgios Albanis, Supervisor: Dr. Konstantinos Kolomvatsos