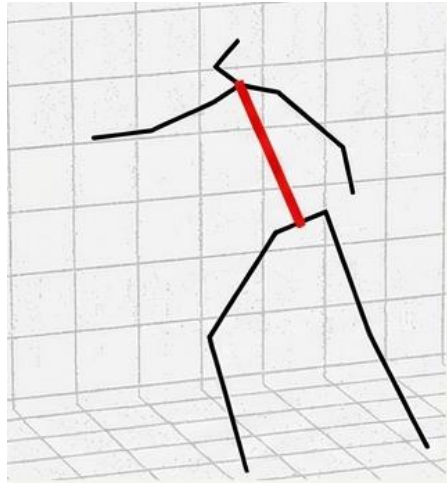


SIMSPINE: A BIOMECHANICS-AWARE SIMULATION FRAMEWORK FOR 3D SPINE MOTION ANNOTATION AND BENCHMARKING

Muhammad Saif Ullah Khan and Didier Stricker
German Research Center for Artificial Intelligence (DFKI)

— THE VISION GAP

Computer Vision Standard
Rigid and Reductive



Biomechanical Reality
Complex and Articulated



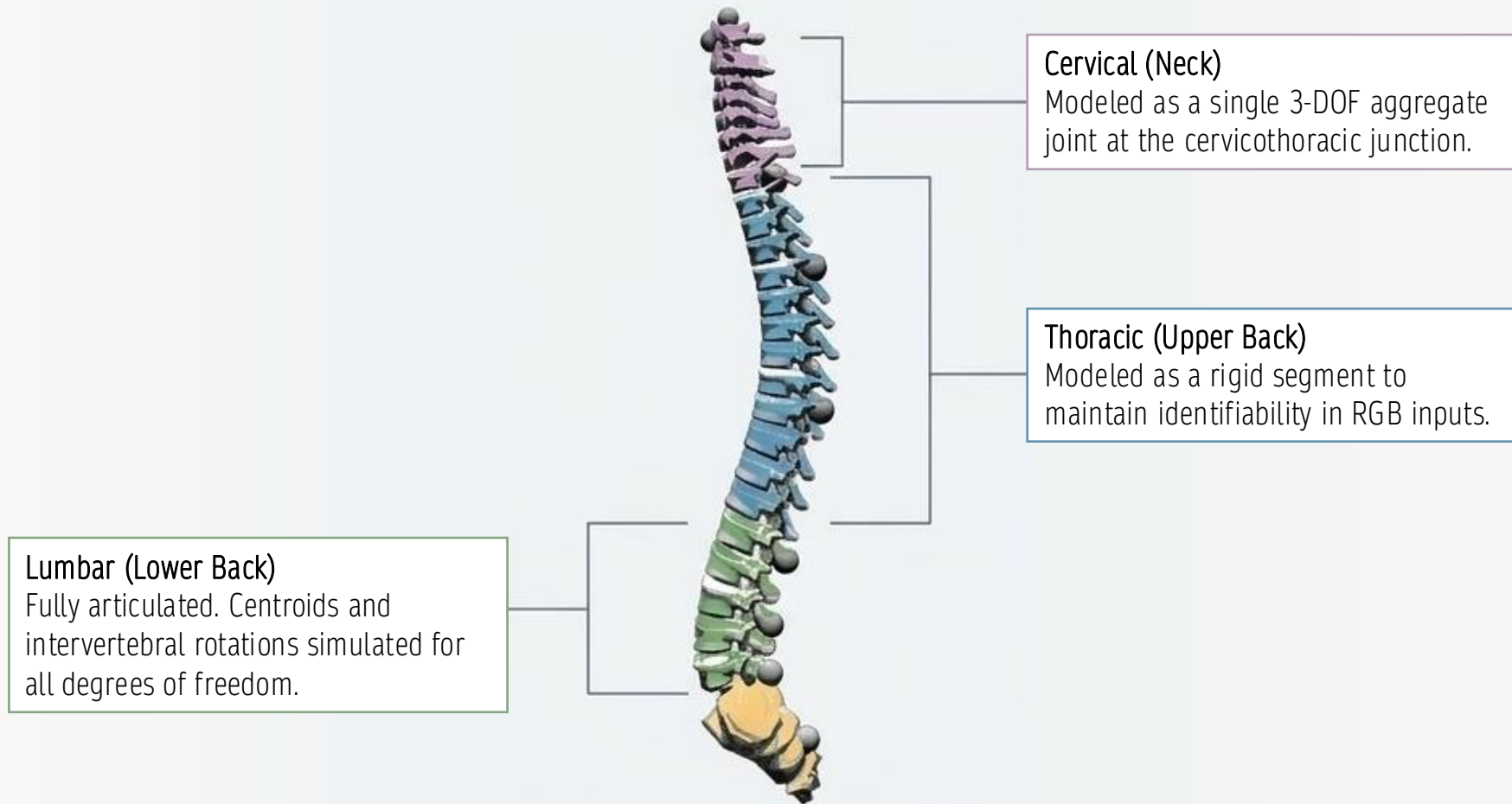
Without modeling the spine's multi-joint kinematics, clinical, ergonomics, and sports biomechanics applications remain fundamentally flawed.

SPINE SIMULATION

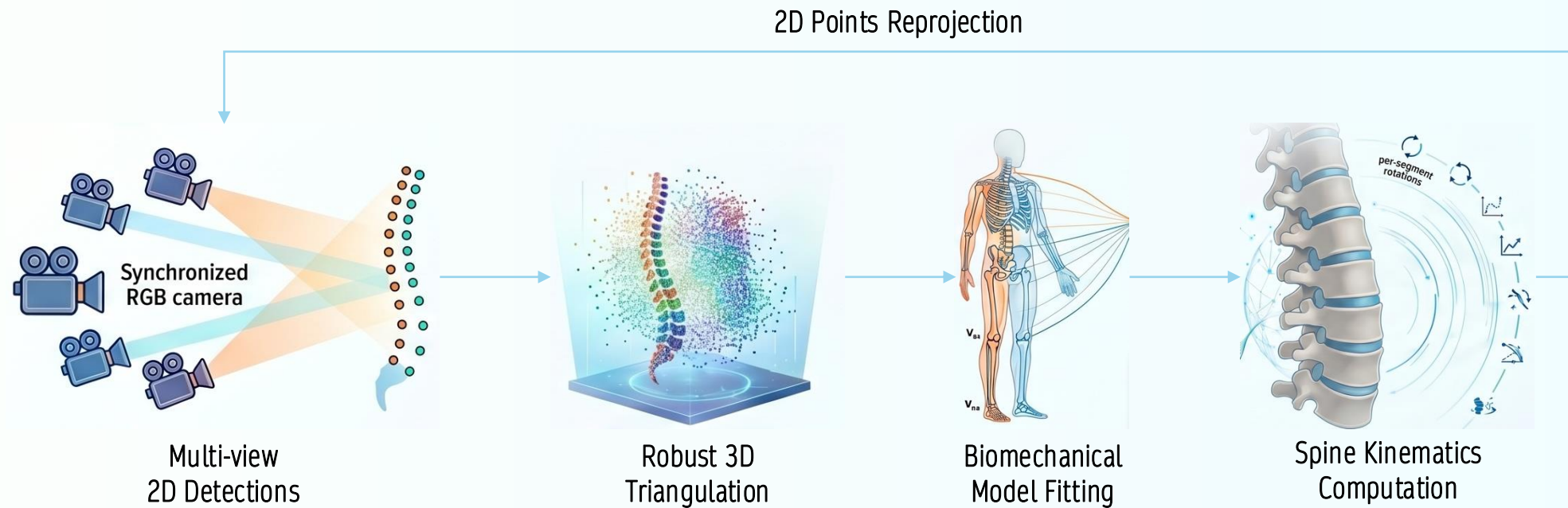
OpenSim-driven automated framework for simulating spine annotations on multi-view RGB video datasets



SPINE MODEL



BIOMECHANICS-AWARE SIMULATION



SIMSPINE DATASET

Provides sparse vertebra-level 3D annotations across 15 different natural full-body actions



SPINE TRACKING DATASETS

Dataset	Year	Input Type	Labels	# Samples	Annotation Type
VerSe	2021	CT	Vertebra centroids and segmentation masks	374 scans	Automated with manual review
Scoliosis1K	2024	Silhouette images	2D body pose, scoliosis diagnosis, demographics	447k frames	Predicted 2D poses, diagnosis by medical experts
SpineTrack	2025	RGB image	2D spine keypoints	33k images	Semi-automated human-in-the-loop annotations, non-expert
SIMSPINE (Ours)	2026	Multi-view RGB video	3D spine kinematics	2.14M frames	Real + biomechanical simulation, indoor scenes, automated labels

SIMSPINE is the only dataset providing biomechanically valid, vertebra-level 3D kinematics directly from standard RGB video sequences.

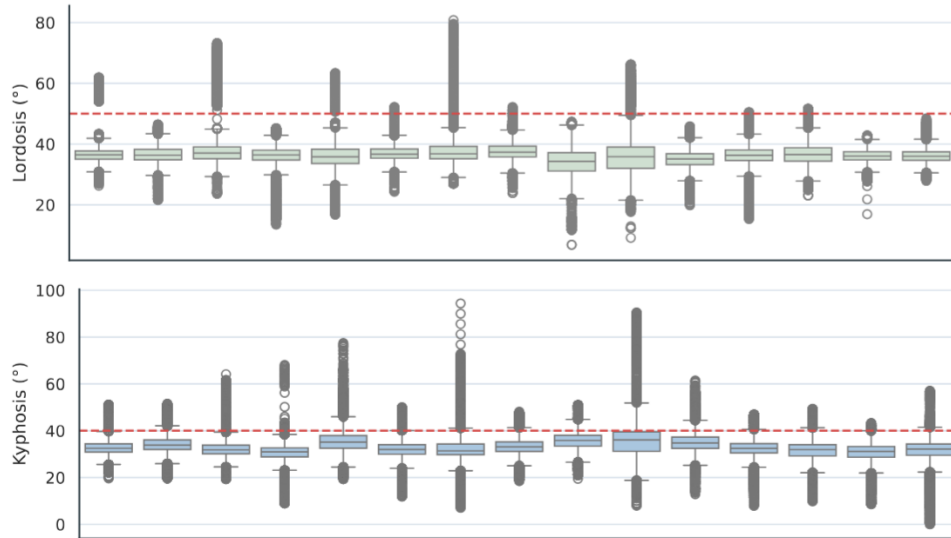
BIOMECHANICAL VALIDITY

Simulated motions show expected action-specific trends and align with clinical ranges for lumbar and thoracic curvatures



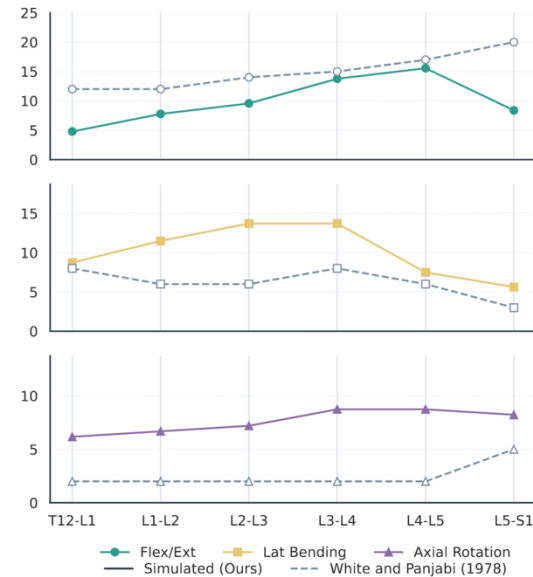
STATISTICS ANALYSIS

Lumbar / thoracic curvatures and lumbar spine articulation fall within reported ranges for nomrative adult populations, aligning SIMSPINE with established biomechanical literature.



Lumbar and Thoracic Curvatures
(SIMSPINE Mean Ranges)

Lumbar Lordosis: 31.7 - 41.3°
Thoracic Kyphosis: 28.1 - 38.6°



Lumbar spine agreement with White and Panjabi (1978)

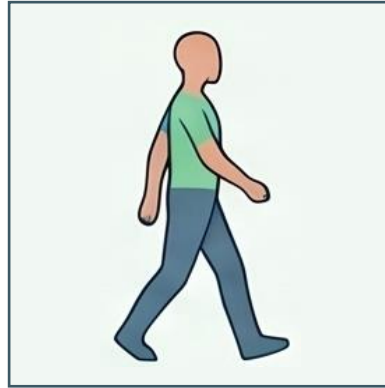
Follows gradients of classical biomechanical literature, respecting physical morphology

ACTION-SPECIFIC SPINAL CURVATURES



Seated Motions
(e.g., Sitting, SittingDown)

*Reduced lumbar lordosis
due to gravitational
effects*



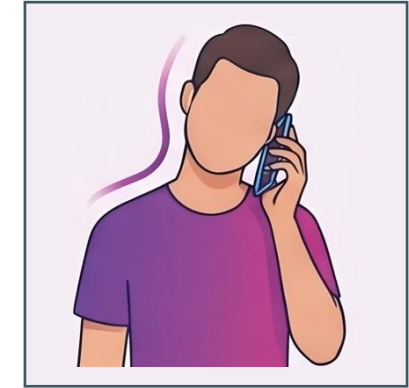
Standing Actions
(e.g., Walking, Dressing)

*Increased lumbar
lordosis and reduced
torso height*



Arm Elevation
(e.g., Photo, Posing)

*Reduced thoracic
curvature as shoulder
displacement flattens
upper back curve*



Neck Movement
(e.g., Phoning)

*Largest lateral bending
range reflecting typical
sideways head tilt during
phone conversations*

Across 15 different actions, SIMSPINE produces action-sensitive kinematics, proving synthetic supervision is a viable proxy for real-world spinal measurement

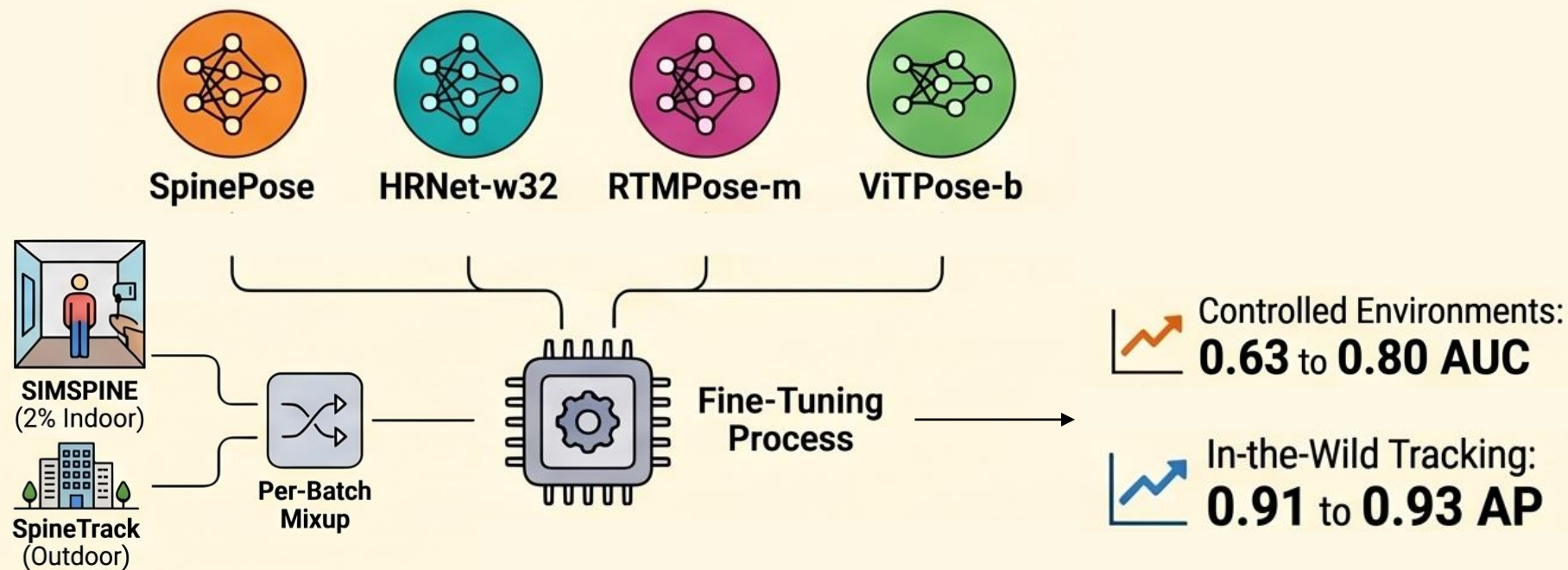
BENCHMARKING

Pretrained baselines for 2D and 3D spine pose estimation



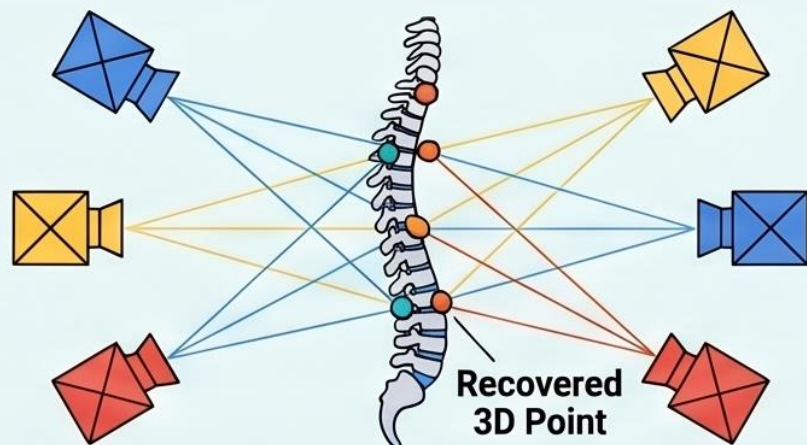
TASK 1: 2D POSE ESTIMATION

Evaluates off-the-shelf CNN and Transformer models fine-tuned on a balanced mix of SpineTrack (outdoor) and 2% SIMSPINE (indoors, ours).



TASK 2: MULTIVIEW 3D RECONSTRUCTION

Uses a weighted linear least-squares method with confidence-based outlier rejection to recover 3D points from synchronized multi-view 2D detections. MPJPE reported.



Action	S_C	S_T	S_L	S	B	All
Mean (GT 2D)	12.36	8.80	5.09	9.00	7.07	7.85
Directions	33.69	39.86	40.16	37.88	27.63	31.79
Discussion	36.35	40.55	41.44	39.39	28.33	32.81
Eating	38.47	43.04	35.40	39.48	27.44	32.32
Greeting	30.66	35.19	33.73	33.29	25.57	28.70
Phoning	42.41	44.64	39.14	42.43	27.23	33.40
Photo	36.40	42.13	42.27	40.25	30.35	34.37
Posing	33.48	37.46	40.14	36.85	28.02	31.60
Purchases	34.81	35.67	39.63	36.44	26.64	30.61
Sitting	40.94	43.16	31.21	39.23	25.42	31.02
SittingDown	34.85	35.95	23.63	32.30	23.31	26.95
Smoking	38.51	44.83	35.66	40.28	27.26	32.54
Waiting	35.17	37.40	33.22	35.54	24.78	29.14
WalkDog	35.53	41.44	40.83	39.31	29.49	33.47
WalkTogether	34.05	43.88	40.40	39.68	29.28	33.50
Walking	33.02	43.80	41.53	39.60	29.06	33.33
Mean (Finetuned)	36.50	41.07	37.13	38.50	27.27	31.82
Mean (Zero-Shot)	54.39	69.90	48.13	58.92	42.74	49.30

TASK 3: MONOCULAR 3D LIFTING

Monocular lifting approach using linear layers with residual connections, optimized with AdamW and a standard Mean Squared Error (MSE) loss function.

Spine-Only vs Full-Body Training



Spine-Only
(15 joints)



Full-Body
(37 joints)

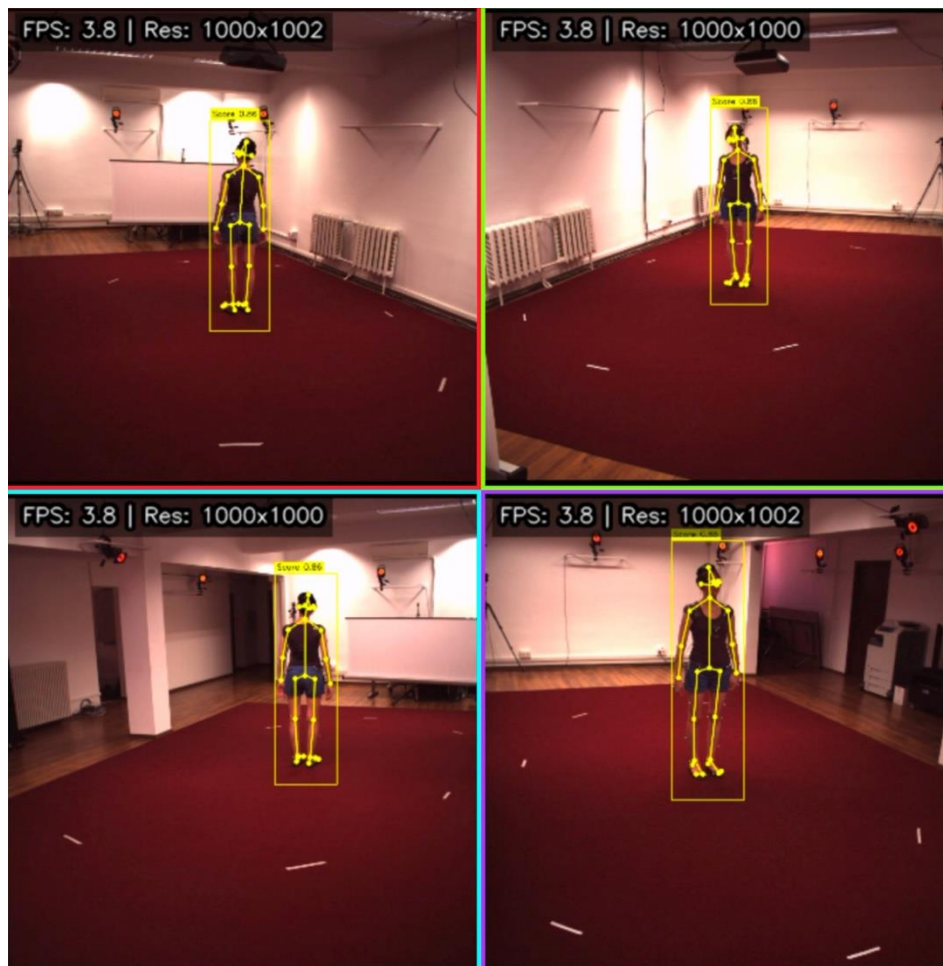
KEY FINDING

Global context improves spine localization

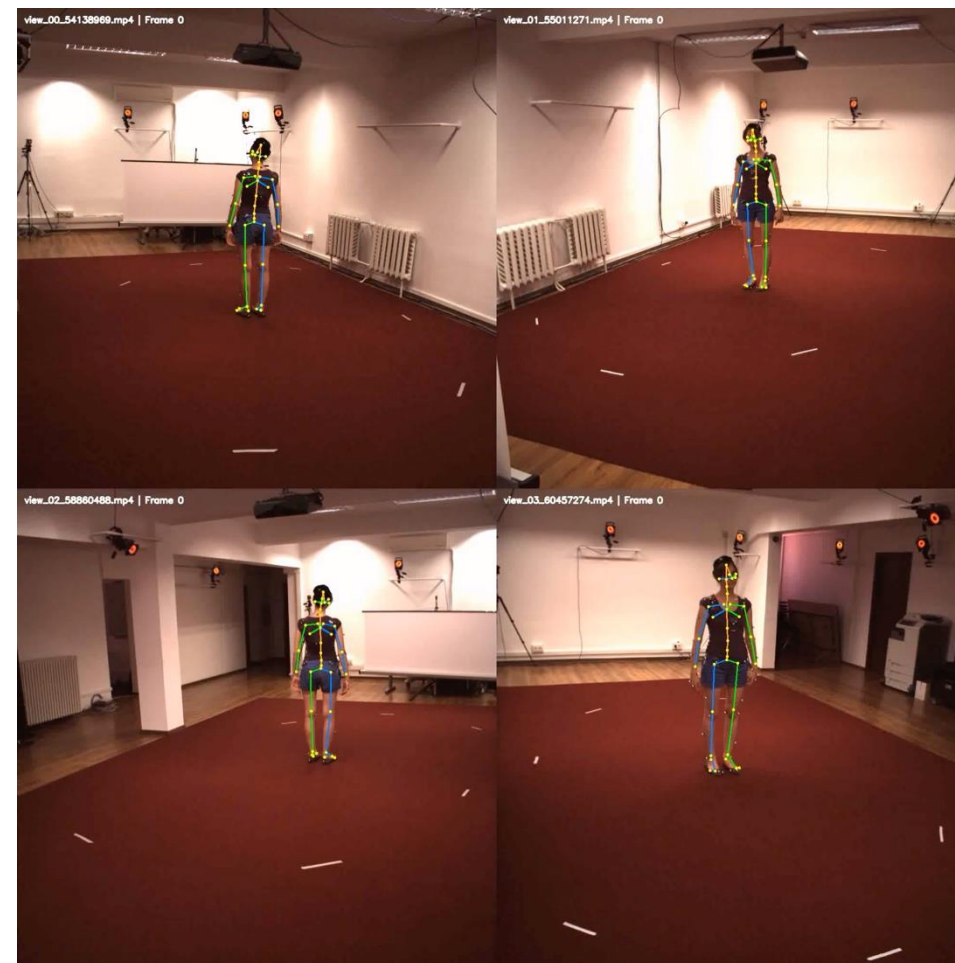
16.3 mm (Full-Body) vs **18.6 mm** (Spine-Only)

2D-to-3D lifting of spine keypoints is more accurate than limb keypoints in general because of relatively smaller and highly correlated vertebral movements with predictable S-curvature.

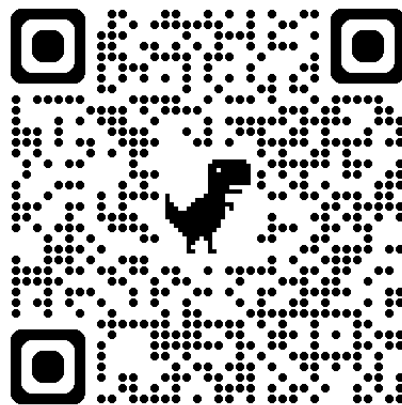
RESULTS



Previous Methods



SIMSPINE



THANK YOU



muhammad_saif_ullah.khan@dfki.de | saifkhichi.com/research/simspine