SimSpine: A Biomechanics-Aware Simulation Framework for 3D Spine Motion Annotation and Benchmarking

Supplementary Material

A. Data Augmentation Curriculum

Motivation. We employ a three-phase augmentation schedule that transitions from heavy appearance and occlusion randomization to geometry-focused fine-tuning. This curriculum combines domain randomization with pose-specific augmentations such as half-body cropping and occlusion masking. The aim is to first promote invariance, then stabilize spatial priors, and finally refine keypoints with clean, low-noise data.

Schedule. Given E fine-tuning epochs (E=10), augmentation stages switch at

$$e_1 = \lfloor 0.33E \rfloor, \quad e_2 = \lfloor 0.67E \rfloor,$$

producing intervals $[0, e_1)$, $[e_1, e_2)$, and $[e_2, E)$ for Stages 1–3.

A.1. Pipeline Overview

All stages share standard top-down cropping and affine normalization to a fixed input size (192×256), followed by target generation from augmented keypoints. Augmentation parameters are applied per sample with fixed random seeds for reproducibility.

Stage 1 (hard) Strong spatial and photometric perturbations for invariance:

- Random flip; half-body crop.
- Bounding-box transform: scale $[0.5, 1.5]$, rotation $\pm 180°$.
- Photometric distortions (exposure, contrast, saturation, hue).
- Occlusion via CoarseDropout (p=1.0, one hole, height/width $\in [0.2, 0.4]$).

Stage 2 (medium) Reduced occlusion and rotation for geometric stability:

- Random flip; half-body crop.
- Scale $[0.5, 1.5]$, rotation $\pm 90°$.
- CoarseDropout (p=0.5, same hole size).

Stage 3 (easy) Geometry-only refinement with minimal noise:

- Random flip only.
- Scale $[0.5, 1.5]$, rotation $\pm 60°$.
- No occlusion, blur, or photometric perturbation.
  Scale factors are relative to person crops.

A.2. Rationale

Stage 1: Large rotations and masking simulate extreme viewpoints and occlusions , driving the model to exploit global skeletal cues and lighting invariance. **Stage 2:** Milder noise allows consolidation of geometric structure without overfitting to easy samples. **Stage 3:** Clean inputs promote precise coordinate optimization and whole-body consistency, improving AUC while maintaining robustness to real-world variation.

Table A1. **Augmentation curriculum** Probabilities (p) are per-sample. CoarseDropout sizes are specified as fractions of the (pre-target) image after `TopdownAffine`.

Parameter	Stage 1	Stage 2	Stage 3
Epochs	$[0, e_1)$	$[e_1, e_2)$	$[e_2, E)$
Rotation	$\pm 180°$	$\pm 90°$	$\pm 60°$
Scale	$[0.5, 1.5]$	$[0.5, 1.5]$	$[0.5, 1.5]$
Shift	(default)	0.0	0.0
HalfBody	✓	✓	✗
Photometric / Occlusion	Photometric; Blur (p=0.1); MedianBlur (p=0.1); CoarseDropout (p=1.0, 1 hole, $h, w \in [0.2, 0.4]$)	Blur (p=0.1); MedianBlur (p=0.1); CoarseDropout (p=0.5, same sizes)	none

B. SimSpine Details

Table B2. Lordosis and Kyphosis stats by Action and Subject

Action	LLA (θ_l)	TKA (θ_k)
Directions	37.38 ± 5.51	33.27 ± 4.32
Discussion	36.42 ± 2.72	34.05 ± 3.63
Eating	38.46 ± 6.79	32.57 ± 4.50
Greeting	35.76 ± 4.25	30.45 ± 5.10
Phoning	37.36 ± 7.25	35.84 ± 5.33
Photo	36.99 ± 2.87	31.97 ± 3.84
Posing	39.09 ± 7.51	32.91 ± 6.76
Purchases	37.69 ± 3.11	33.35 ± 4.10
Sitting	34.43 ± 4.54	35.57 ± 3.94
SittingDown	35.96 ± 6.00	35.75 ± 7.68
Smoking	34.94 ± 3.21	35.14 ± 4.38
Waiting	36.21 ± 3.60	32.28 ± 4.13
WalkDog	36.70 ± 3.67	31.39 ± 5.13
WalkTogether	35.97 ± 2.18	30.70 ± 4.41
Walking	36.10 ± 2.38	31.60 ± 5.17
Avg	36.63 ± 4.37	33.12 ± 4.83

Subject	LLA (θ_l)	TKA (θ_k)
S1	36.24 ± 5.98	32.51 ± 4.25
S5	38.11 ± 2.86	31.88 ± 3.99
S6	37.48 ± 3.30	37.35 ± 4.63
S7	37.37 ± 4.80	32.63 ± 6.14
S8	35.61 ± 5.50	33.95 ± 5.44
S9	34.71 ± 6.02	32.70 ± 5.61
S11	35.08 ± 2.97	34.27 ± 3.52
Avg	36.37 ± 4.49	33.61 ± 4.80

Table B3. **SimSpine Anatomical Keypoints, Kinematic Axes, and Hierarchical Parent Links.** The dataset defines 37 anatomical keypoints covering full-body and spine-specific markers, grouped into Lumbar, Thoracic, Cervical, and Peripheral regions. Kinematic axes correspond to OpenSim-style rotational and translational degrees of freedom in the musculoskeletal model. Parent indices follow the hierarchical linkage used in forward kinematics. *Note: Due to licensing reasons, only 15 spine annotations are included in the released dataset. Body and feet annotations can be obtained from Human3.6M and H3WB datasets, respectively.*

ID	Name	Region	Parent ID	Swap / Side	Associated Kinematic Axes (if any)
0	Nose	Cervical	17	–	neck_flexion, neck_bending, neck_rotation
1	LEye	Body	0	REye	–
2	REye	Body	0	LEye	–
3	LEar	Body	1	REar	–
4	REar	Body	2	LEar	–
5	LShoulder	Body	33	RShoulder	arm_flex_l, arm_add_l, arm_rot_l
6	RShoulder	Body	34	LShoulder	arm_flex_r, arm_add_r, arm_rot_r
7	LElbow	Body	5	RElbow	elbow_flex_l, pro_sup_l
8	RElbow	Body	6	LElbow	elbow_flex_r, pro_sup_r
9	LWrist	Body	7	RWrist	wrist_flex_l, wrist_dev_l
10	RWrist	Body	8	LWrist	wrist_flex_r, wrist_dev_r
11	LHip	Body	19	RHip	hip_flexion_l, hip_adduction_l, hip_rotation_l
12	RHip	Body	19	LHip	hip_flexion_r, hip_adduction_r, hip_rotation_r
13	LKnee	Body	11	RKnee	knee_angle_l, knee_angle_l_beta
14	RKnee	Body	12	LKnee	knee_angle_r, knee_angle_r_beta
15	LAnkle	Body	13	RAnkle	ankle_angle_l, subtalar_angle_l, mtp_angle_l
16	RAnkle	Body	14	LAnkle	ankle_angle_r, subtalar_angle_r, mtp_angle_r
17	Head	Cervical	36	–	–
18	Neck	Cervical	30	–	neck_flexion, neck_bending, neck_rotation
19	Hip (Root)	Lumbar	-1	–	pelvis_tx, pelvis_ty, pelvis_tz, pelvis_tilt, pelvis_list, pelvis_rotation
20	LBigToe	Body	15	RBigToe	–
21	RBigToe	Body	16	LBigToe	–
22	LSmallToe	Body	20	RSmallToe	–
23	RSmallToe	Body	21	LSmallToe	–
24	LHeel	Body	15	RHeel	–
25	RHeel	Body	16	LHeel	–
26	Spine_01	Lumbar	19	–	L5_S1_Flex_Ext, L5_S1_Lat_Bending, L5_S1_axial_rotation, L4_L5_Flex_Ext, L4_L5_Lat_Bending, L4_L5_axial_rotation
27	Spine_02	Lumbar	26	–	L3_L4_Flex_Ext, L3_L4_Lat_Bending, L3_L4_axial_rotation, L2_L3_Flex_Ext, L2_L3_Lat_Bending, L2_L3_axial_rotation
28	Spine_03	Lumbar	27	–	L1_L2_Flex_Ext, L1_L2_Lat_Bending, L1_L2_axial_rotation
29	Spine_04	Thoracic	28	–	–
30	Spine_05	Thoracic	29	–	–
31	LLatissimus	Thoracic	29	RLatissimus	–
32	RLatissimus	Thoracic	29	LLatissimus	–
33	LClavicle	Thoracic	30	RClavicle	–
34	RClavicle	Thoracic	30	LClavicle	–
35	Neck_02	Cervical	18	–	L1_T12_Flex_Ext, L1_T12_Lat_Bending, L1_T12_axial_rotation
36	Neck_03	Cervical	35	–	–

Spine markers by region:
Lumbar (4): Hip, Spine_01, Spine_02, Spine_03.
Thoracic (6): Spine_04, Spine_05, 2×Latissimus, 2×Clavicles.
Cervical (5): Nose, Head, Neck, Neck_02, Neck_03.
Total keypoints: 37 (**Spine:** 15 **Body:** 22).

C. Additional Results

Figure C1 shows additional qualitative comparisons with the previous SOTA 2D tracker for spine pose estimation [31].

Left: On an internet frontal-view image with lateral bending, [31]'s tracker produces noisy, implausible locations for lower spine, upper neck, and clavicle points (circled in red), which our tracker corrects. *Middle:* On H36M validation samples using 2D tracking + multiview triangulation, [31] underestimates sagittal curves, and places clavicles behind the spine, which is anatomically impossible. *Right:* On two self-recorded front-facing bicycle views (novel motion), we compare sagittal curvature in 3D from [31] with our 2D-based reconstruction: [31] exhibits discontinuities, while our model yields a more realistic back posture, showing that our method and data go beyond simple 2D pose [31].

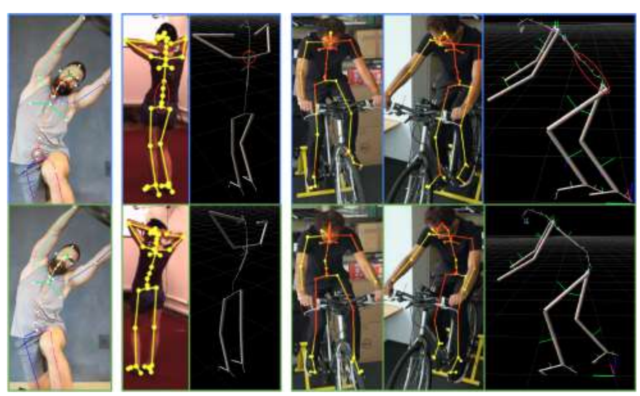
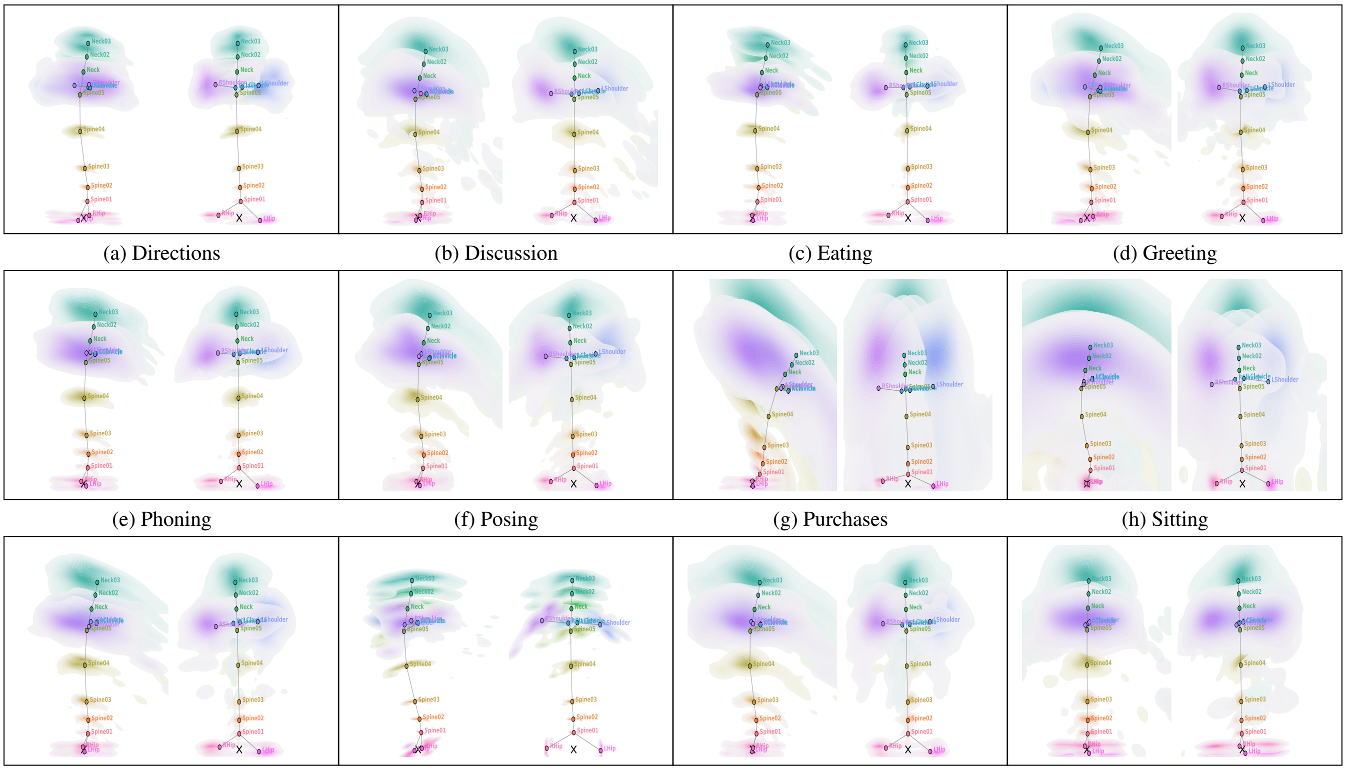


Figure C1. **Qualitative Evidence:** Comparison of the 2D tracker from [31] (top, blue) and our fine-tuned tracker on SimSpine (bottom, green). *Best viewed zoomed-in on a screen.*



(a) Directions (b) Discussion (c) Eating (d) Greeting

(e) Phoning (f) Posing (g) Purchases (h) Sitting

(i) Smoking (j) TakingPhoto (k) Waiting (l) Walking

Figure C2. **Sagittal (Y–Z) and Frontal (X–Z) Distributions of Upper-Body Joint Positions Across Actions.** Each pair of panels shows the sagittal (left) and frontal (right) kernel-density contours of root-relative joint coordinates aggregated over all motion-capture frames of that action. Colored circles denote mean joint centers, and connecting lines depict the average kinematic chain. Sagittal views highlight action-specific curvature and torso inclination—e.g., pronounced flexion in *Sitting* and near-vertical alignment in *Walking*—whereas frontal views emphasize lateral symmetry and limited sideward spread, reflecting the bilateral consistency of the upper torso.

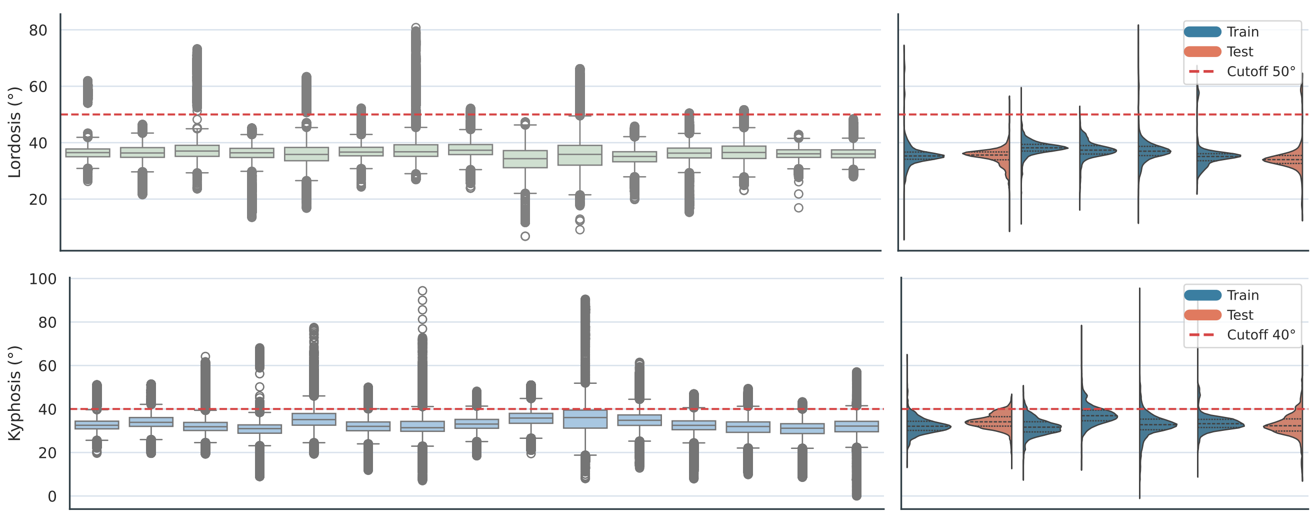


Figure C3. **Spine curvature boxplots to supplement the curvature summary** provided in Fig. 3 (left). We provide per-subejct and per-action distributions and a reference cutoff value for "normal" lordosis and kyphosis in literature.