

Synthetic CCTV Fight and Interpersonal Violence Dataset

Simulated fights and close interactions between two to six people, observed by 12 CCTV cameras each, with multi-person keypoints, roles, interaction phases and contact events. Built to train and test skeleton-based fight detectors and to measure false alarms on friendly contact.

**1,045**

simulated events

12,540sequences (12 cameras
each)**26.3 h**

2.84 M frames at 30 fps

20classes: 10 fights, 10 hard
negatives

● Fight classes (579 events)

punch_head	153	tackle	39
punch_body	61	takedown	58
kick_body	48	shove	60
kick_leg	33	grab_strike	35
group_fight	58	grapple	34

● Hard negatives (466 events)

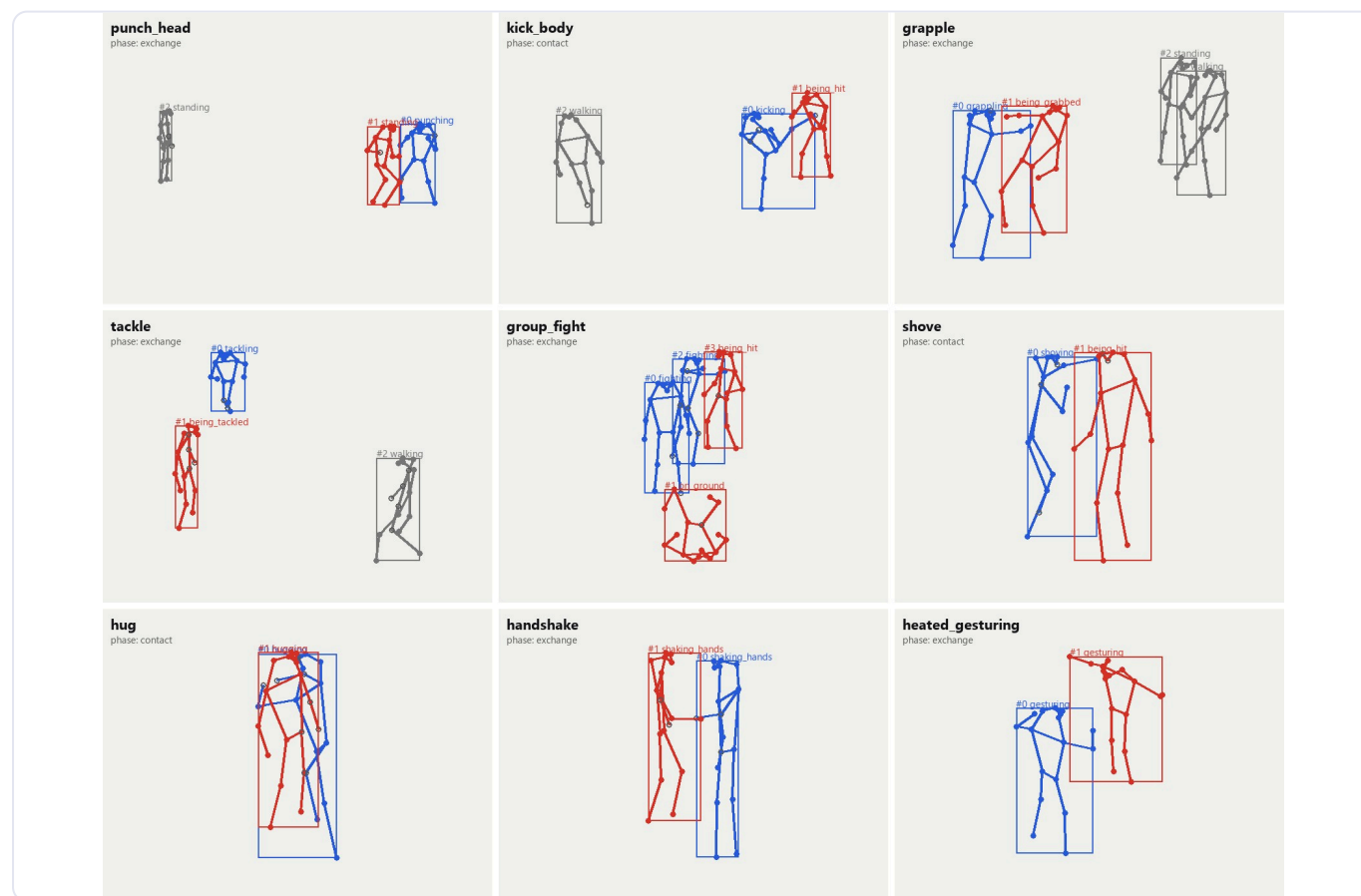
hug	48	play_push	47
handshake	48	heated_gesturing	57
high_five	45	walk_past	50
back_pat	39	run_together	41
dance_pair	44	crowd_normal	47

● Annotations

- **Per person and frame:** track id, role (initiator of contact, receiver, bystander), action, 17 COCO keypoints with visibility (visible / occluded / outside the frame), bounding box, share occluded by other people.
- **Per frame:** interaction phase: approach, contact, exchange, separation, aftermath.
- **Per event:** fight start and end, first contact, struck body part, striking limb, knock-down impulse; 3D skeleton (23 joints) of every person; camera intrinsics and extrinsics.

Xentim is a brand of TATO STUDIO sp. z o.o., ul. 28 Czerwca 1956 r. 174/28, 61-507 Poznan, Poland. Generated with Unreal® Engine (Epic Games, Inc.) and NVIDIA Kimodo (NVIDIA Open Model License; text encoder built with Meta Llama 3). Not affiliated with, sponsored or endorsed by Epic Games, NVIDIA or Meta.

● Labels at the moment of contact



● Scenes and cameras

Scenes	open hall 420, railway platform 315, metro platform 310 events
People	2 to 6 per event; 470 events with bystanders
Cameras	3 fixed + 9 randomized CCTV per event; 1.5 to 8 m high, 1 to 11 m away, 12 to 28 mm lens, Super 35 sensor
Split	train 8,664 / val 2,040 / test 1,836 sequences; about 15 % of each class in val and test
Formats	Parquet, CSV, JSON (ZIP)

● How it was made

- Strikes, grabs and reactions: NVIDIA Kimodo text-to-motion, paired and timed in Unreal Engine 5.8 so each strike lands on the body.
- Knock-downs, tackles and takedowns end in a physics ragdoll.
- Automatic quality control on every person: teleports, joint pops, foot sliding, floor penetration, bone stretch, strikes that miss, bodies that overlap. 239 of 1,284 runs removed.
- No real people, no personal data, no internet footage.

● Licence and pricing

Research: one academic institution, non-commercial	€3,000
Standard: one company, one product, all sites, updates for 12 months	€15,000
Enterprise: several products and affiliates, priority support	€38,000
Bundle of three Xentim datasets (falls, railway, fights)	25 % off

Free evaluation sample: 30 simulations x 12 cameras with preview videos. Keypoint annotations only, no RGB video. Licensed for discriminative models; use for generative AI is prohibited.