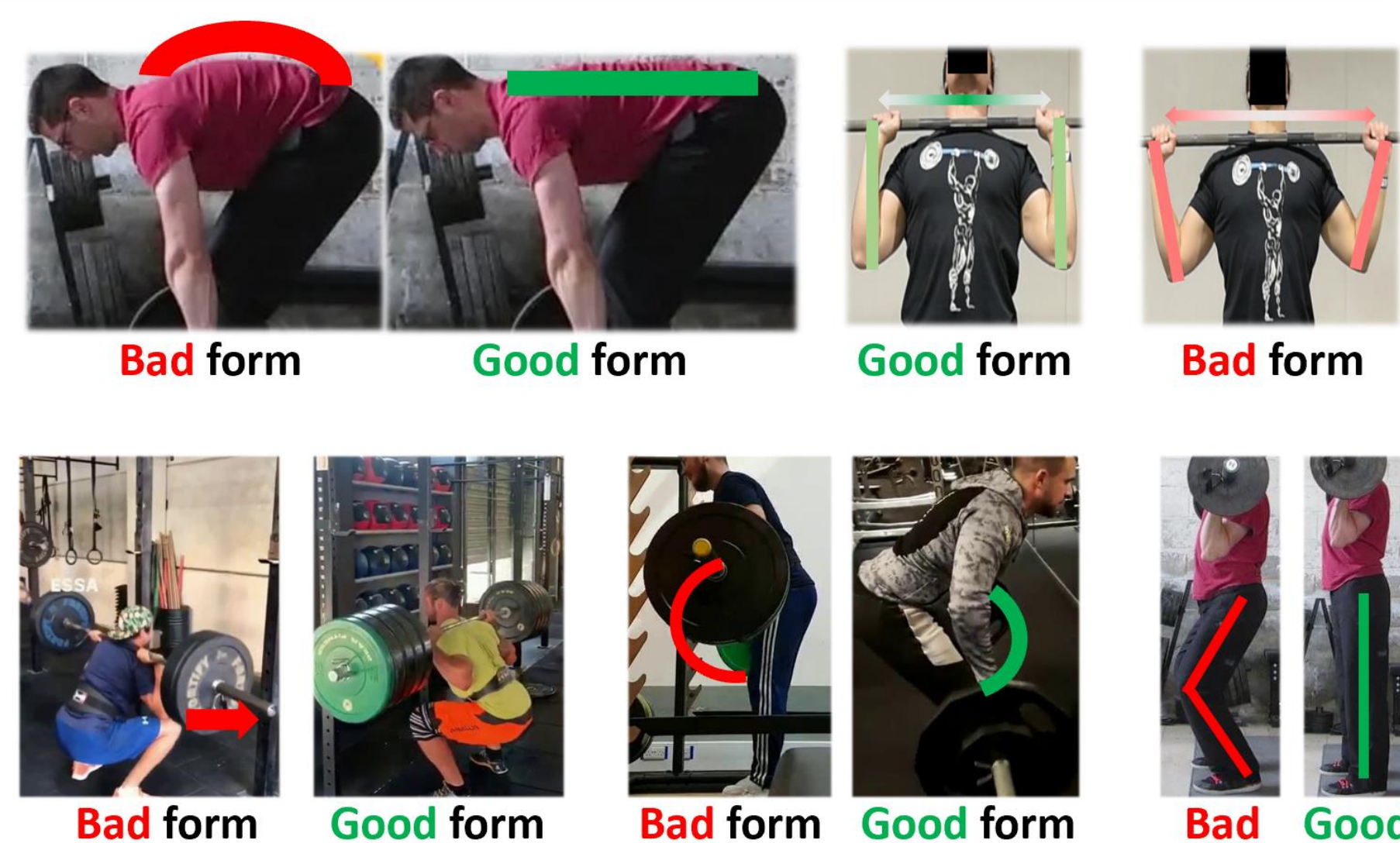


Task Definition

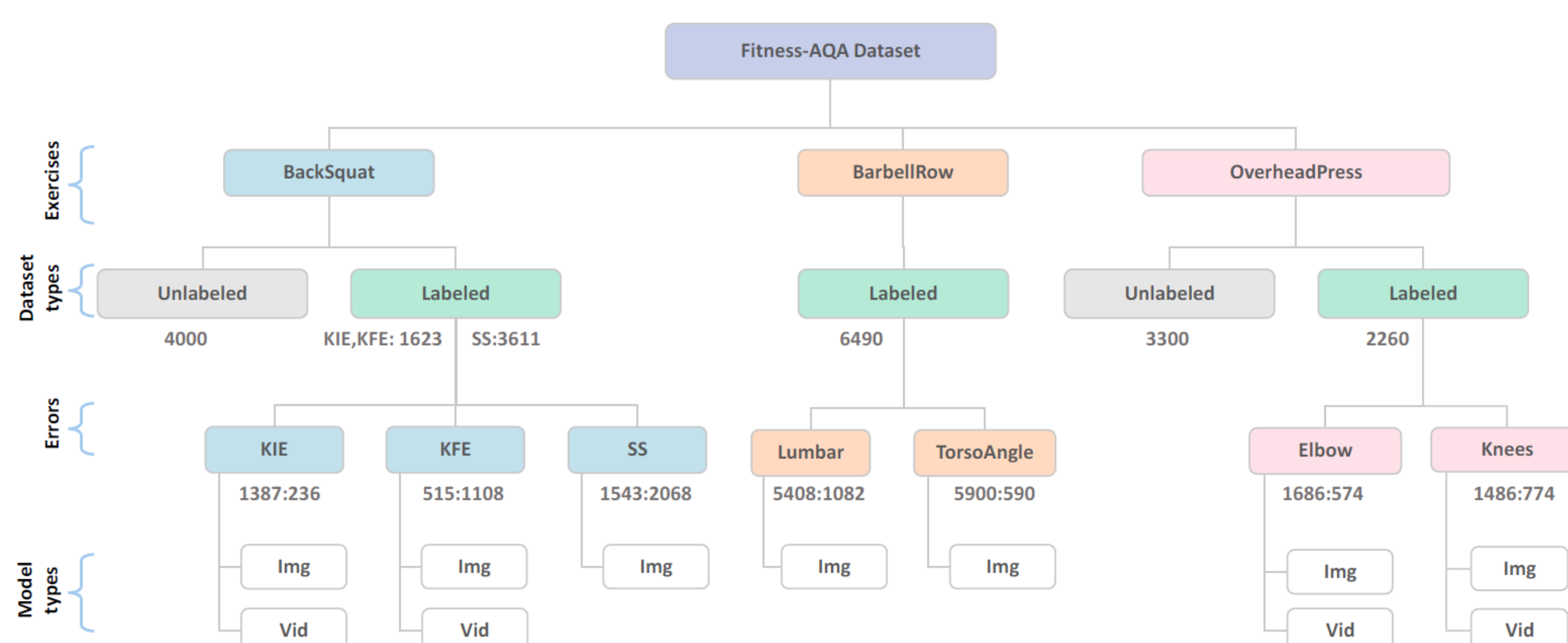
Detecting errors (**Bad** forms) in Workout Form in Real-World Scenarios



Shortcomings of Current Work

- Academic research is limited to controlled conditions
- Use off-the-shelf 2D/3D Pose Estimators
 - Good for simple, controlled conditions
 - But does not fare well in real-world, in-the-wild conditions
- Dataset void: No suitable in-the-wild datasets available

Contribution 1: Our Fitness-AQA Dataset



- Real-world videos
- People making errors under the impact of actual weights
- Occlusions
- Various types of clothing, background, illumination
- Unusual poses
- Severe to Subtle, Finegrained action errors

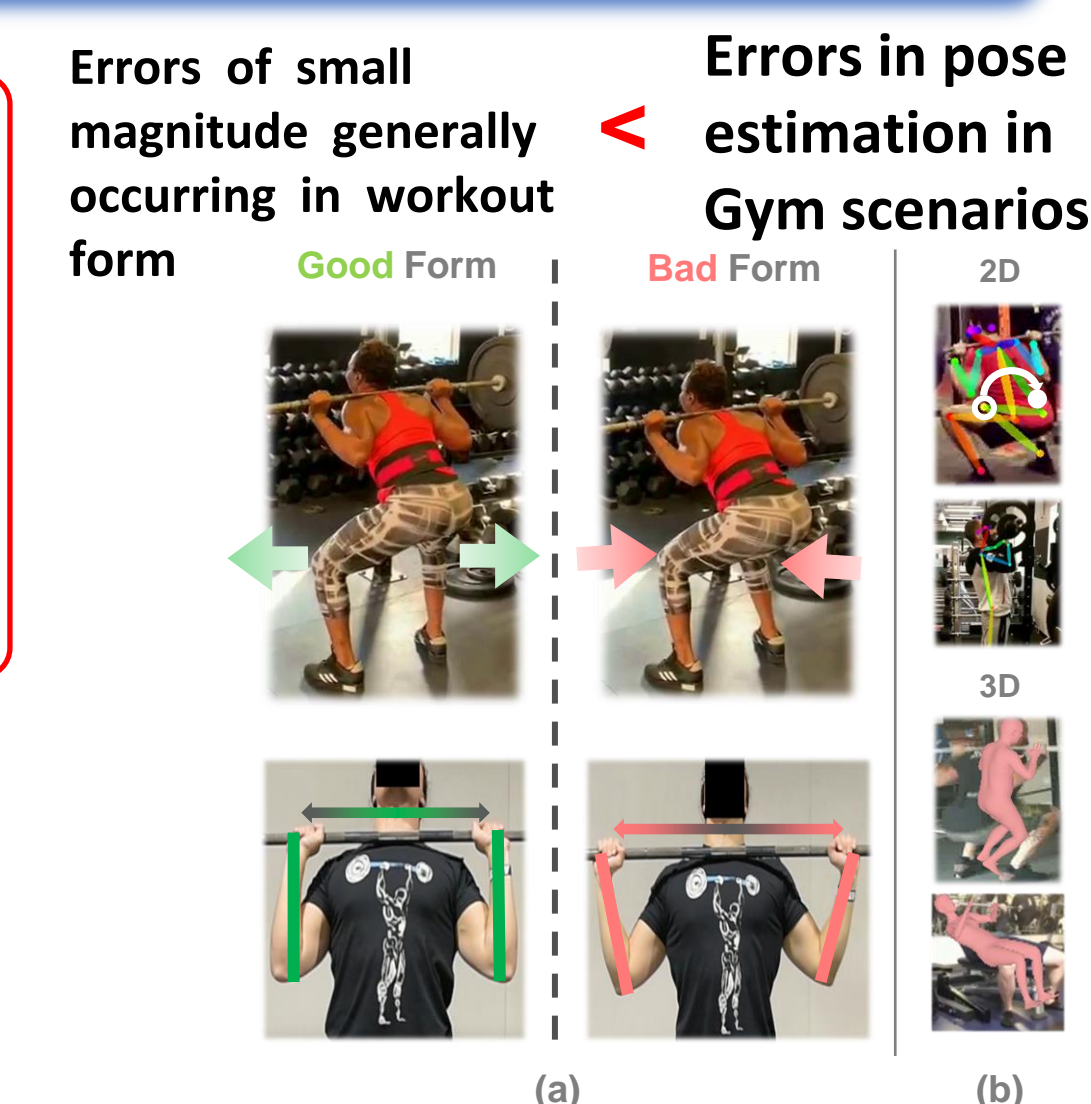
- Characteristics of exercises in the dataset:
 - Compound exercises – more likely to cause injuries than isolation exercises
 - Upper & Lower bodies covered
 - Targeting injury prone & complex joints: shoulders, knees, hips, spine, wrists



Challenges of Fitness-AQA Dataset

People record themselves by using their cellphone cameras placed somewhere in the vicinity

- camera angles
- occlusion from gym equipment
- illumination
- clothing

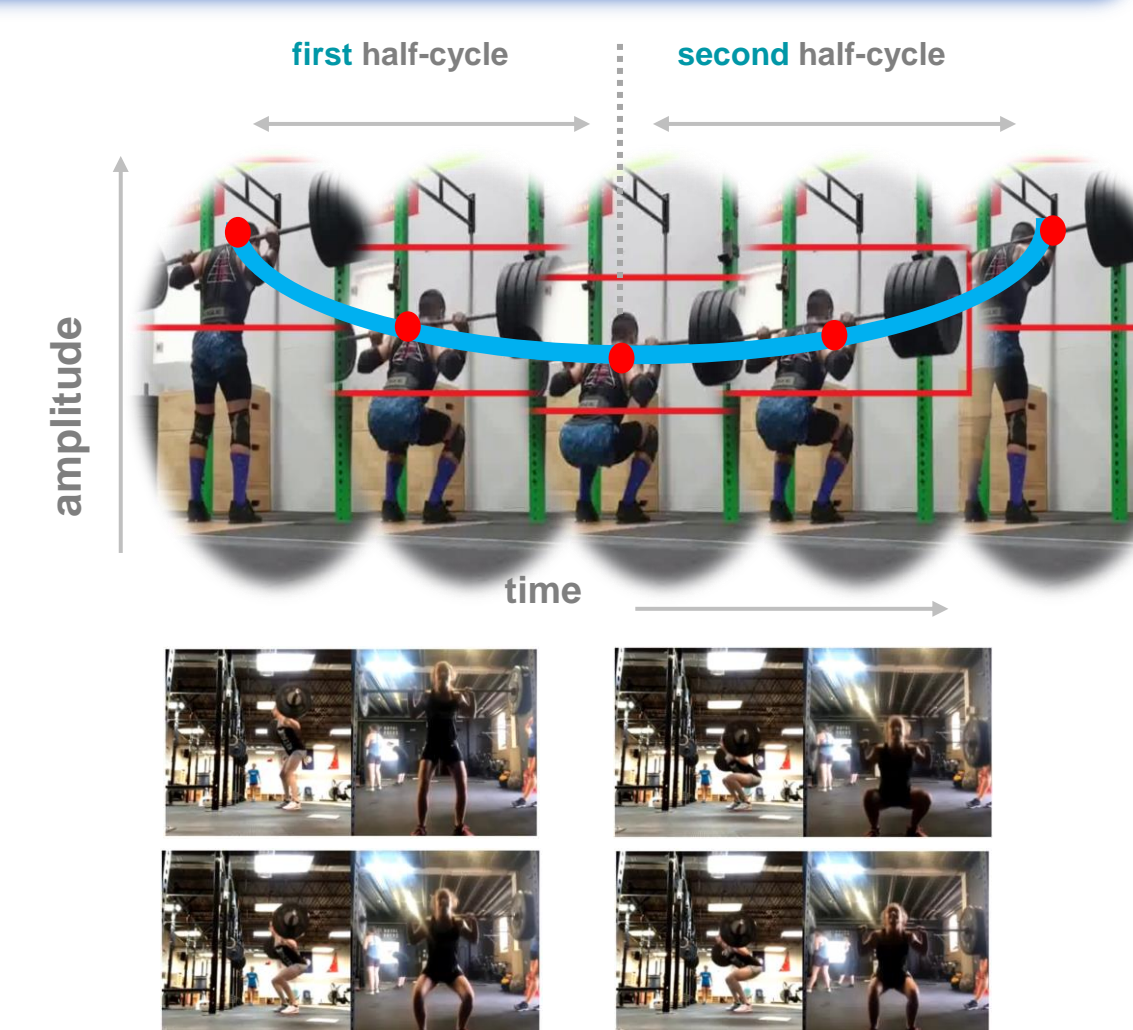


Our Proposal

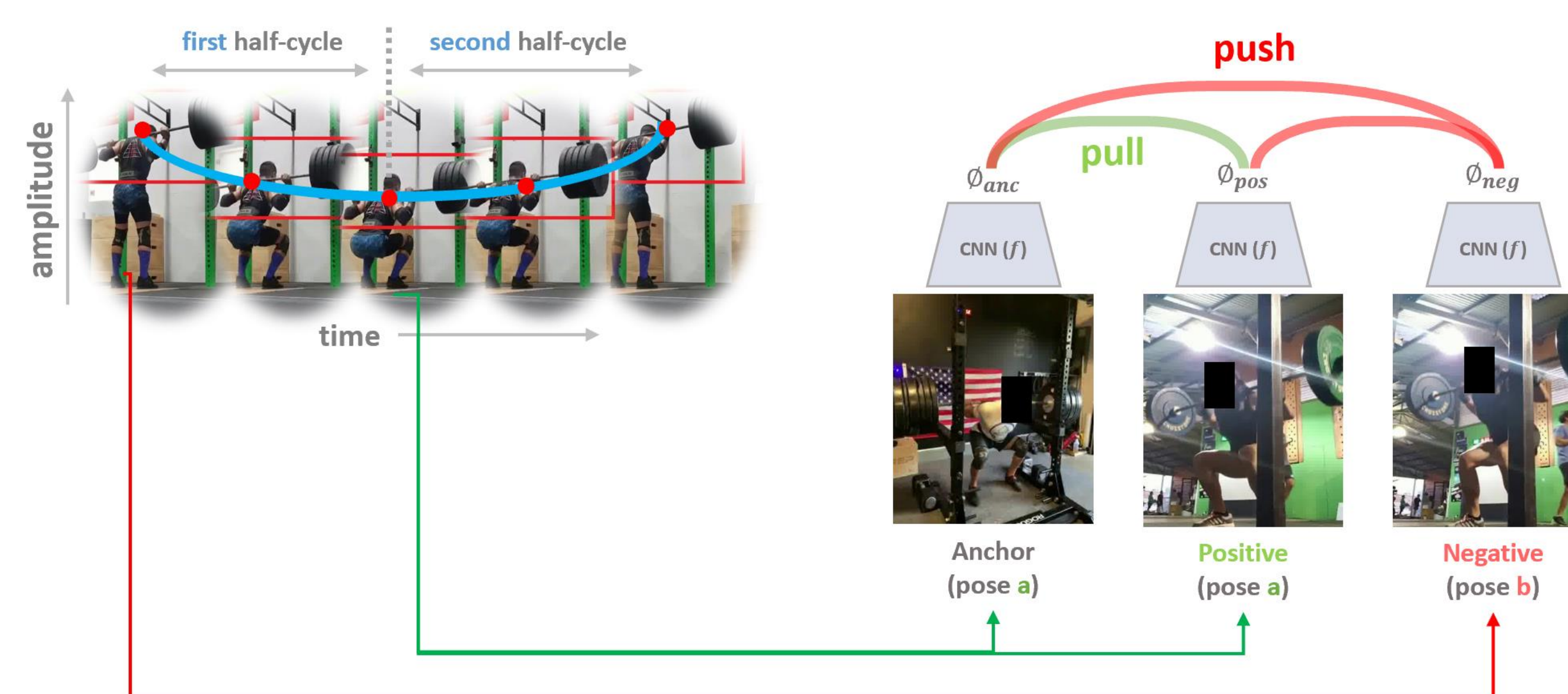
- Replace error-prone pose estimators with Self-Supervised Pose-sensitive representations learned from unlabeled real-world videos
- Map these self-supervised representations to errors-labels using smaller labeled datasets

Contribution 2: Quasi-Synchronizing Videos

- Steps to Quasi-synchronize videos:
 - Track the weight to get trajectories along the time direction
 - Normalize the amplitudes of these trajectories
- At any given amplitude, the people doing the same exercise would approximately be in the same pose

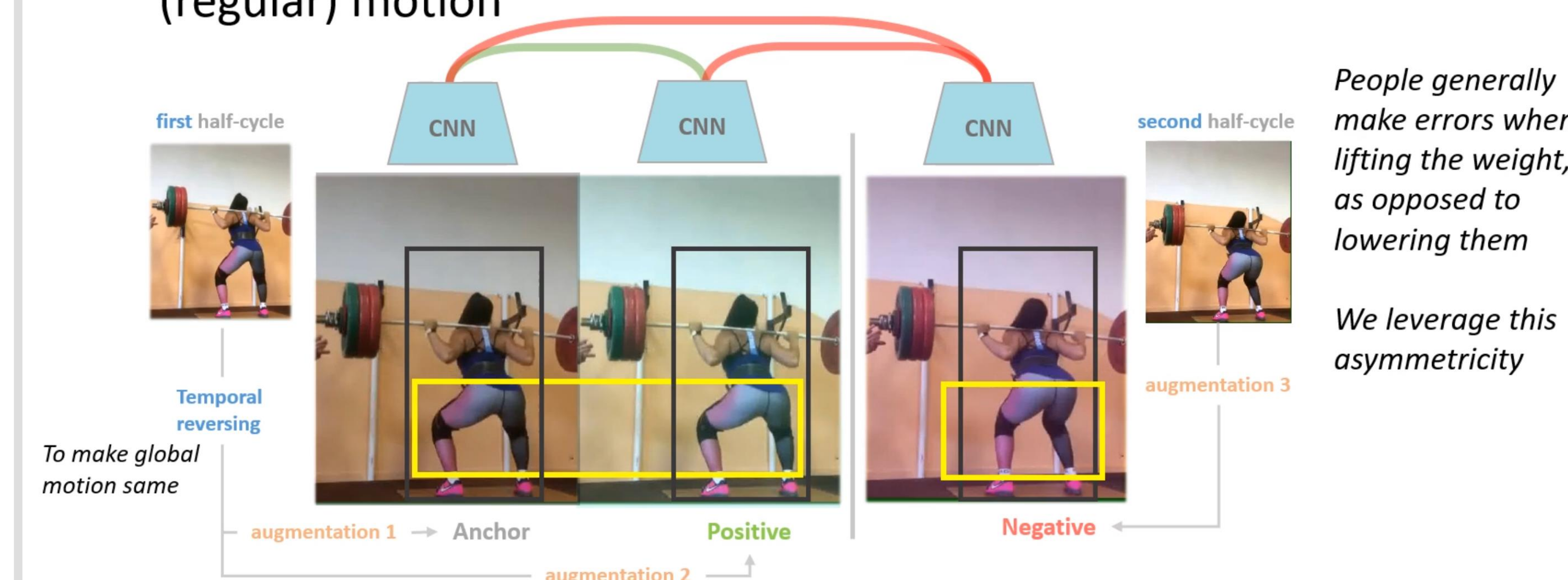


Contribution 3: Self-Supervised Pose Contrastive Learning



Contribution 3: Self-Supervised Motion Disentangling

- Objective: Separate **local** (irregular/erroneous) motion from **global** (regular) motion



Exp. 1

Features Extraction	Accuracies (%)					
	KIE	CVRB	CCRB	SS	KFE	Avg.
HMR-TDM [21]	89.80	98.65	93.05	87.30	83.58	89.08
Ours CVCSPC	95.92	91.89	94.44	77.77	89.55	89.92

Exp. 2

Feature extraction model	Modality	F-score ↑	
		KIE	KFE
OpenPose-TDM [2, 27]	2D Pose	0.4143	0.8123
OpenPose-TDM* [2, 27]	2D Pose	0.3186	0.7968
SPIN-TDM [22, 27]	3D Pose	0.2878	0.7761
ImageNet [39]	Image	0.1923	0.7725
SimSiam [5]	Image	0.2270	0.7868
Ours PAD	Image	0.3180	0.7784
Ours Vanilla PC	Image	0.4118	0.7965
Ours CVCSPC	Image	0.5195	0.8286
Kinetics [20]	Video	0.2970	0.8184
VideoSpeed-1 [1]	Video	0.3095	0.8155
VideoSpeed-2	Video	0.3617	0.8000
VideoRot [18]	Video	0.3333	0.8138
TemporalXform [17]	Video	0.3414	0.8319
Ours TemporalXform-1	Video	0.3457	0.8097
Ours TemporalXform-2	Video	0.2286	0.8184
Ours MD	Video	0.4186	0.8338
Ours MD + CVCSPC	Image, Video	0.5263	0.8468

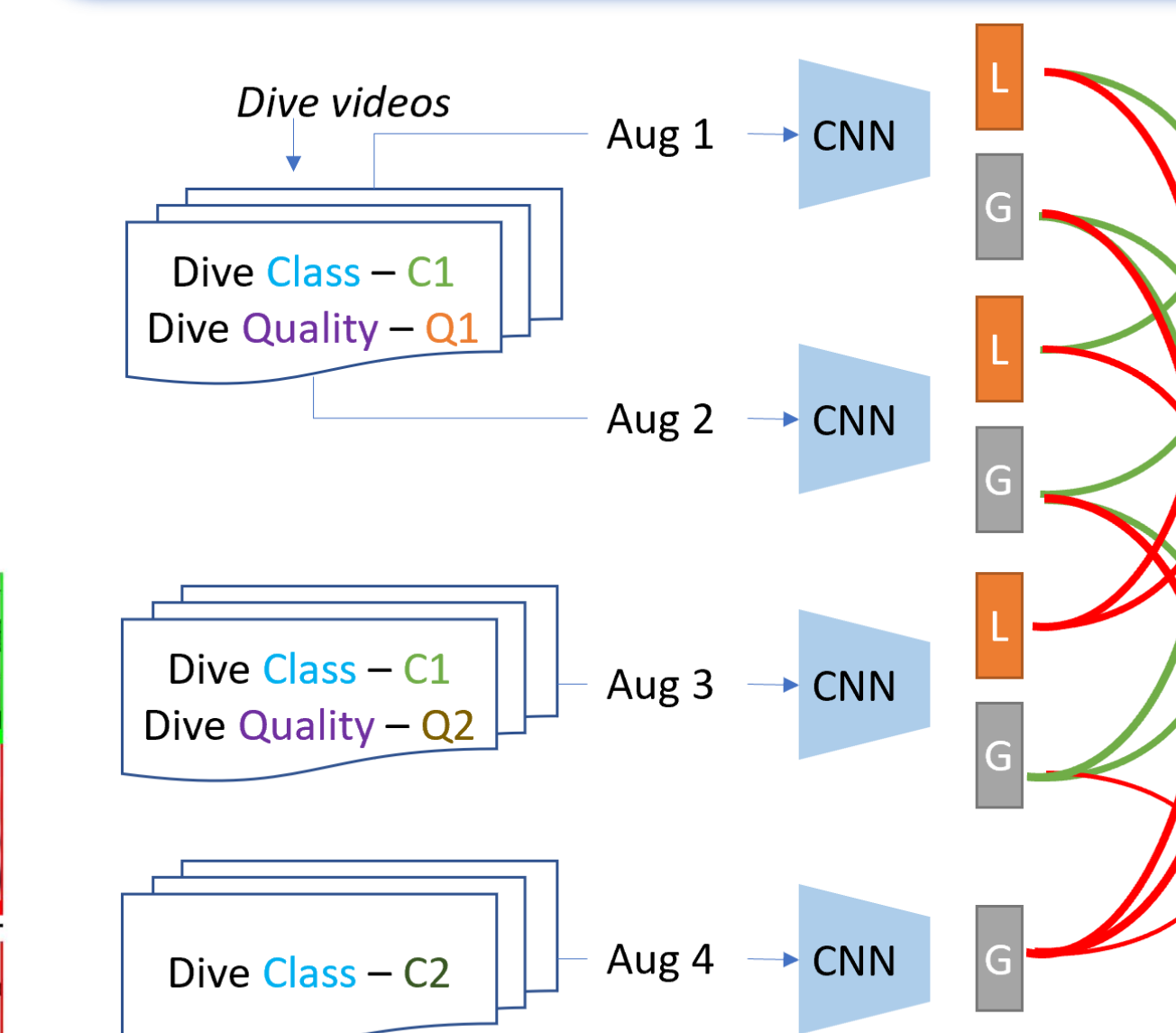
Exp. 3

Feature extraction model	Modality	F-score ↑		
		Elbow Err.	Knees Err.	
OpenPose-TDM [2, 27]	2D Pose	0.4265	0.7131	
SimSiam [5]	Image	0.4145	0.5301	
Ours CVCSPC	Image	0.4522	0.7203	
TemporalXform [17]	Video	0.4138	0.8416	
Ours MD	Video	0.4552	0.8452	

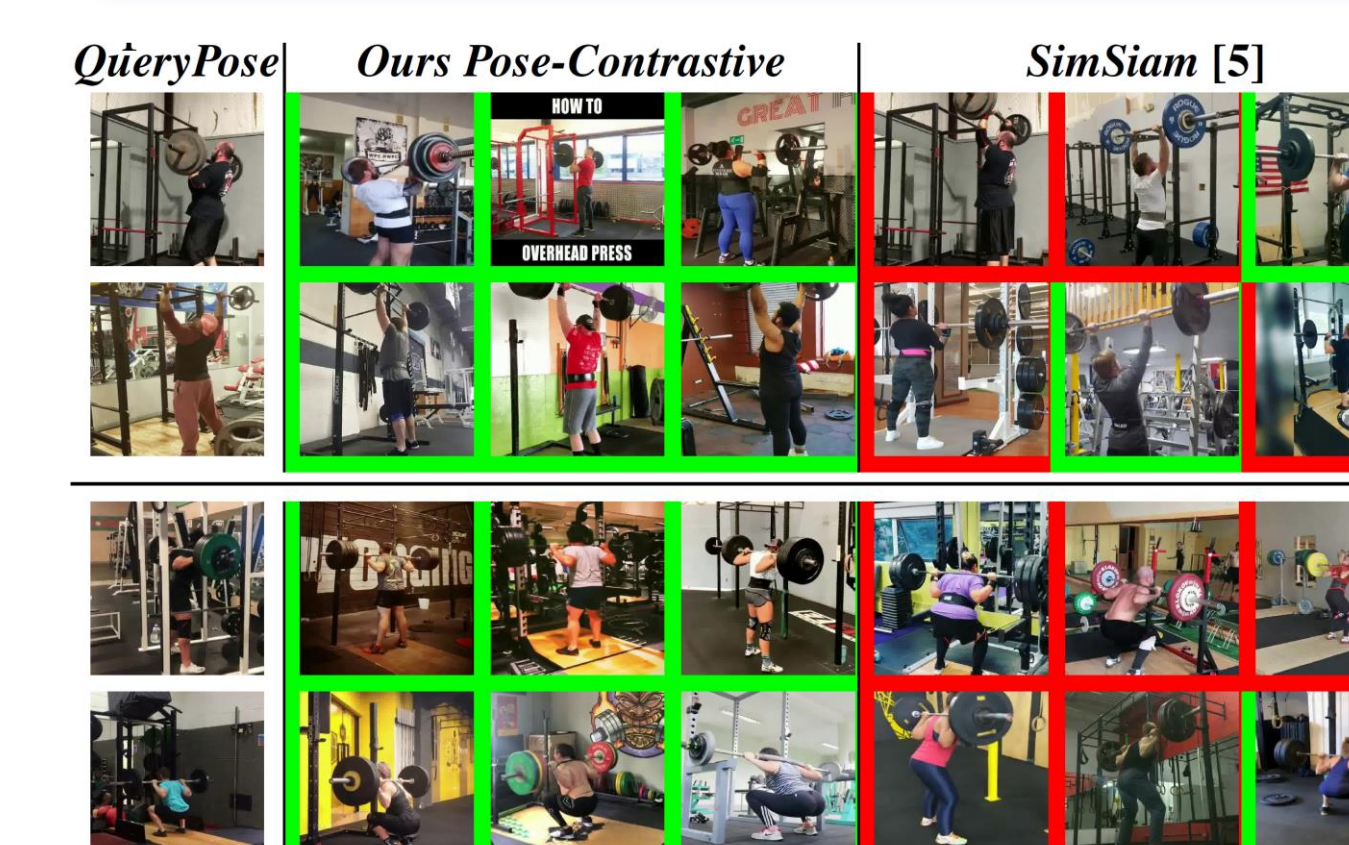
Exp. 4 – Xfer Repr.

Feature extraction model	Modality	F-score ↑		
		Lumbar Err.	Torso Err.	
OpenPose-TDM [2, 27] (SQ→BR)	2D Pose	0.5422	0.4060	
SimSiam [5] (SQ→BR)	Image	0.5934	0.4543	
Ours CVCSPC (SQ→BR)	Image	0.6057	0.4800	
Ours CVCSPC (OHP→BR)	Image	0.5760	0.4675	
Ours CVCSPC (SQ+OHP→BR)	Image	0.6338	0.5261	

Motion Disentangling for Scoring Olympic Dives



Pose Retrieval Exp.



Model	SSL SoTA [38]	Ours baseline	Ours MD
Sp. Corr.	0.7700	0.5665	0.7763