A. PennAction Breakdown

In this section, we present a detailed breakdown of **LRProp** performance on each action in the PennAction dataset. The results are summarized in Table 4.

Dataset	τ	*Progress*	*AP@*							*Classification@*						*DTW-A*
			K=1	*K=5*	*K=10*	*K=15*	K=20	K=25	K=30	5	*10*	*25*	*50*	75	*100*	
squat	99.58	97.08	90.9	90.73	90.22	90.11	89.95	89.83	89.74	91.19	90.43	91	91.07	91	91.57	87.4
pushup	99.08	92.7	92.58	92.53	92.39	92.32	92.2	92.12	92.02	90.77	89.27	91.86	93.17	93.25	93.19	92.35
tennis serve	98.66	95.54	90.05	89.44	89.2	89.03	88.76	88.58	88.47	89.4	91.31	91.52	91.17	91.74	91.69	87.85
tennis forhand	98.85	88.28	92.14	91.03	90.89	90.61	90.47	90.36	90.3	92.53	92.44	93.14	92.68	93.14	93.08	89.66
situp	99.33	95.39	96.83	96.81	96.7	96.63	96.51	96.41	96.4	96.81	96.72	97.04	96.97	97.08	97.08	95.74
pullup	99.23	96.8	96.73	96.52	96.46	96.39	96.33	96.23	96.18	96.4	96.34	95.66	95.54	96.13	96.27	95.25
jumping jacks	98.63	97.02	91.91	91.84	91.74	91.67	91.44	91.03	90.68	92.11	93.41	93.58	93.58	93.7	93.76	90.08
golf swing	98.92	98.19	95.02	95.3	94.98	94.64	94.46	94.26	94.14	94.92	94.92	94.97	95.28	95.37	95.4	92.69
bowl	98.84	76.21	92.33	90.65	89.73	89.13	88.56	88.22	87.85	84.57	85.02	84.53	84.9	85.25	85.68	83.88
benchpress	99.32	95.14	94.33	93.96	93.9	93.75	93.65	93.55	93.49	91.38	93.81	94.98	95.11	95.19	95.33	92.57
baseball swing	99.2	94.9	91.61	92.11	91.87	91.88	91.78	91.63	91.49	97.05	89.59	92.49	93.03	93.19	93.11	90.07
baseball pitch	99.35	98.05	92.92	92.91	92.73	92.59	92.55	92.43	92.32	92.18	93.11	94.62	94.94	95.18	95.29	90.93
clean and jerk	99.2	87.64	88.67	88.24	87.84	87.66	87.5	87.33	87.2	87.89	88.62	90.31	91.14	90.99	90.8	83.59

Table 4. **PennAction breakdown**. A breakdown of the performance of **LRProp** on each of the actions contained in the PennAction dataset, using various evaluation metrics: Phase Classification@% (Classification@%), Phase Progression (Progress), Kendall's Tau (τ), Average Precision@K (AP@K) and DTW Accuracy (DTW A).

B. Implementation Details

In our model, we use a ResNet-50 [21] pretrained by BYOL [16] as a frame-wise spatial encoder. We use a 7-layer Transformer encoder [39] with a hidden size of 256 and 8 heads to model temporal context. We set $\sigma^2 = 10$, recall Equations (2), (7), and $\tau = 0.1$, recall Equation (1). We use the Adam optimizer with a learning rate of 10^{-4} and a weight decay of 10^{-5}, and we apply a cosine decay schedule without restarts [29] to the learning rate. We fix a batch size of 2, and train for 300 epochs all considered datasets. We use the same spatial augmentations and temporal sampling technique as in [7, 8]. The number of sampled frames, T, and the regularization parameters, λ_1, λ_2, recall Equation (12), are given in Table 5.

Dataset	*Action class*	*T*	λ_1	λ_2
PennAction	jumping jacks	20	0.01	0.4
	tennis serve	40		0.003
	tennis forhand			
	baseball swing			0.001
	bowl	60		
	situp	80		0.4
	benchpress			
	baseball pitch			
	golf swing			1.6
	pushup			
	squat			0.8
	pullup			
	clean and jerk			
Pouring	-	240		

Table 5. Hyper-parameters of **LRProp** regarding each dataset.