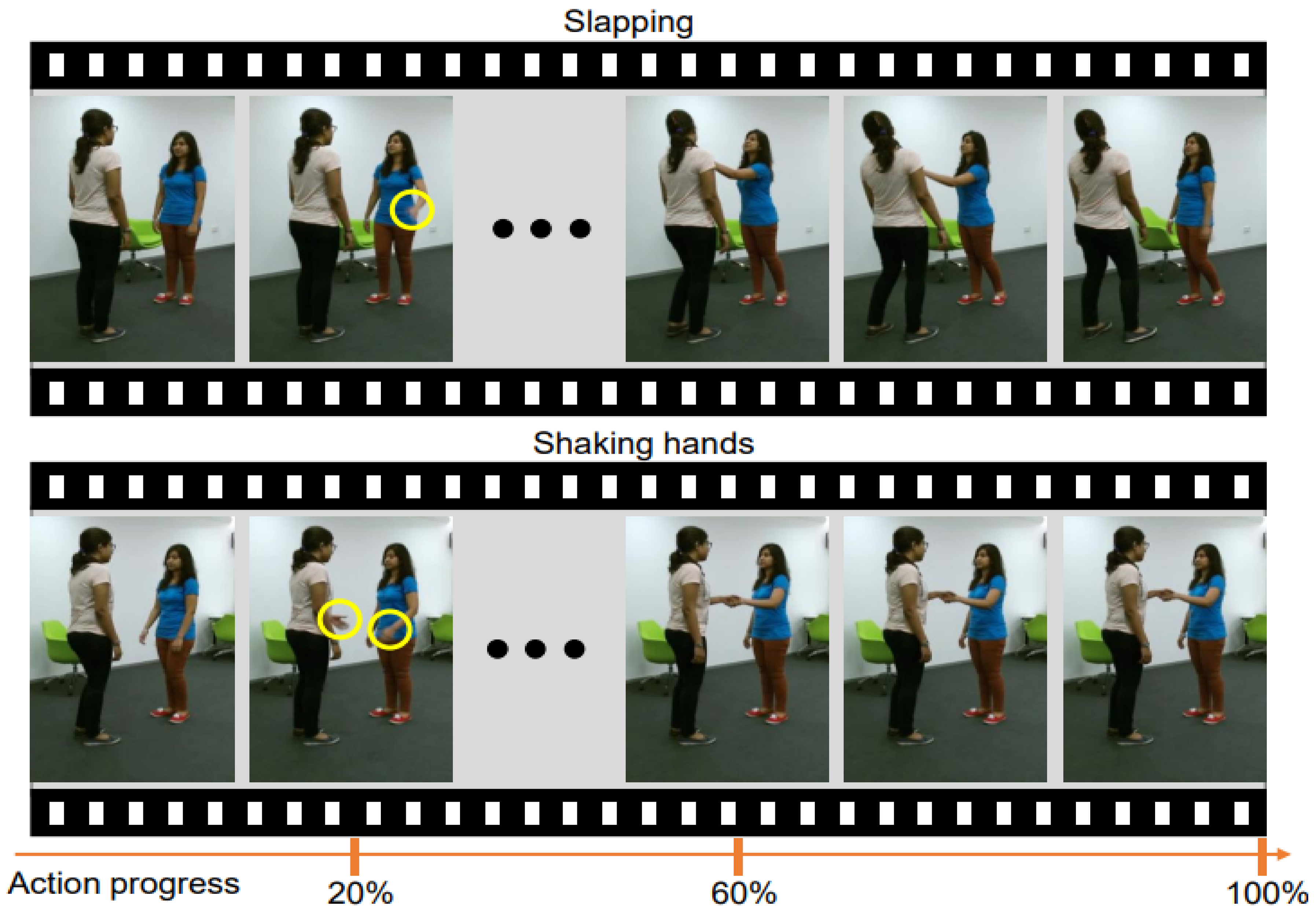
ERA: Expert Retrieval and Assembly for Early Action Prediction

Lin Geng Foo[1*], Tianjiao Li[1*], Hossein Rahmani[2], Qiuhong Ke[3], and Jun Liu[1]
[1] Singapore University of Technology and Design, [2] Lancaster University, [3] University of Monash

Introduction



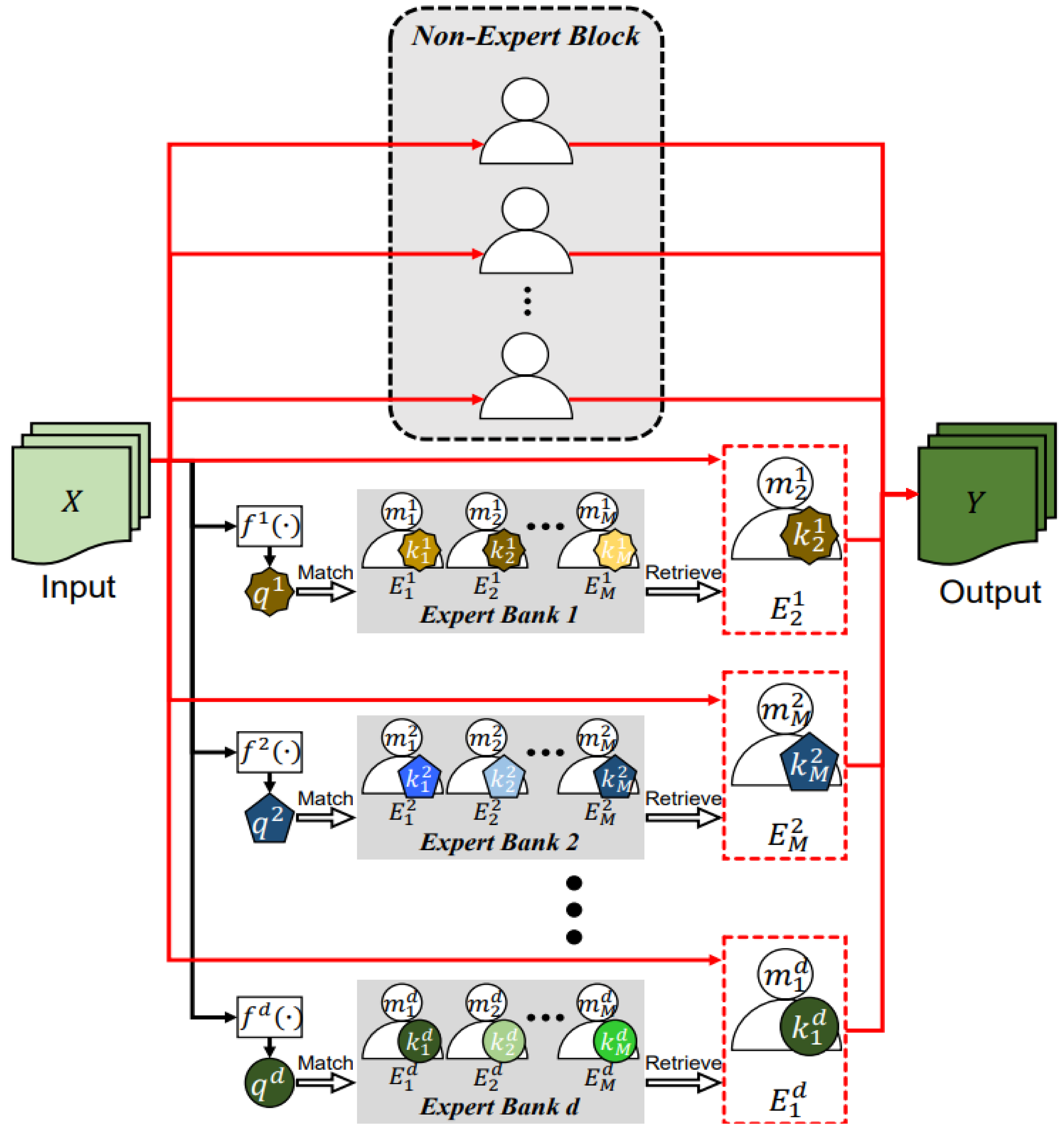
Goal: To successfully predict the action class of an action before it is fully performed.

Motivation: As shown in the diagram, early action prediction is a quite challenging problem because the beginning stages of different actions can be very similar, with only minor subtle differences for discrimination.

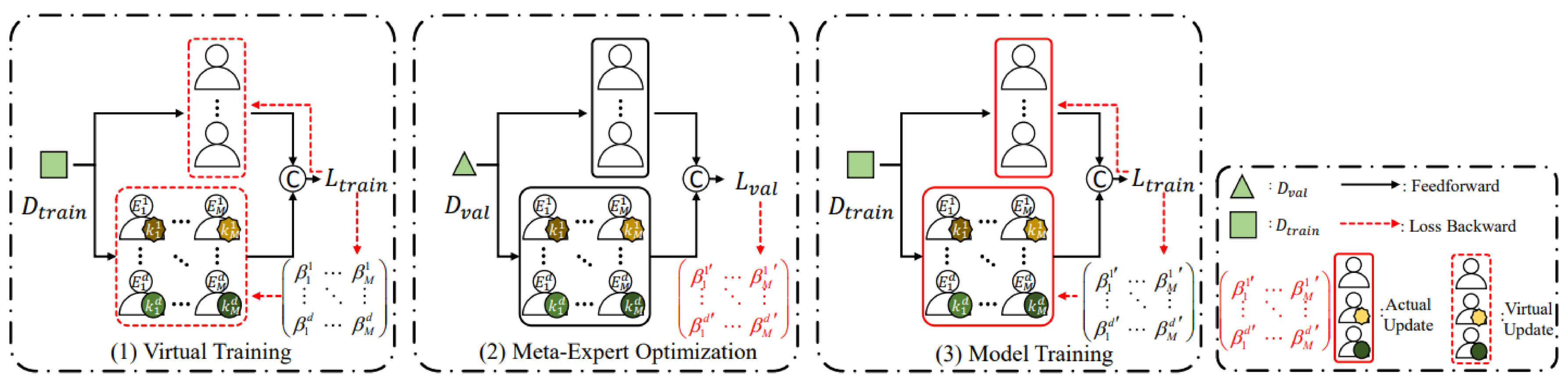
ERA-Net

In this work, to improve the performance on early action prediction, we propose an Expert Retrieval and Assembly (ERA) module that contains non-experts and experts. Unlike non-experts that contain parameters which are shared across all samples and capture general patterns that exist in many samples, experts are only trained on a subset of the data and contain parameters that focus on encoding subtle differences to distinguish between highly similar samples.

$$q^p = f^p(X) \Rightarrow \begin{matrix} s_i^p = q^p \cdot k_i^p \\ I^p = Argmax_i(\{s_i^p\}_{i=1}^M) \end{matrix} \Rightarrow W_{expert} = Concat(\{m_{I^p}^p\}_{p=1}^d)$$



Expert Learning Rate Optimization



Experts in ERA modules are end-to-end trainable using backpropagation. However, due to the uneven distribution of samples across experts, some experts might be selected by more samples and be better trained than others, possibly causing imbalanced training that limits the performance of our ERA module. To mitigate this effect, we design an Expert Learning Rate Optimization (ELRO) method that optimizes the training among experts, leading to improved early action prediction accuracy.

Experimental Results

Methods	Observation Ratios on NTU60						Observation Ratios on SYSU					
	20%	40%	60%	80%	100%	AUC	20%	40%	60%	80%	100%	AUC
Jain *et al.* [18]	7.07	18.98	44.55	63.84	71.09	37.38	31.61	53.37	68.71	73.96	75.53	57.23
Ke *et al.* [21]	8.34	26.97	56.78	75.13	80.43	45.63	26.76	52.86	72.32	79.40	80.71	58.89
Kong *et al.* [26]	-	-	-	-	-	-	51.75	58.83	67.17	73.83	74.67	61.33
Ma *et al.* [35]	-	-	-	-	-	-	57.08	71.25	75.42	77.50	76.67	67.85
Weng *et al.* [55]	35.56	54.63	67.08	72.91	75.53	57.51	-	-	-	-	-	-
Aliakbarian *et al.* [41]	27.41	59.26	72.43	78.10	79.09	59.98	56.11	71.01	78.39	80.31	78.50	69.12
Hu *et al.* [16]	-	-	-	-	-	-	56.67	75.42	80.42	82.50	79.58	71.25
Wang *et al.* [54]	35.85	58.45	73.86	80.06	82.01	60.97	63.33	75.00	81.67	86.25	87.92	74.31
Pang *et al.* [39]	33.30	56.94	74.50	80.51	81.54	61.07	-	-	-	-	-	-
Tran *et al.* [50]	24.60	57.70	76.90	85.70	88.10	62.80	-	-	-	-	-	-
Ke *et al.* [22]	32.12	63.82	77.02	82.45	83.19	64.22	58.81	74.21	82.18	84.42	83.14	72.55
HARD-Net [30]	42.39	72.24	82.99	86.75	87.54	70.56	-	-	-	-	-	-
Baseline	38.09	66.36	78.67	83.29	84.10	66.43	60.71	73.04	77.81	83.88	84.32	72.20
ERA-Net w/o ELRO	43.94	73.23	84.53	87.61	87.97	71.62	63.50	80.82	82.70	86.33	87.10	75.78
ERA-Net	**53.98**	**74.34**	**85.03**	**88.35**	**88.45**	**73.87**	**65.30**	**81.27**	**85.67**	**89.17**	**89.38**	**77.73**

Methods	Observation Ratios on NTU120						Observation Ratios on UCF101					
	20%	40%	60%	80%	100%	AUC	10%	30%	50%	70%	90%	AUC
MSRNN [16]	-	-	-	-	-	-	68.01	88.71	89.25	89.92	90.23	80.89
Wu *et al.* [56]	-	-	-	-	-	-	80.24	84.55	86.28	87.53	88.24	80.57
Wu *et al.* [57]	-	-	-	-	-	-	82.36	88.97	91.32	92.41	93.02	84.66
Wang *et al.* [54]	-	-	-	-	-	-	83.32	88.92	90.85	91.28	91.31	89.64
Baseline	23.14	32.49	59.07	75.61	81.18	50.03	82.88	89.02	89.64	91.12	91.96	89.30
ERA-Net w/o ELRO	29.60	43.45	65.14	78.03	82.01	55.52	86.99	91.49	93.63	94.24	94.40	92.51
ERA-Net	**31.73**	**45.67**	**67.08**	**78.84**	**82.43**	**57.02**	**89.14**	**92.39**	**94.29**	**95.45**	**95.72**	**93.64**

To validate effectiveness of our ERA module for early action prediction, we conduct extensive experiments on both skeletal and RGB datasets. We experiment on the NTU RGB+D 60 (NTU60), NTU RGB+D 120 (NTU120) and SYSU datasets for skeletal data, and the UCF-101 (UCF101) dataset for RGB data. As shown in the tables above, our ERA-Net achieves the state-of-the-art performances on all datasets.

Conclusion

In this paper, we have proposed a novel plug-and-play ERA module for early action prediction. To encourage the experts to effectively use subtle differences for early action prediction, we push them to discriminate exclusively among similar samples. An Expert Learning Rate Optimization algorithm is further proposed to balance the training among numerous experts, which improves performance.