Action Shuffle Alternating Learning for Unsupervised Action Segmentation

Jun Li, Sinisa Todorovic

Oregon State University

CVPR VIRTUAL JUNE 19-25

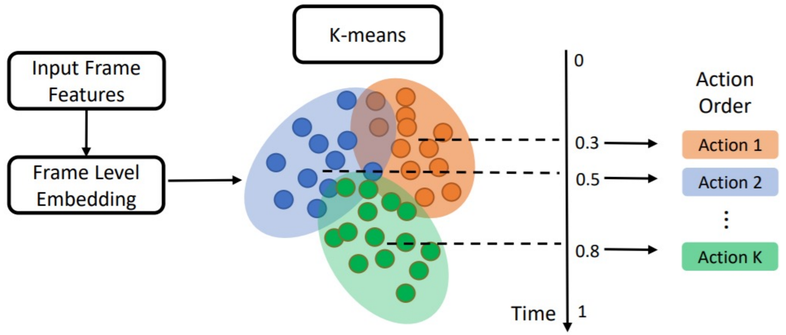
Problem:

Localize salient latent actions when there is no ground truth provided in training.

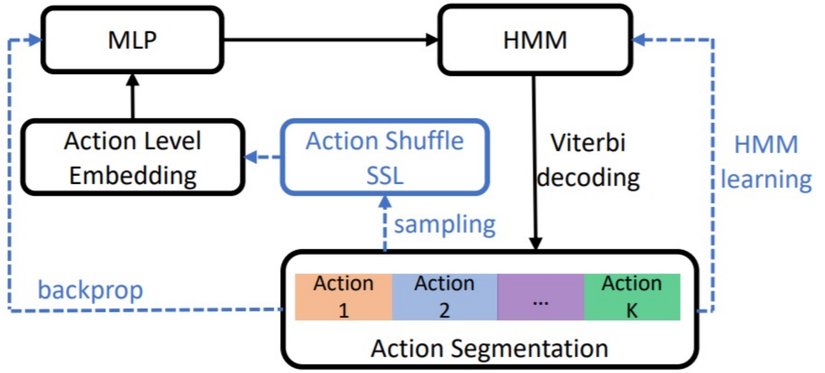
Key Ideas for Unsupervised Training:

- Alternate prediction of latent actions and training of HMM by using the predicted latent actions as pseudo-ground truth.
- Self-supervised learning of feature embedding by a temporal shuffling of the predicted action segments, rather than individual frames.

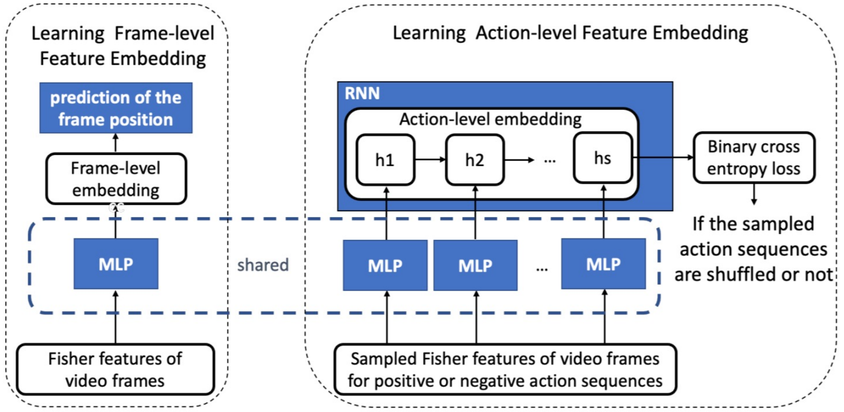
Stage 1: Temporal Self-Supervised Learning



Stage 2: Alternate Learning of the Model and Embedding



Details of Action Shuffle Alternating Learning



Results:

YouTube Instructions		
Unsupervised		
	F1-score	MoF
Frank-Wolfe [2]	24.4	-
Mallow [26]	27.0	27.8
CTE [18]	28.3	39.0
VTE-UNET [30]	29.9	-
Our ASAL	**32.1**	**44.9**

Summary:

- New self-supervised learning as a verification of the temporal ordering of action segments, not frames.
- Unified learning of the action-level embedding and HMM within the Generalized EM framework.
- Our approach outperforms the state of the art on challenging datasets.

Acknowledgement.

DARPA XAI N66001-17-2-4029,
DARPA MCSN66001-19-2-4035.