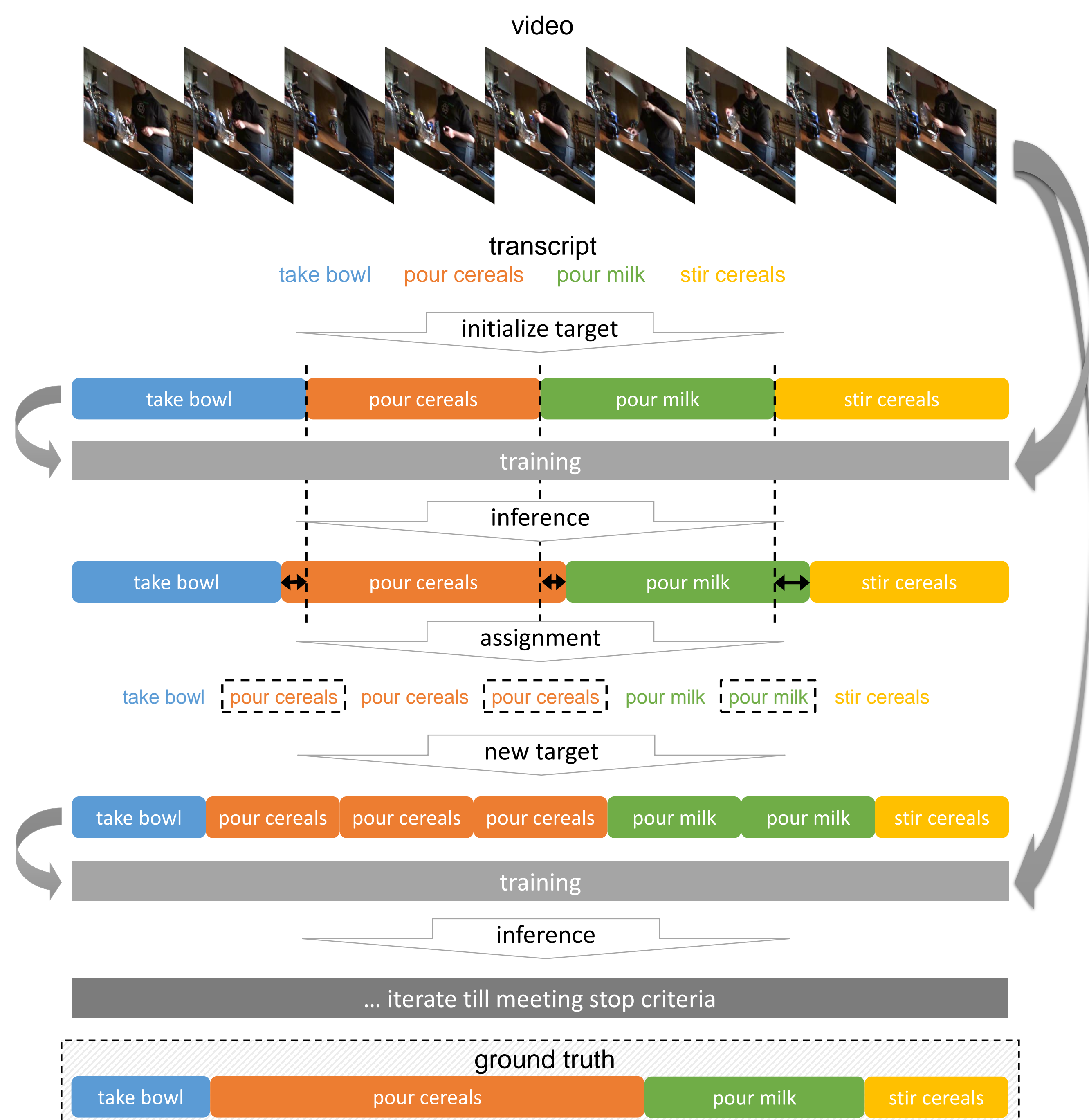


Overview of the Weakly-Supervised Action Segmentation System

- We iteratively train a temporal segmentation network with target generated from action transcript, and refine the action transcript based on the inference of current network.



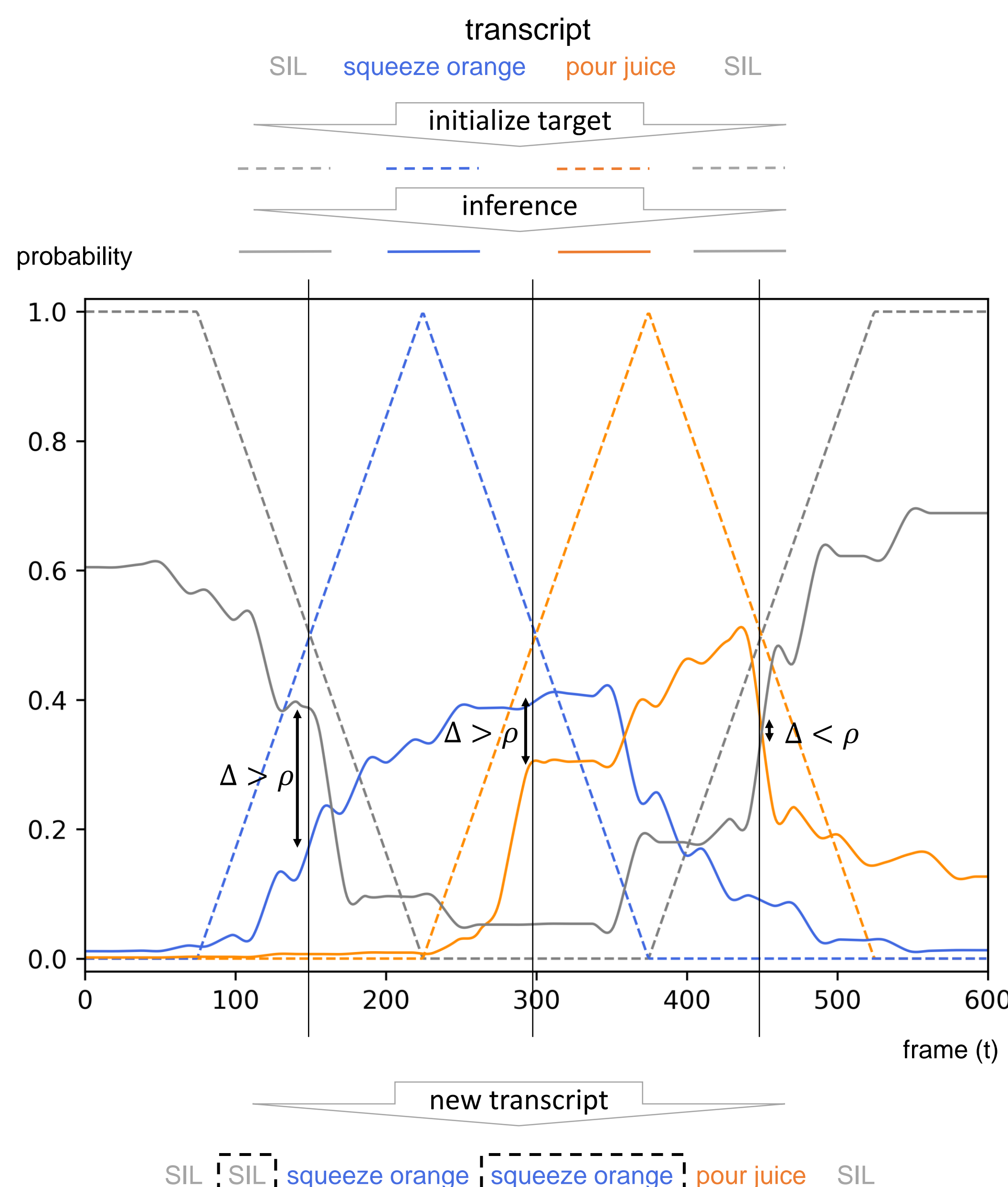
Iterative Soft Boundary Assignment (ISBA)

Algorithm 1 Transcript Refinement

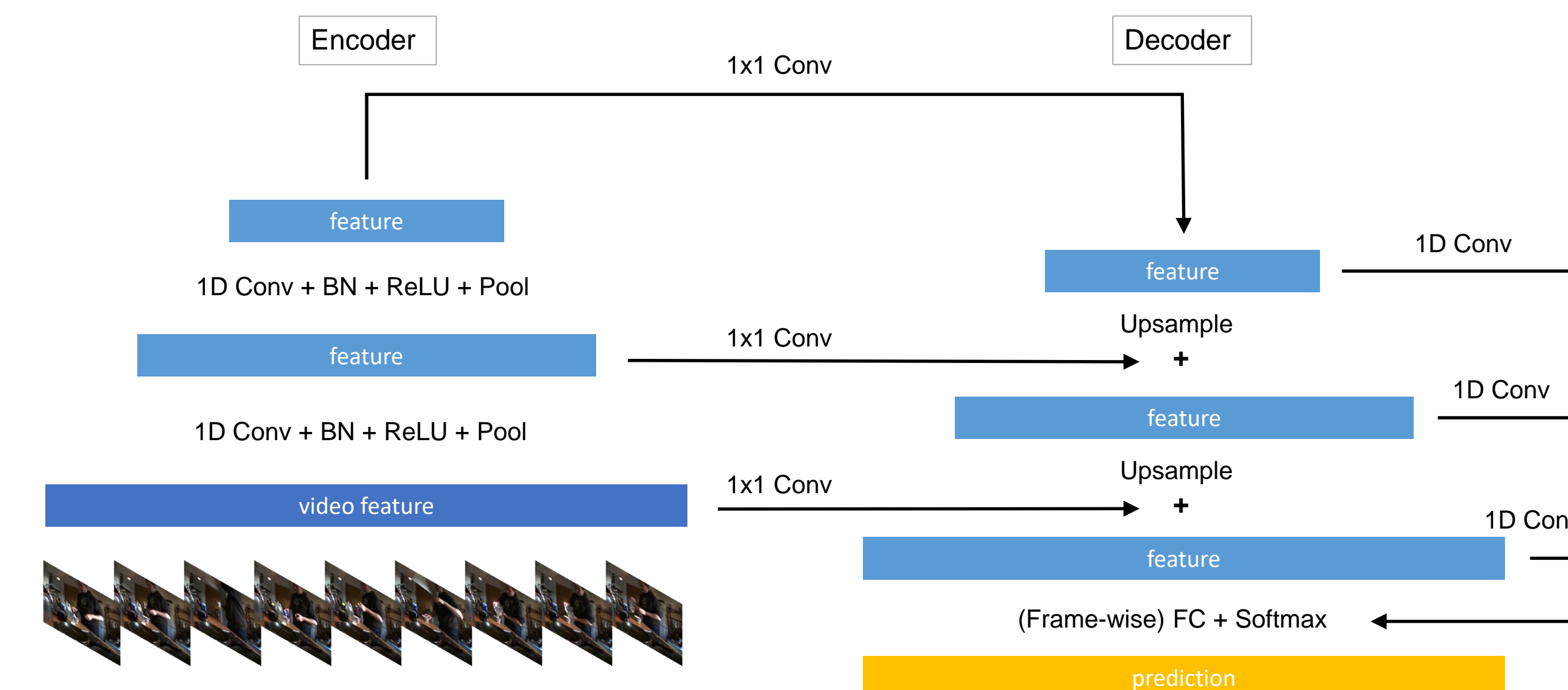
```

Given transcript in training set
for each action in length(transcript) do
  t = boundary_frame
  if (action ≠ next_action) then
    if  $P_t(action) - P_t(next\_action) > \rho$  then
      insert action into transcript at current location
    else if  $P_t(next\_action) - P_t(action) > \rho$  then
      insert next_action into transcript at current location
  end for
  
```

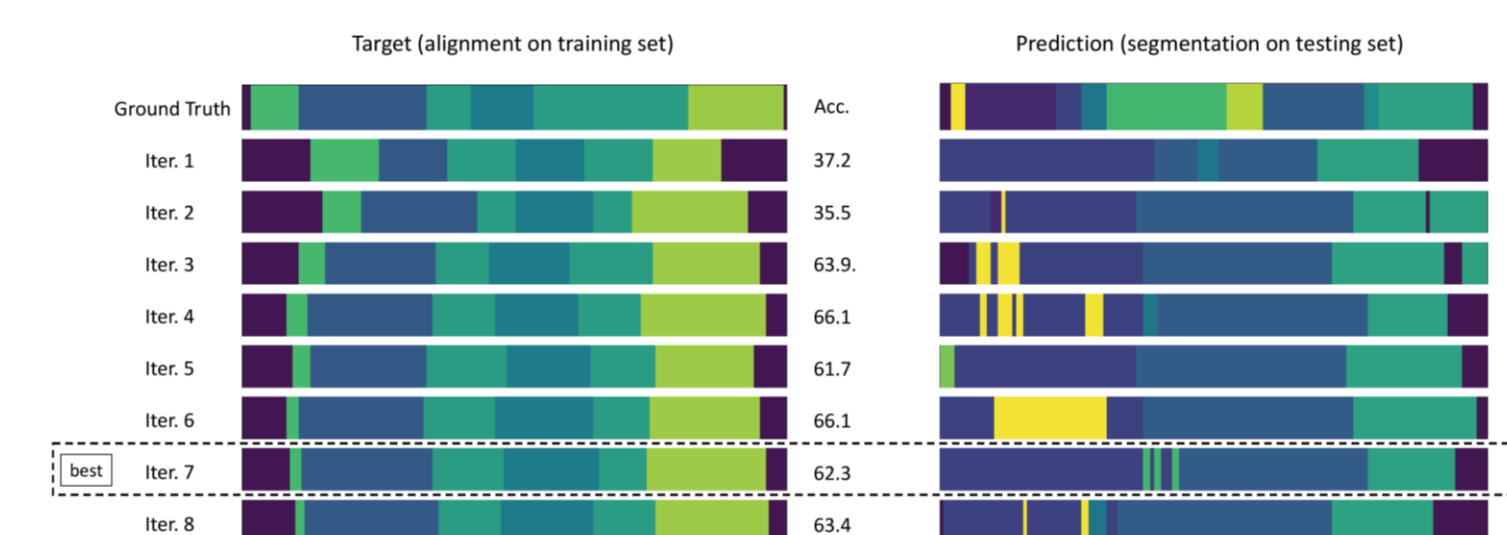
- ISBA uses a simple-yet-effective algorithm, trying to generate a reasonable probabilistic distribution of different actions through numbers of iterations.



Temporal Convolutional Feature Pyramid Network (TCFPN)



Experiment



Breakfast	Acc.	Acc.-b.g.	IoU	IoD
OCDC [2]*	8.9	-	-	-
HTK [10]	25.9	-	9.8	-
ECTC [6]	27.7	-	-	-
GRU reest. [16]	33.3	-	-	-
ED-TCN + ISBA	32.0	28.8	18.4	30.6
TCFPN + ISBA	38.4	38.4	24.2	40.6

Hollywood Ext.	Acc.	Acc.-b.g.	IoU	IoD
HTK [10]	33.0	-	8.6	-
GRU reest. [16]	-	-	11.9	-
ED-TCN + ISBA	27.8	29.4	8.9	10.7
TCFPN + ISBA	28.7	34.5	12.6	18.3

Conclusion

- We propose ISBA as a novel training strategy for weakly-supervised sequence learning, and TCFPN as an advanced temporal convolutional network for supervised action recognition.
- The whole system TCFPN+ISBA outperforms state-of-the-art on both action segmentation task and alignment tasks.
- Our training strategy for weakly-supervised sequence learning is general and can be extended to other tasks, such as speech recognition, video segmentation, etc.

Code



About

