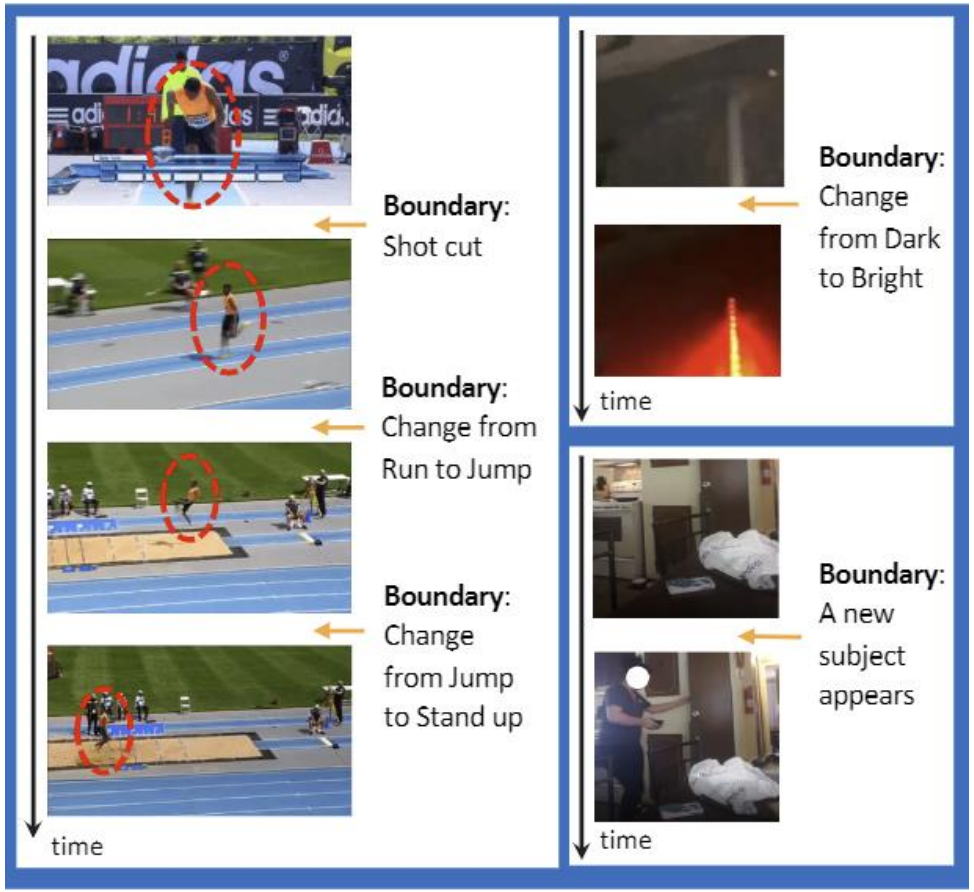
[CVPR-2022] UBoCo : Unsupervised Boundary Contrastive Learning for Generic Event Boundary Detection

2022-07-18

GEBD?



Generic Event Boundary Detection (GEBD), aiming at detecting generic, taxonomy-free event boundaries that segment a whole video into chunks.

GEBD?

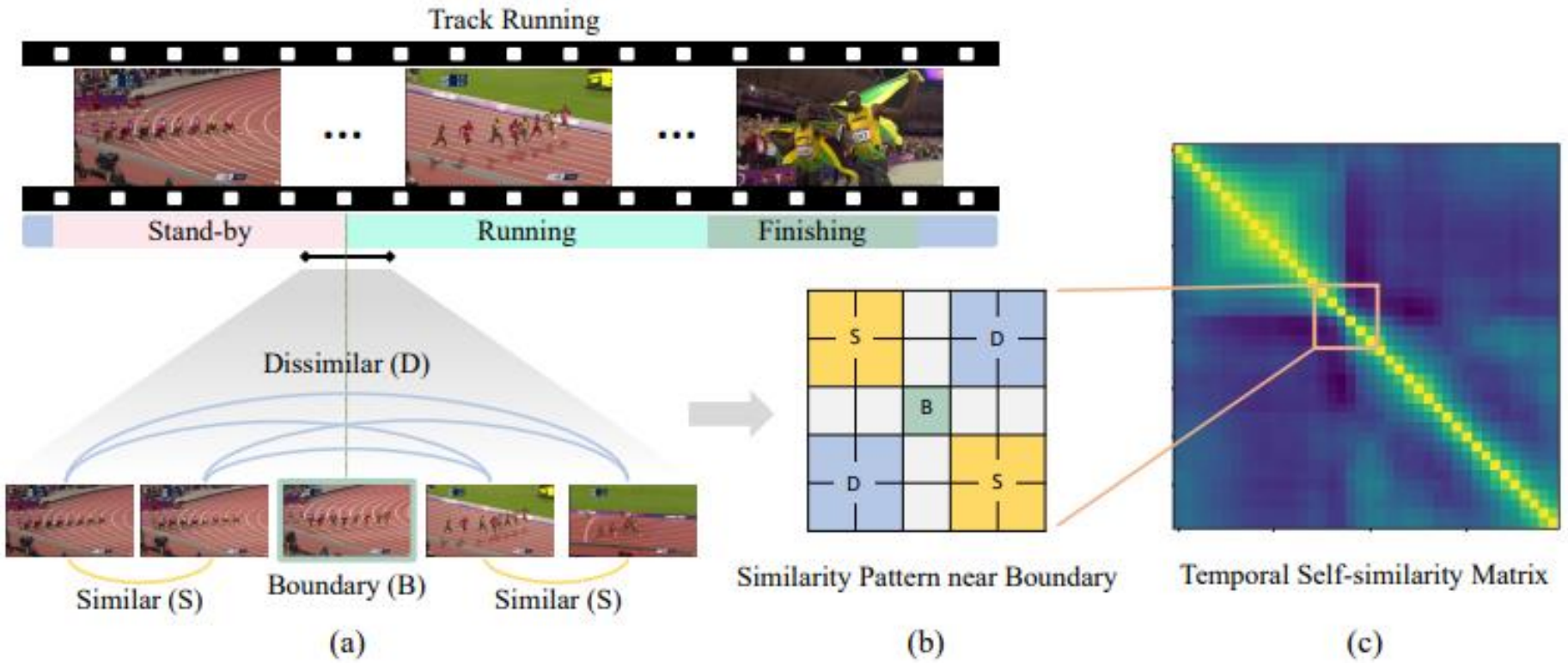
1. Change of Object
2. Change of Object of Interaction (Shot Change)
3. Change of Action

Contribution

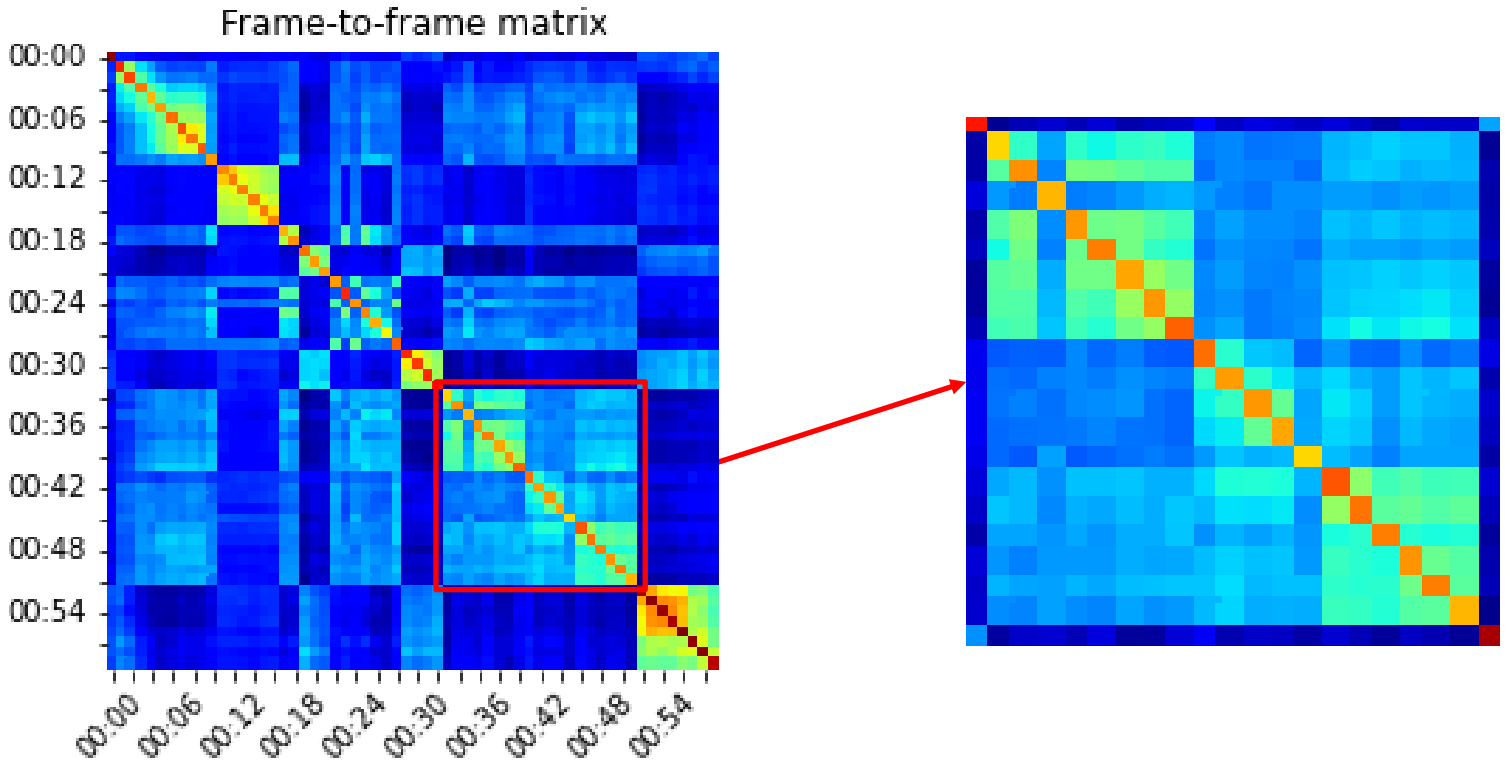
- Found that the properties of the Self-Similarity Matrix work well with GEBD, and propose to use this methodology in GEBD
- Propose Recursive TSM Parsing (RTP) algorithms, using boundary patterns, based on divide and conquer algorithm
- Propose BoCo Loss applicable to RTP, showing good performance in unsupervised learning methodology

Method

Method


Track Running
Stand-by
Running
Finishing
Dissimilar (D)
Similar (S)
Boundary (B)
Similar (S)
(a)
S
D
B
D
S
Similarity Pattern near Boundary
(b)
Temporal Self-similarity Matrix
(c)

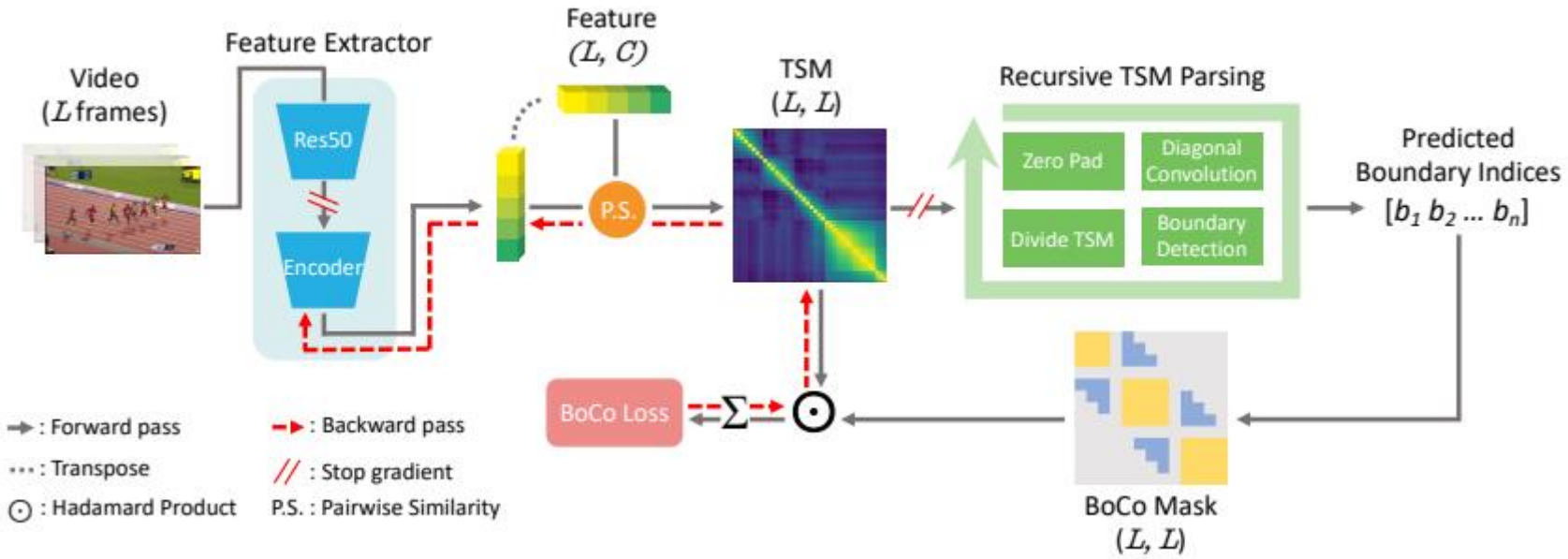
Method



Features that can be observed on the Self-similarity map

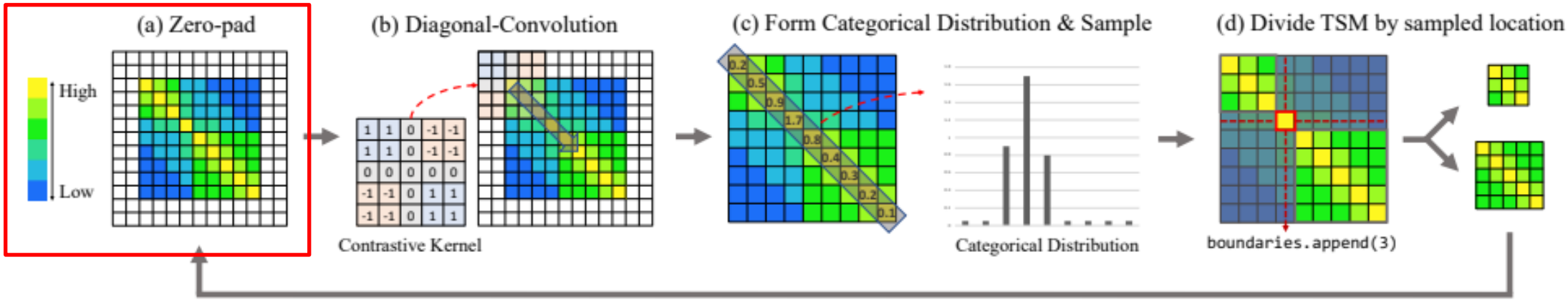
- A diagonal line is formed
- Boxes appear at the interface where the event changes

Method



- Extra encoder in Feature Extractor
- Calculate Self-similarity in TSM
- Predict boundaries with RTP
- Calculate BoCo Loss with BoCo Mask

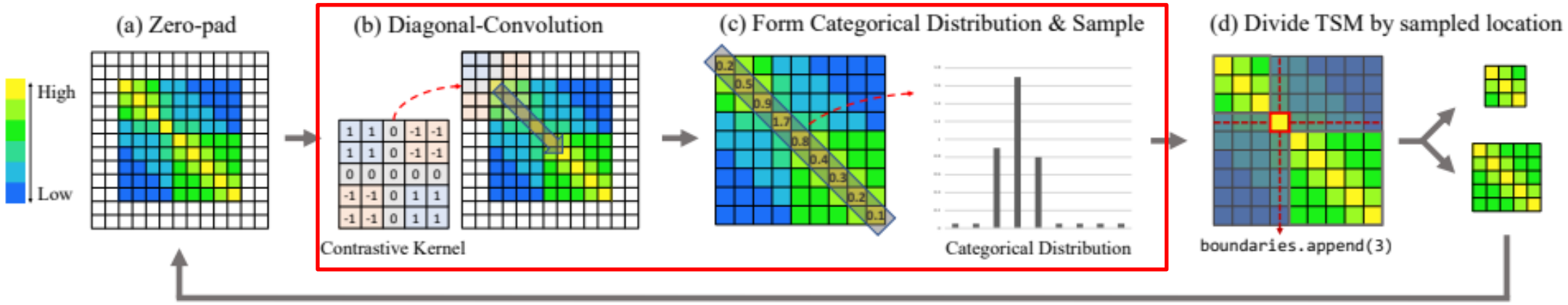
Method



- The beginning and end of the video are less likely to belong to the boundary.
- The beginning and end of the video do not belong to the boundary, because RTP divides the video based on boundary score.

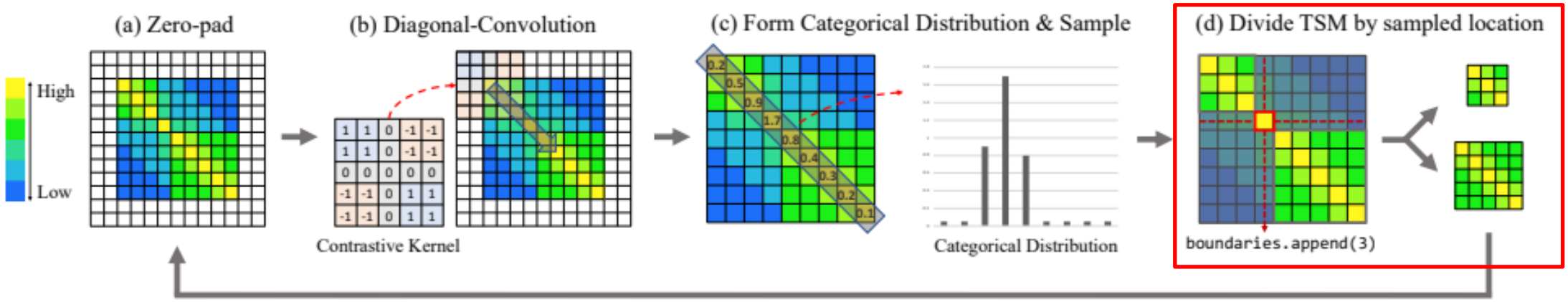
=〉 Adding Zero-pad significantly reduces the probability that the start and end are detected as boundary.

Method



- Diagonal Conv & score calculation
  - Contrastive Kernel based on similarity pattern
  - Based on knowledge that pattern is associated with near by frames
- In the case of boundary, a point with a probability of Top K% is selected as boundary

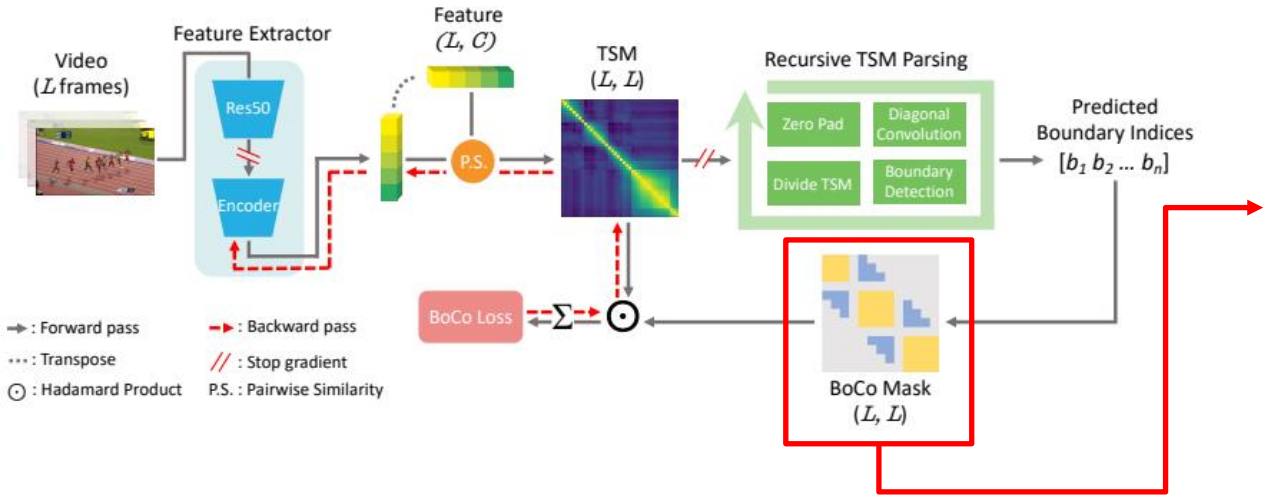
Method



TSM is based on “Divide and Conquer” algorithm. Thus, it has an exit status.

- Length of TSM is lower than pre-defined T1
- Difference between predicted score is lower than pre-defined T2

Method



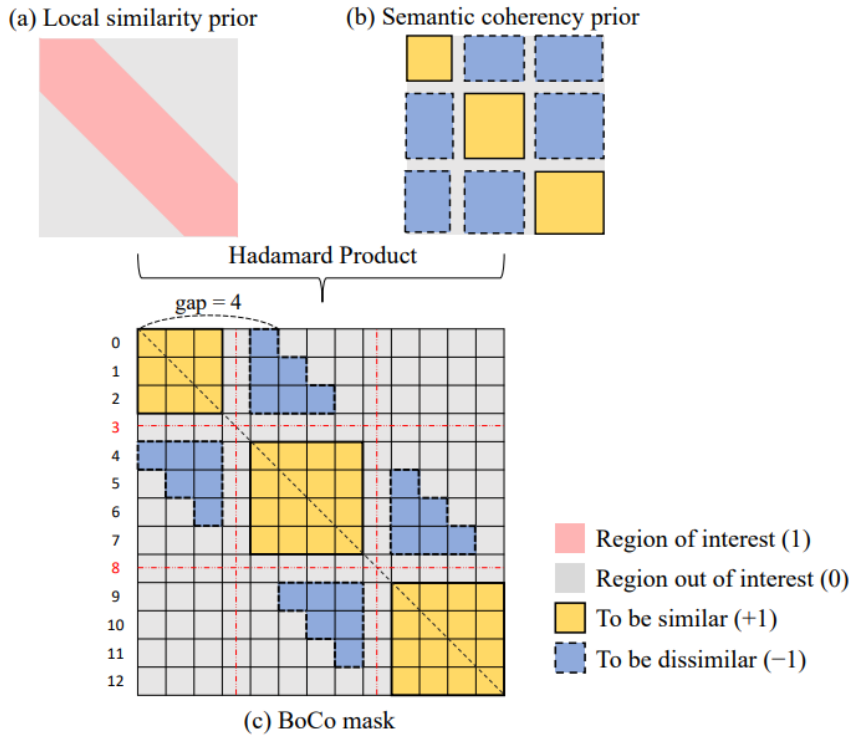
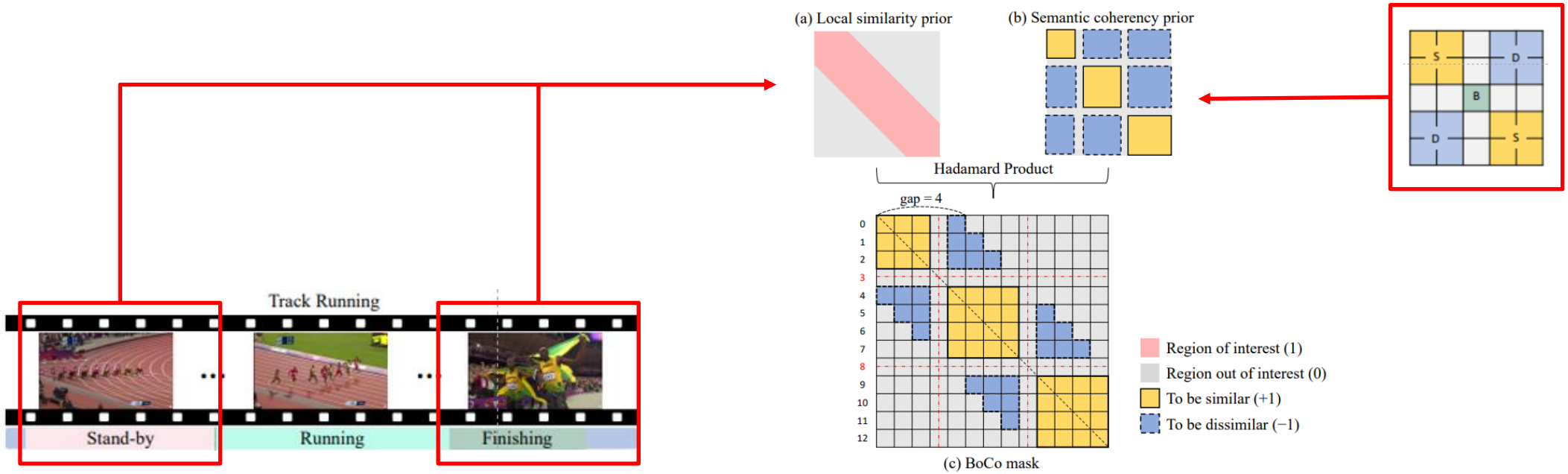


Figure 5. Given the list of boundary indices ([3,8] in the above example), (c) represents the mask to compute BoCo loss. Both prior (a), (b) are merged by element-wise multiplication.

Method



BoCo mask can be defined as a pattern map.

(a) To train boundaries, it is necessary to train frames nearby.

(b) Use semantic coherency priority based on patterns observed

Method

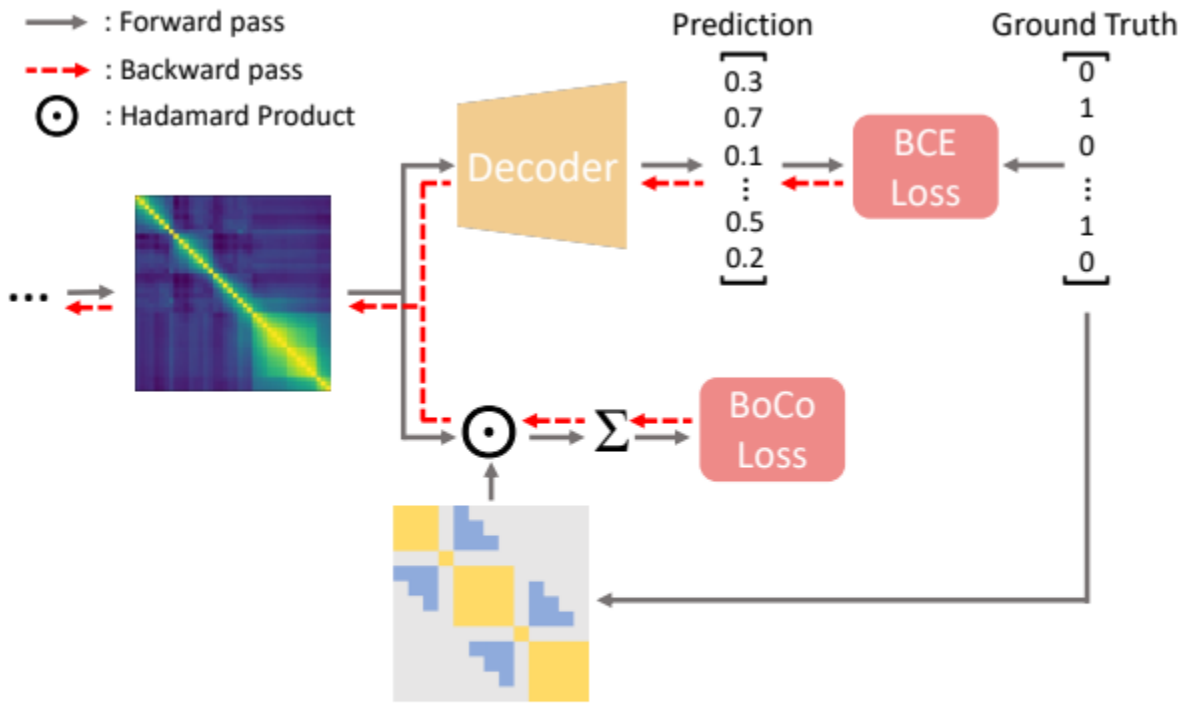


Figure 6. Supervised GEBD framework with a decoder. With supervision, we can replace RTP with a neural decoder and improve the GEBD performance with additional BCE loss.

Experiments

Experiments

	Method	F1@0.05	Average F1
Unsupervised	SceneDetect	27.5	31.9
	PA-Random	33.6	50.6
	PA	39.6	52.7
	UBoCo-Res50 (ours)	**70.3**	**86.7**
	UBoCo-TSN (ours)	**70.2**	**86.7**
Supervised	BMN	18.6	22.3
	BMN-StartEnd	49.1	64.0
	TCN-TAPOS	46.4	62.7
	TCN	58.8	68.5
	PC	62.5	81.7
	Ours	**0.743**	**0.865**
	SBoCo-Res50 (ours)	**73.2**	**86.6**
	SBoCo-TSN (ours)	**78.7**	**89.2**

Table 1. Results on Kinetics-GEBD for unsupervised (top) and supervised (bottom) methods. The scores of previous methods are from [37].

Another supervised-based paper from CVPR 2022.

Experiments

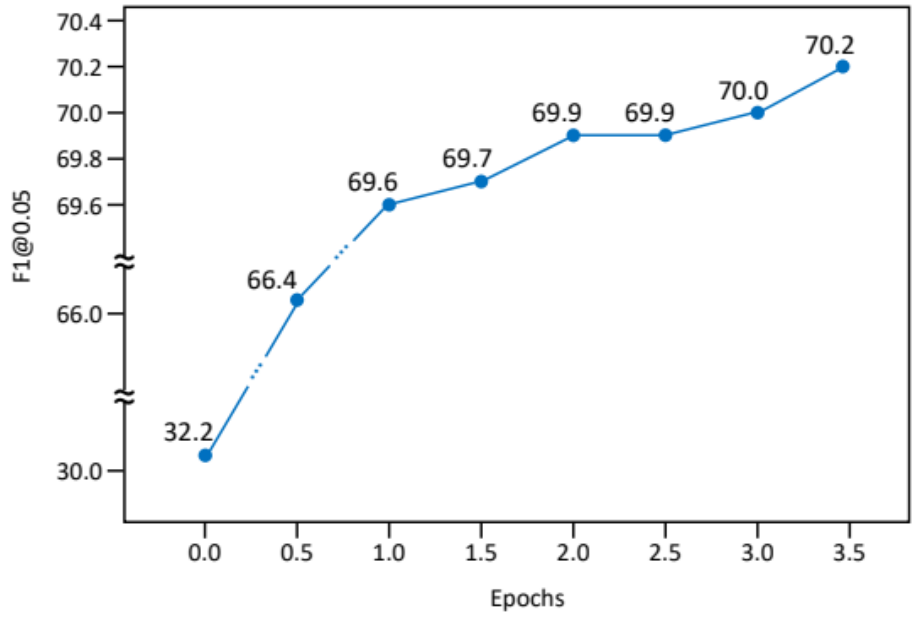
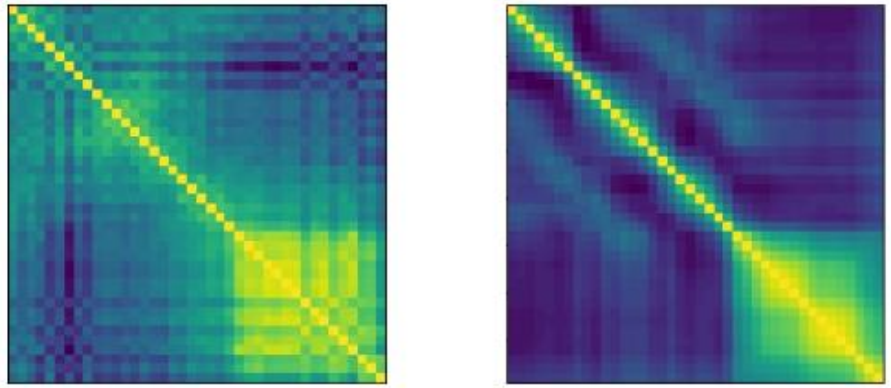


Figure 7. Improvement of UBoCo as the encoder of the model is trained on pseudo-labels in a self-supervised manner.



(a) Trained without BoCo loss (b) Trained with BoCo loss

Figure 9. BoCo loss in supervised model makes the TSM more interpretable and informative. Boundary patterns in (b) is much more distinguishable than those in (a).

Experiments

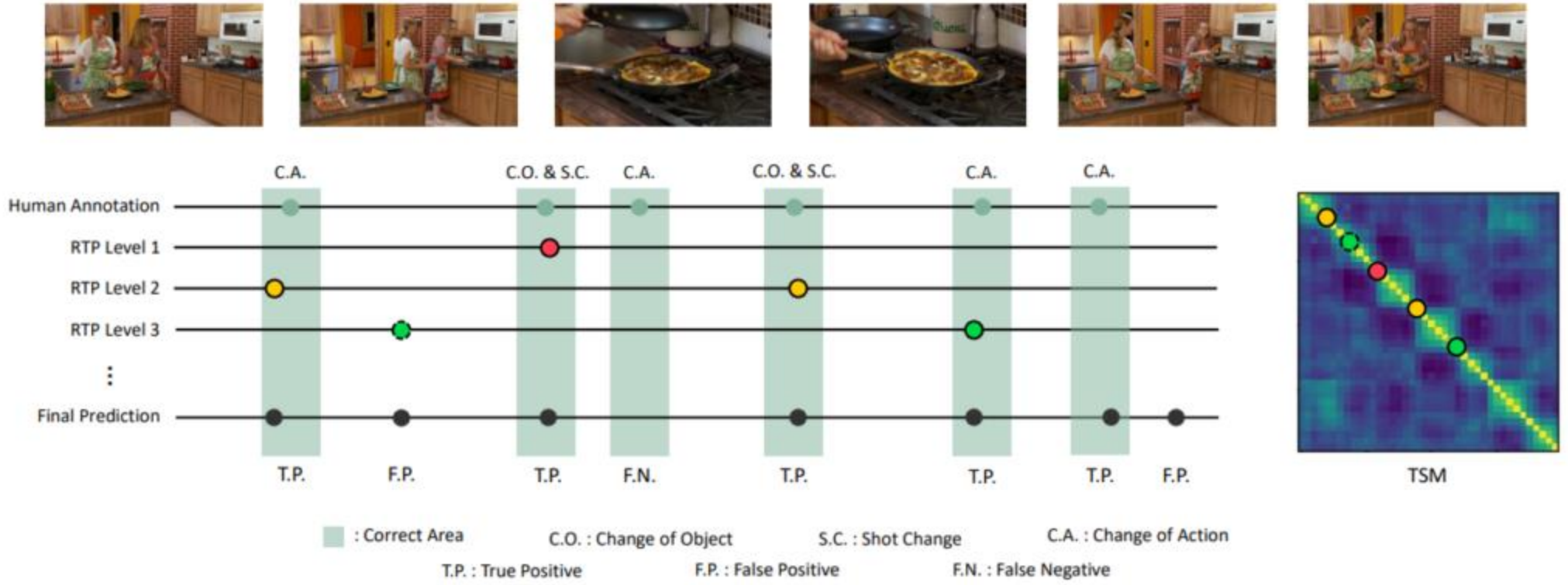


Figure 8. Above figure illustrates how RTP detects event boundaries from the given TSM. As shown in the figure, apparent boundaries including shot change are captured at the early level of RTP, while more subtle boundaries are deferred to the last level.

Q&A

Q&A