

# Bottom-Up Temporal Action Localization with Mutual Regularization

## supplementary materials

Anonymous ECCV submission

Paper ID 622

### A Qualitative Analysis: Classification

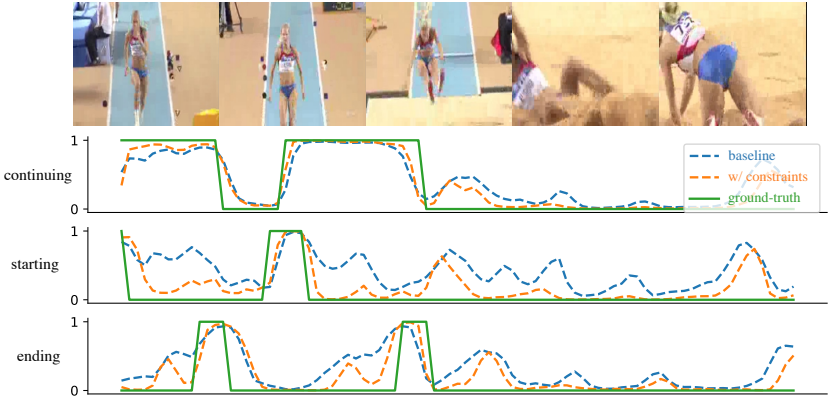
As mentioned in Section 3.2 (motivation), we show the three classification results in terms of **False Alarm** on THUMOS14 dataset. It reveals that we can not obtain a convincing results even in training set because of the ambiguous labels for action starting, continuing, and ending. So we provide more comprehensive results in Table 1 to show effectiveness of our proposed IntraC and InterC. To a certain extent, the introduction of these constraints can alleviate the classification errors caused by semantic ambiguity at a single temporal location. We find both False Alarm (FA) and Missing Alarm (MA) are reduced by IntraC and InterC in training set. That means it helps to find a better local-optimal solution for training these classification tasks. Although the MA (S) and MA (E) are increased (between baseline and IntraC+InterC) in testing stage, the FA (S) and FA (E) decrease, which has a greater impact on the final result.

**Table 1.** Additional quantitative results of the event classification. FA and MA are false alarm and missing alarm of the classification metrics. S, E and C, in parentheses, denote starting, ending, and continuing, respectively.

	FA (S)	MA (S)	FA (E)	MA (E)	FA (C)	MA (C)
Baseline Train	0.68	0.16	0.64	0.15	0.28	0.13
Baseline Test	0.75	0.28	0.69	0.27	0.36	0.20
IntraC+InterC Train	0.50	0.12	0.49	0.12	0.22	0.09
IntraC+InterC Test	0.63	0.36	0.59	0.36	0.30	0.20

## B Quantitative Analysis: Visualization

This supplementary materials also provide more visualization results on two datasets, THUMOS14 and ActivityNet1.3. Shown as follows are the predicted probability events  $\mathbf{p}^C$ ,  $\mathbf{p}^S$ , and  $\mathbf{p}^E$  on test set. Same as the Section 4.4, the green, blue, and orange lines stand for the ground-truth, the event predicted by baseline model, and the event predicted by our IntraC+InterC equipped model, respectively. Overall, adding proposed regularization terms (IntraC+InterC) generate more stable probabilities inside foreground and background regions. Besides, the crucial evidence for performance improvement is that most false positives in starting event  $\mathbf{p}^S$  and ending event  $\mathbf{p}^E$  are suppressed, so that these constraints can reduce many wrong candidate proposals by these false positive starting and ending points. For each video, we also provide the final localization results with the corresponding scores and their IoU ratios with ground-truth.



**Fig. 1.** Example on THUMOS14 dataset: “video\_test\_0000242.mp4”.

**Table 2.** Action localization results of Figure 1. The Green table stands for ground-truth annotation. The Blue table stands for baseline prediction, while the Orange table stands for ours results. The order is ranked by “Score” if it is larger than 0.5. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(0.0, 6.7)	-	-	LongJump

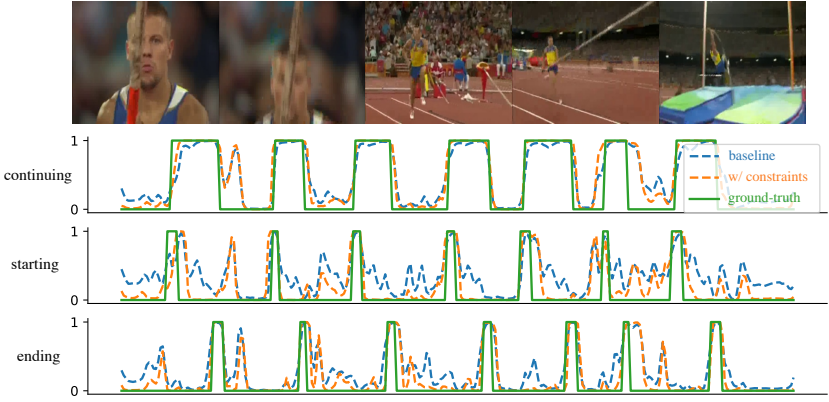
Segment (sec)	Score	IoU	Action
(12.3, 18.8)	0.77	0.72	LongJump
(5.3, 7.5)	0.71	0.19	LongJump
(5.9, 7.7)	0.62	0.10	LongJump
(5.9, 7.1)	0.62	0.11	LongJump
(5.3, 6.5)	0.62	0.19	LongJump
(17.0, 19.9)	0.61	0.32	LongJump
(5.9, 6.5)	0.55	0.09	LongJump
(4.3, 7.1)	0.54	0.33	LongJump

Segment (sec)	Score	IoU	Action
(11.8, 20.7)	-	-	LongJump

Segment (sec)	Score	IoU	Action
(43.3, 45.8)	0.54	0.00	LongJump
(6.4, 7.5)	0.54	0.04	LongJump
(6.4, 7.1)	0.53	0.05	LongJump
(16.6, 18.8)	0.52	0.24	LongJump
(5.3, 6.0)	0.52	0.12	LongJump
(12.3, 16.2)	0.51	0.43	LongJump
(5.9, 8.5)	0.51	0.09	LongJump
(16.0, 19.4)	0.51	0.37	LongJump

Segment (sec)	Score	IoU	Action
(12.2, 20.2)	0.99	0.90	LongJump

Segment (sec)	Score	IoU	Action
(1.1, 7.3)	0.88	0.77	LongJump



**Fig. 2.** Example on THUMOS14 dataset: “video\_test\_0000444.mp4”.

**Table 3.** Action localization results of Figure 2. The **Green** table stands for ground-truth annotation. The **Blue** table stands for baseline prediction, while the **Orange** table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(12.4, 22.9)	-	-	PoleVault
(36.2, 42.7)	-	-	PoleVault
(55.0, 63.0)	-	-	PoleVault
(77.0, 85.7)	-	-	PoleVault

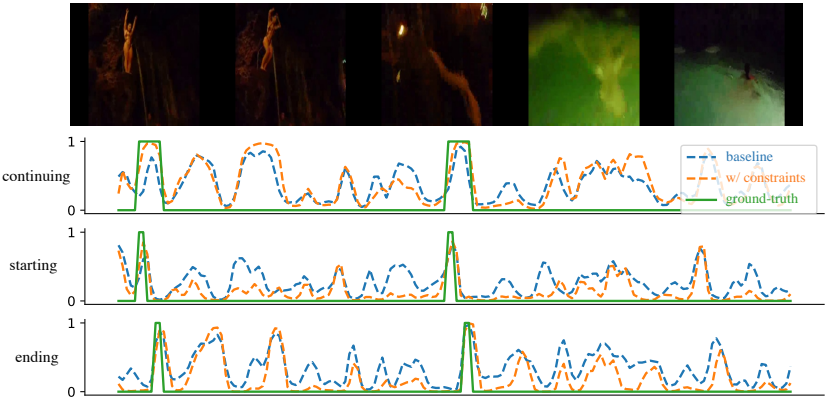
Segment (sec)	Score	IoU	Action
(94.4, 105.3)	-	-	PoleVault
(113.1, 118.0)	-	-	PoleVault
(129.6, 138.6)	-	-	PoleVault

Segment (sec)	Score	IoU	Action
(36.1, 43.6)	0.98	0.86	PoleVault
(55.4, 63.7)	0.97	0.87	PoleVault
(77.9, 86.1)	0.96	0.86	PoleVault
(14.4, 22.9)	0.96	0.81	PoleVault
(130.7, 139.2)	0.95	0.83	PoleVault
(94.7, 105.7)	0.94	0.94	PoleVault
(113.1, 118.4)	0.90	0.92	PoleVault
(26.2, 28.3)	0.77	0.00	PoleVault
(110.7, 117.9)	0.71	0.66	PoleVault

Segment (sec)	Score	IoU	Action
(26.6, 28.0)	0.68	0.00	PoleVault
(133.3, 139.0)	0.65	0.57	PoleVault
(116.4, 118.7)	0.58	0.28	PoleVault
(17.9, 22.9)	0.57	0.48	PoleVault
(25.7, 28.0)	0.54	0.00	PoleVault
(100.2, 105.7)	0.51	0.46	PoleVault
(113.1, 115.8)	0.51	0.54	PoleVault
(119.8, 120.4)	0.50	0.00	PoleVault
(116.4, 117.9)	0.50	0.29	PoleVault

Segment (sec)	Score	IoU	Action
(36.1, 43.3)	1.00	0.90	PoleVault
(14.6, 23.0)	1.00	0.78	PoleVault
(55.4, 63.7)	1.00	0.87	PoleVault
(94.7, 105.5)	1.00	0.95	PoleVault
(130.2, 139.1)	1.00	0.88	PoleVault
(77.9, 85.7)	0.99	0.90	PoleVault
(110.7, 118.5)	0.91	0.62	PoleVault
(26.4, 43.1)	0.83	0.39	PoleVault

Segment (sec)	Score	IoU	Action
(113.1, 119.5)	0.83	0.76	PoleVault
(26.4, 28.3)	0.70	0.00	PoleVault
(112.6, 117.8)	0.59	0.88	PoleVault
(126.9, 138.8)	0.53	0.76	PoleVault
(119.6, 120.2)	0.52	0.00	PoleVault
(119.6, 119.9)	0.52	0.00	PoleVault
(119.6, 120.6)	0.50	0.00	PoleVault



**Fig. 3.** Example on THUMOS14 dataset: “video\_test\_0000785.mp4”.

**Table 4.** Action localization results of Figure 3. The **Green** table stands for ground-truth annotation. The **Blue** table stands for baseline prediction, while the **Orange** table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(3.4, 5.6)	-	-	CliffDiving

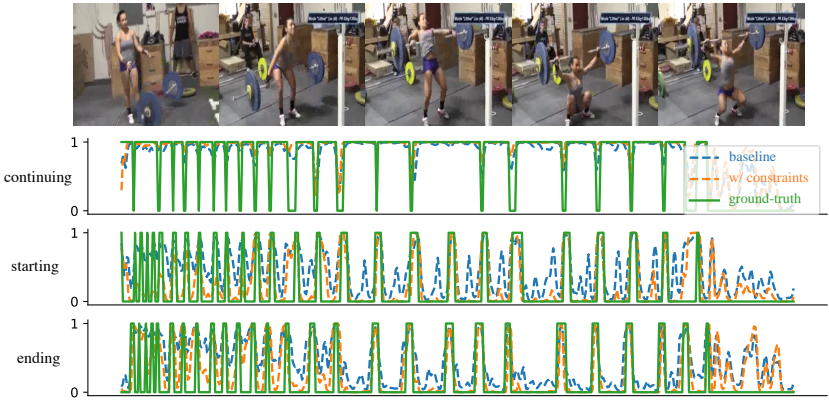
Segment (sec)	Score	IoU	Action
(44.0, 45.7)	0.82	0.66	Diving
(0.5, 5.9)	0.71	0.41	Diving
(75.9, 78.0)	0.60	0.00	Diving
(4.2, 5.9)	0.57	0.55	Diving

Segment (sec)	Score	IoU	Action
(43.7, 46.0)	0.86	0.71	Diving
(3.6, 5.9)	0.77	0.80	Diving
(0.5, 5.9)	0.66	0.41	Diving

Segment (sec)	Score	IoU	Action
(43.3, 45.6)	-	-	CliffDiving

Segment (sec)	Score	IoU	Action
(44.3, 46.2)	0.56	0.46	Diving
(16.9, 20.7)	0.53	0.00	Diving
(37.6, 45.7)	0.51	0.28	Diving

Segment (sec)	Score	IoU	Action
(3.6, 11.3)	0.52	0.26	Diving
(43.7, 52.5)	0.50	0.20	Diving



**Fig. 4.** Example on THUMOS14 dataset: “video\_test\_0001270.mp4”.

**Table 5.** Action localization results of Figure 4. The **Green** table stands for ground-truth annotation. The **Blue** table stands for baseline prediction, while the **Orange** table stands for ours results. The order is ranked by “Score” if it is larger than **0.7**. IoU is reported with ground-truth annotation for each instance.

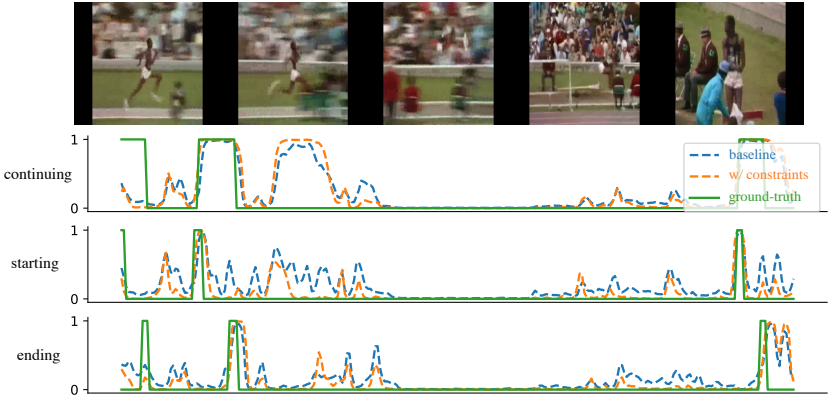
Segment (sec)	Score	IoU	Action	Segment (sec)	Score	IoU	Action
(0.0, 4.8)	-	-	CleanAndJerk	(69.3, 75.5)	-	-	CleanAndJerk
(5.8, 7.5)	-	-	CleanAndJerk	(77.8, 84.7)	-	-	CleanAndJerk
(8.2, 10.1)	-	-	CleanAndJerk	(88.1, 100.3)	-	-	CleanAndJerk
(10.7, 12.4)	-	-	CleanAndJerk	(101.4, 113.7)	-	-	CleanAndJerk
(12.9, 14.6)	-	-	CleanAndJerk	(115.3, 129.5)	-	-	CleanAndJerk
(15.9, 20.4)	-	-	CleanAndJerk	(130.0, 141.1)	-	-	CleanAndJerk
(21.3, 24.6)	-	-	CleanAndJerk	(143.2, 152.4)	-	-	CleanAndJerk
(26.2, 30.2)	-	-	CleanAndJerk	(155.9, 173.2)	-	-	CleanAndJerk
(31.2, 36.3)	-	-	CleanAndJerk	(175.5, 186.8)	-	-	CleanAndJerk
(37.4, 41.1)	-	-	CleanAndJerk	(188.8, 199.8)	-	-	CleanAndJerk
(41.9, 46.4)	-	-	CleanAndJerk	(201.4, 212.6)	-	-	CleanAndJerk
(48.2, 52.2)	-	-	CleanAndJerk	(214.2, 222.1)	-	-	CleanAndJerk
(53.9, 57.9)	-	-	CleanAndJerk	(226.9, 230.5)	-	-	CleanAndJerk
(59.6, 65.5)	-	-	CleanAndJerk				

Segment (sec)	Score	IoU	Action
(175.0, 186.9)	0.95	0.95	CleanAndJerk
(143.0, 153.0)	0.94	0.92	CleanAndJerk
(226.6, 231.1)	0.94	0.81	CleanAndJerk
(42.5, 46.9)	0.90	0.78	CleanAndJerk
(53.5, 58.2)	0.90	0.85	CleanAndJerk
(188.1, 199.9)	0.89	0.93	CleanAndJerk
(201.6, 213.6)	0.88	0.90	CleanAndJerk
(37.9, 41.5)	0.87	0.79	CleanAndJerk
(26.1, 29.2)	0.87	0.75	CleanAndJerk
(130.5, 141.5)	0.87	0.92	CleanAndJerk
(59.3, 64.6)	0.86	0.80	CleanAndJerk
(26.6, 30.0)	0.86	0.86	CleanAndJerk
(88.5, 100.9)	0.85	0.92	CleanAndJerk
(224.3, 231.1)	0.85	0.53	CleanAndJerk
(55.1, 58.2)	0.82	0.66	CleanAndJerk
(21.5, 25.1)	0.82	0.82	CleanAndJerk
(214.1, 220.7)	0.81	0.81	CleanAndJerk
(0.5, 5.8)	0.80	0.74	CleanAndJerk
(116.8, 129.8)	0.80	0.87	CleanAndJerk
(47.9, 51.2)	0.80	0.69	CleanAndJerk
(210.2, 213.1)	0.79	0.20	CleanAndJerk
(222.3, 222.9)	0.79	0.00	CleanAndJerk
(78.2, 84.2)	0.79	0.87	CleanAndJerk
(13.7, 15.2)	0.79	0.42	CleanAndJerk
(6.7, 8.3)	0.78	0.33	CleanAndJerk
(17.6, 20.6)	0.77	0.61	CleanAndJerk
(13.7, 14.7)	0.75	0.53	CleanAndJerk
(222.3, 223.2)	0.75	0.00	CleanAndJerk
(232.6, 234.2)	0.75	0.00	CleanAndJerk

Segment (sec)	Score	IoU	Action
(225.6, 230.9)	0.99	0.68	CleanAndJerk
(201.4, 213.5)	0.99	0.92	CleanAndJerk
(77.9, 85.2)	0.99	0.92	CleanAndJerk
(53.1, 58.6)	0.98	0.73	CleanAndJerk
(187.8, 200.0)	0.98	0.90	CleanAndJerk
(130.2, 141.9)	0.98	0.92	CleanAndJerk
(142.8, 153.4)	0.98	0.87	CleanAndJerk
(59.6, 65.7)	0.97	0.96	CleanAndJerk
(174.7, 187.2)	0.97	0.90	CleanAndJerk
(214.0, 222.8)	0.96	0.90	CleanAndJerk
(86.9, 101.0)	0.96	0.87	CleanAndJerk

Segment (sec)	Score	IoU	Action
(170.4, 173.9)	0.75	0.15	CleanAndJerk
(16.2, 20.6)	0.74	0.90	CleanAndJerk
(8.8, 10.6)	0.74	0.57	CleanAndJerk
(222.9, 223.2)	0.73	0.00	CleanAndJerk
(31.5, 33.9)	0.73	0.47	CleanAndJerk
(26.1, 28.0)	0.73	0.44	CleanAndJerk
(221.8, 222.9)	0.72	0.03	CleanAndJerk
(65.8, 66.2)	0.72	0.00	CleanAndJerk
(7.1, 9.0)	0.72	0.25	CleanAndJerk
(21.3, 22.1)	0.72	0.25	CleanAndJerk
(6.7, 9.6)	0.72	0.40	CleanAndJerk
(7.1, 8.0)	0.72	0.19	CleanAndJerk
(81.2, 84.9)	0.72	0.49	CleanAndJerk
(26.6, 28.8)	0.71	0.54	CleanAndJerk
(49.3, 51.8)	0.71	0.61	CleanAndJerk
(13.7, 16.7)	0.71	0.25	CleanAndJerk
(21.3, 22.7)	0.71	0.42	CleanAndJerk
(6.1, 9.0)	0.71	0.44	CleanAndJerk
(60.7, 64.1)	0.71	0.58	CleanAndJerk
(6.7, 7.7)	0.71	0.44	CleanAndJerk
(6.1, 8.0)	0.71	0.63	CleanAndJerk
(15.8, 16.7)	0.71	0.17	CleanAndJerk
(69.5, 74.2)	0.71	0.76	CleanAndJerk
(149.9, 153.0)	0.71	0.25	CleanAndJerk
(78.2, 81.2)	0.71	0.43	CleanAndJerk
(14.0, 14.7)	0.71	0.35	CleanAndJerk
(60.2, 65.6)	0.71	0.89	CleanAndJerk
(65.9, 66.2)	0.70	0.00	CleanAndJerk

Segment (sec)	Score	IoU	Action
(115.8, 130.2)	0.94	0.92	CleanAndJerk
(68.7, 75.9)	0.91	0.86	CleanAndJerk
(155.1, 174.1)	0.85	0.91	CleanAndJerk
(41.9, 47.5)	0.85	0.81	CleanAndJerk
(59.5, 59.4)	0.82	0.00	CleanAndJerk
(232.4, 234.5)	0.81	0.00	CleanAndJerk
(30.9, 37.0)	0.78	0.84	CleanAndJerk
(0.6, 6.3)	0.75	0.66	CleanAndJerk
(25.9, 30.5)	0.75	0.86	CleanAndJerk
(225.9, 234.3)	0.73	0.43	CleanAndJerk
(21.1, 25.5)	0.71	0.76	CleanAndJerk



**Fig. 5.** Example on THUMOS14 dataset: “video\_test\_0001219.mp4”.

**Table 6.** Action localization results of Figure 5. The **Green** table stands for ground-truth annotation. The **Blue** table stands for baseline prediction, while the **Orange** table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(0.4, 5.9)	-	-	LongJump
(18.1, 26.1)	-	-	LongJump

Segment (sec)	Score	IoU	Action
(141.0, 148.0)	0.94	0.74	LongJump
(18.9, 26.7)	0.93	0.84	LongJump
(146.3, 148.0)	0.60	0.00	LongJump
(146.3, 147.4)	0.55	0.00	LongJump

Segment (sec)	Score	IoU	Action
(140.9, 147.9)	0.98	0.76	LongJump
(18.8, 27.0)	0.98	0.82	LongJump

Segment (sec)	Score	IoU	Action
(140.8, 146.3)	-	-	LongJump

Segment (sec)	Score	IoU	Action
(149.3, 151.1)	0.55	0.00	LongJump
(25.3, 26.7)	0.53	0.10	LongJump
(25.3, 26.1)	0.50	0.10	LongJump

Segment (sec)	Score	IoU	Action
(10.7, 26.1)	0.67	0.52	LongJump



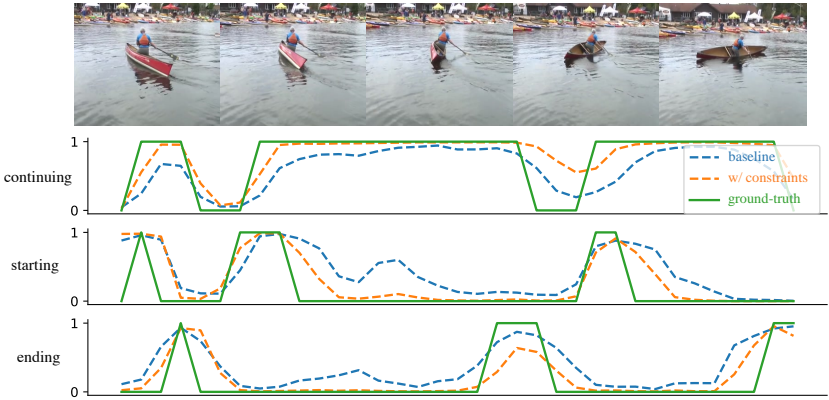


Fig. 6. Example on ActivityNet1.3 dataset: “0iIY3HLF3lU.mp4”.

Table 7. Action localization results of Figure 6. The Green table stands for ground-truth annotation. The Blue table stands for baseline prediction, while the Orange table stands for ours results. The order is ranked by “Score” if it is larger than 0.5. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(5.3, 10.4)	-	-	Canoeing
(20.3, 54.2)	-	-	Canoeing

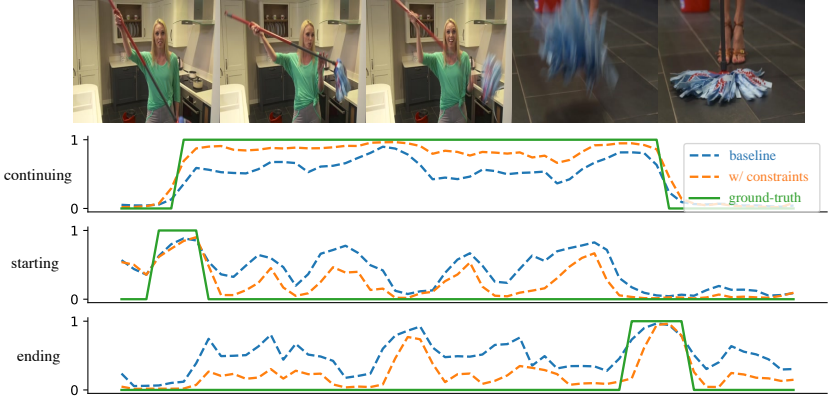
Segment (sec)	Score	IoU	Action
(22.2, 88.1)	0.93	0.47	Canoeing
(5.3, 9.8)	0.89	0.88	Canoeing
(22.2, 54.3)	0.85	0.94	Canoeing
(66.8, 88.1)	0.84	0.90	Canoeing

Segment (sec)	Score	IoU	Action
(21.3, 87.7)	0.95	0.49	Canoeing
(4.3, 9.9)	0.91	0.76	Canoeing
(65.5, 87.7)	0.87	0.94	Canoeing

Segment (sec)	Score	IoU	Action
(64.6, 88.2)	-	-	Canoeing

Segment (sec)	Score	IoU	Action
(5.3, 54.3)	0.84	0.69	Canoeing
(4.4, 8.8)	0.58	0.59	Canoeing
(37.0, 54.3)	0.52	0.50	Canoeing

Segment (sec)	Score	IoU	Action
(21.3, 54.9)	0.63	0.95	Canoeing
(4.3, 54.9)	0.63	0.67	Canoeing



**Fig. 7.** Example on ActivityNet1.3 dataset: “1U8y7e22SQg.mp4”.

**Table 8.** Action localization results of Figure 7. The Green table stands for ground-truth annotation. The Blue table stands for baseline prediction, while the Orange table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

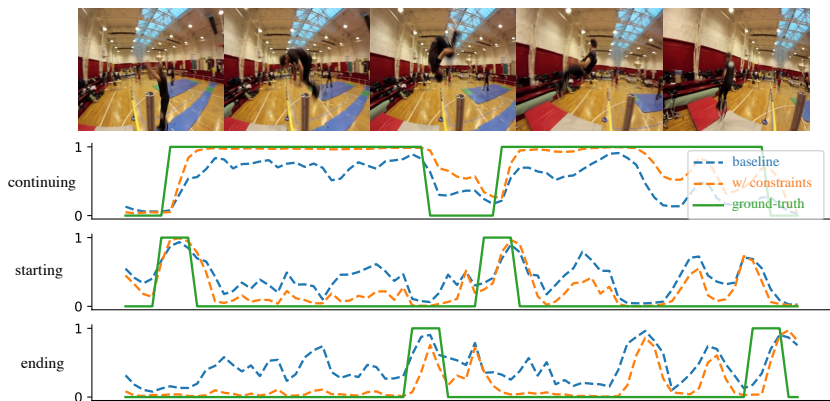
Segment (sec) Score IoU Action			
(14.2, 113.5)	-	-	Mooping floor

Segment (sec)	Score	IoU	Action
(15.9, 112.6)	0.86	0.97	Mooping floor
(15.9, 63.2)	0.82	0.48	Mooping floor
(99.0, 112.6)	0.80	0.14	Mooping floor
(48.3, 112.6)	0.75	0.65	Mooping floor
(48.3, 63.2)	0.72	0.15	Mooping floor
(15.9, 33.6)	0.71	0.18	Mooping floor
(93.0, 112.6)	0.67	0.20	Mooping floor
(15.9, 21.2)	0.66	0.05	Mooping floor
(17.2, 84.3)	0.65	0.68	Mooping floor

Segment (sec)	Score	IoU	Action
(73.6, 112.6)	0.65	0.39	Mooping floor
(31.1, 63.2)	0.59	0.32	Mooping floor
(48.3, 84.3)	0.59	0.36	Mooping floor
(97.7, 109.4)	0.57	0.12	Mooping floor
(50.7, 60.4)	0.53	0.10	Mooping floor
(99.0, 127.3)	0.53	0.13	Mooping floor
(44.4, 60.4)	0.52	0.16	Mooping floor
(31.1, 33.6)	0.52	0.03	Mooping floor
(73.6, 84.3)	0.51	0.11	Mooping floor

Segment (sec)	Score	IoU	Action
(16.2, 114.1)	0.87	0.97	Mooping floor
(16.2, 62.4)	0.69	0.47	Mooping floor

Segment (sec)	Score	IoU	Action
(98.7, 114.1)	0.64	0.15	Mooping floor
(73.1, 114.1)	0.51	0.40	Mooping floor



**Fig. 8.** Example on ActivityNet1.3 dataset: “2AQg1DDVYHI.mp4”.

**Table 9.** Action localization results of Figure 8. The Green table stands for ground-truth annotation. The Blue table stands for baseline prediction, while the Orange table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(16.2, 87.5)	-	-	Powerbocking

Segment (sec)	Score	IoU	Action
(113.0, 150.4)	0.86	0.50	Powerbocking
(18.5, 89.0)	0.85	0.95	Powerbocking
(113.0, 189.6)	0.81	0.91	Powerbocking
(132.8, 150.4)	0.77	0.23	Powerbocking
(18.5, 59.3)	0.69	0.57	Powerbocking
(166.0, 189.6)	0.66	0.24	Powerbocking
(180.4, 189.6)	0.64	0.06	Powerbocking
(182.7, 191.2)	0.60	0.03	Powerbocking
(74.5, 150.4)	0.59	0.37	Powerbocking

Segment (sec)	Score	IoU	Action
(109.8, 185.3)	-	-	Powerbocking

Segment (sec)	Score	IoU	Action
(132.8, 168.7)	0.59	0.47	Powerbocking
(135.1, 191.2)	0.58	0.62	Powerbocking
(74.5, 89.0)	0.56	0.18	Powerbocking
(18.5, 31.2)	0.54	0.18	Powerbocking
(166.0, 168.7)	0.54	0.03	Powerbocking
(113.0, 118.4)	0.51	0.07	Powerbocking
(18.5, 46.1)	0.51	0.39	Powerbocking
(166.0, 170.2)	0.51	0.05	Powerbocking
(184.4, 189.6)	0.50	0.01	Powerbocking

Segment (sec)	Score	IoU	Action
(17.9, 191.2)	0.96	0.44	Powerbocking
(112.4, 191.2)	0.94	0.89	Powerbocking
(112.4, 151.1)	0.83	0.51	Powerbocking

Segment (sec)	Score	IoU	Action
(17.9, 90.0)	0.75	0.94	Powerbocking
(179.5, 191.2)	0.71	0.07	Powerbocking
(165.6, 191.2)	0.54	0.24	Powerbocking

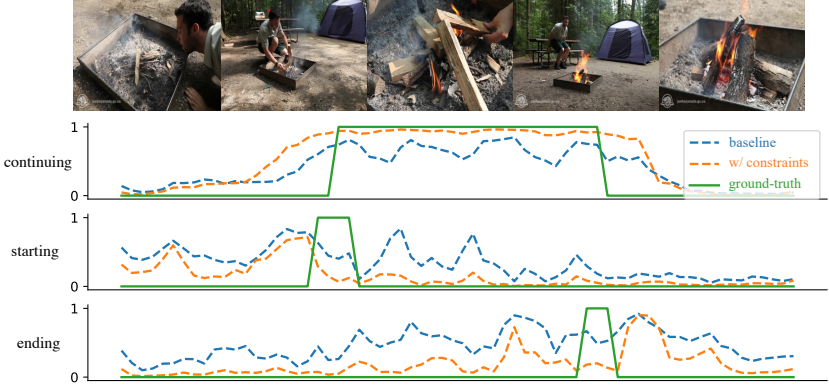


Fig. 9. Example on ActivityNet1.3 dataset: “E\_9te0nq3A.mp4”.

Table 10. Action localization results of Figure 9. The Green table stands for ground-truth annotation. The Blue table stands for baseline prediction, while the Orange table stands for ours results. The order is ranked by “Score” if it is larger than 0.5. IoU is reported with ground-truth annotation for each instance.

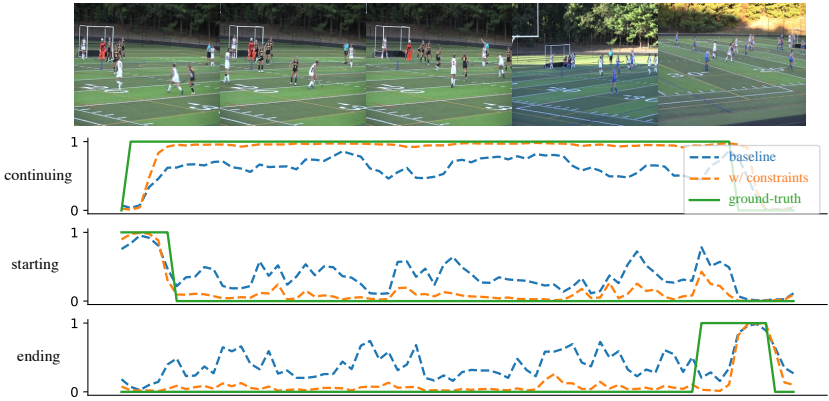
Segment (sec) Score IoU				Action
(55.4, 120.6)				Starting a campfire

Segment (sec)	Score	IoU	Action
(72.2, 130.7)	0.78	0.64	Starting a campfire
(72.2, 101.3)	0.76	0.45	Starting a campfire
(43.8, 101.3)	0.75	0.60	Starting a campfire
(89.5, 101.3)	0.69	0.18	Starting a campfire
(72.2, 76.1)	0.68	0.06	Starting a campfire
(43.8, 76.1)	0.67	0.27	Starting a campfire
(89.5, 129.2)	0.65	0.42	Starting a campfire

Segment (sec)	Score	IoU	Action
(15.4, 101.3)	0.60	0.44	Starting a campfire
(43.8, 62.4)	0.58	0.09	Starting a campfire
(70.4, 76.1)	0.56	0.09	Starting a campfire
(43.8, 127.4)	0.56	0.78	Starting a campfire
(89.5, 107.1)	0.54	0.27	Starting a campfire
(72.2, 77.9)	0.54	0.09	Starting a campfire
(72.2, 81.8)	0.51	0.15	Starting a campfire

Segment (sec)	Score	IoU	Action
(48.0, 131.3)	0.65	0.78	Starting a campfire

Segment (sec)	Score	IoU	Action
(48.0, 100.6)	0.52	0.62	Starting a campfire



**Fig. 10.** Example on ActivityNet1.3 dataset: “\_wITx73-BXw.mp4”.

**Table 11.** Action localization results of Figure 10. The **Green** table stands for ground-truth annotation. The **Blue** table stands for baseline prediction, while the **Orange** table stands for ours results. The order is ranked by “Score” if it is larger than **0.5**. IoU is reported with ground-truth annotation for each instance.

Segment (sec)	Score	IoU	Action
(6.0, 172.3)	-	-	Playing field hockey

Segment (sec)	Score	IoU	Action
(163.4, 177.8)	0.77	0.05	Playing lacrosse
(145.8, 177.8)	0.71	0.15	Playing lacrosse
(9.5, 71.9)	0.71	0.38	Playing lacrosse
(9.5, 136.4)	0.69	0.76	Playing lacrosse

Segment (sec)	Score	IoU	Action
(93.6, 177.8)	0.63	0.46	Playing lacrosse
(9.5, 37.1)	0.63	0.17	Playing lacrosse
(168.4, 177.8)	0.56	0.02	Playing lacrosse

Segment (sec)	Score	IoU	Action
(9.3.0, 177.9)	0.83	0.95	Playing lacrosse