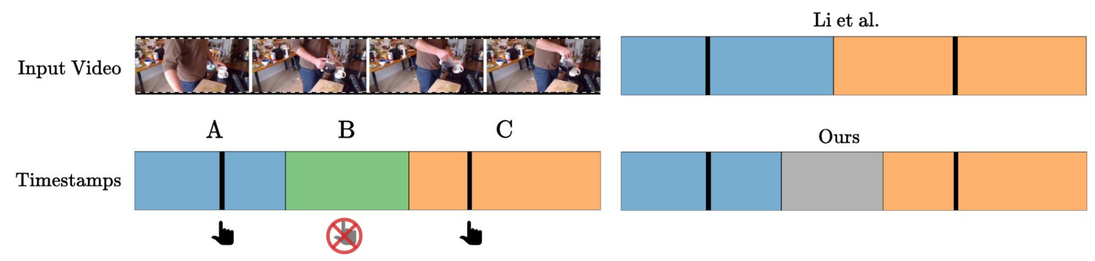
Robust Action Segmentation from Timestamp Supervision

Yaser Souri[1,4,*], Yazan Abu Farha[2,*], Emad Bahrami[1,*], Gianpiero Francesca[3], Jürgen Gall[1]
[1]University of Bonn, [2]Birzeit University, [3]Toyota Motor Europe, [4] Microsoft Research
*Contributed equally to this work

Introduction

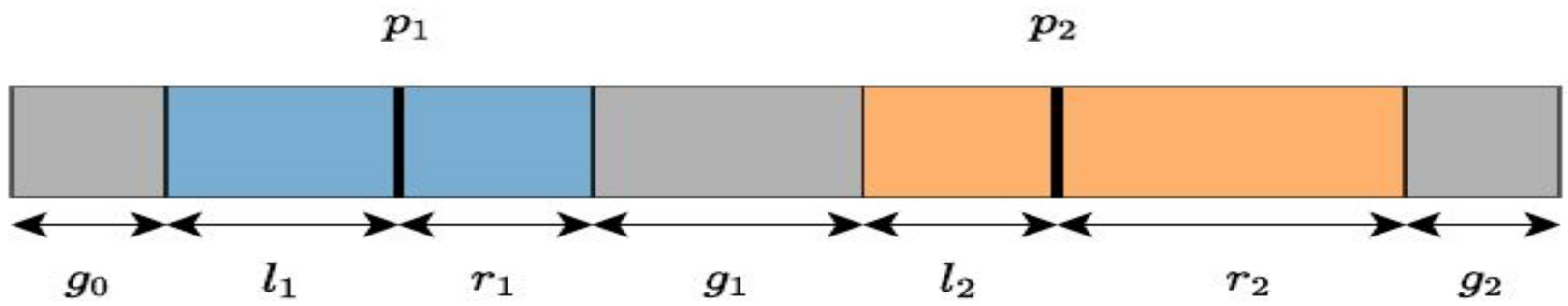
- Timestamps are currently the best weak supervision for action segmentation
- In timestamp supervision, only a single frame per action segments is annotated, annotator however can miss an action
- **Objective:** Design an approach to which is robust to missing timestamps



Dealing with Missing Timestamps

We aim to identify the start and the end of the actions that corresponds to the timestamps

- Divide the frames between two consecutive timestamps $p_1$and p_2 into three parts r_1, g_1, l_2



- The problem is formulated as constrained optimization to jointly minimize r_1, g_1, l_2 and solved using gradient descent

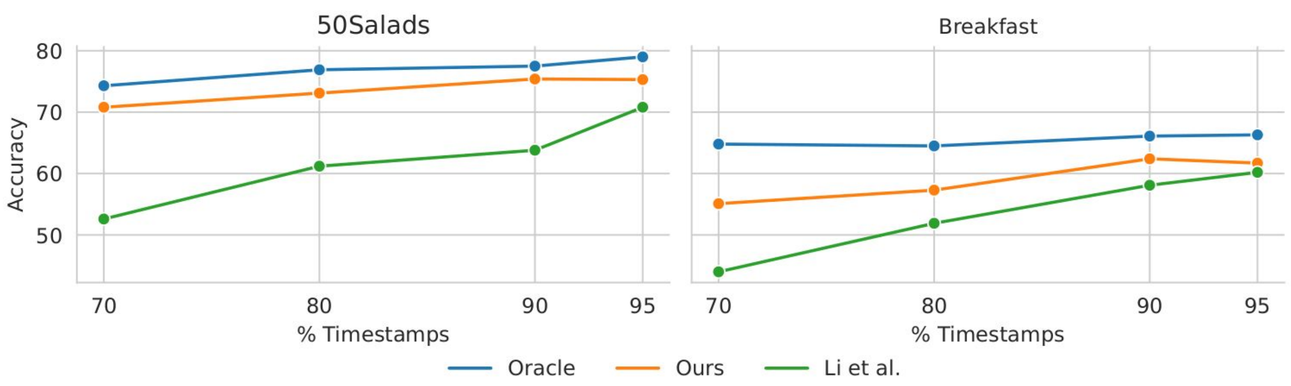
$$\{r_i^*, g_i^*, l_i^*\} = \underset{r_i, g_i, l_i}{\operatorname{argmin}} \sum_{i=1}^{N} \left(\sum_{t=p_i-l_i}^{p_i+r_i} -\log \tilde{y}_t[y_{p_i}] \right) + \beta \sum_{i=0}^{N} g_i$$

$\hat{y}_t[y_{p_i}]$: predicted probabilities for timestamp p_i

$$\text{s.t. } p_{i+1} - p_i = r_i + g_i + l_{i+1}$$
$$r_i \geq 0, \;\; g_i \geq 0, \;\; l_i \geq 0.$$

Quantitative Results

- 100% timestamps corresponds to having one timestamp per action segment
- The 'Oracle' is an upper bound of our approach, and corresponds to training in a fully supervised setting



% Timestamps	Method	Breakfast					50Salads				
		F1@{10, 25, 50}			Edit	Acc	F1@{10, 25, 50}			Edit	Acc
95%	Li *et al.*	67.3	59.7	42.7	68.2	60.2	70.9	67.4	53.4	63.8	70.8
	Ours	**70.2**	**62.4**	**44.8**	**69.7**	**61.7**	**72.9**	**69.6**	**57.5**	**64.2**	**75.3**
	Oracle	71.0	65.3	51.4	70.2	66.3	74.8	72.0	64.1	67.9	79.0
90%	Li *et al.*	65.0	56.5	39.7	66.8	58.1	63.9	59.6	44.3	57.6	63.8
	Ours	**69.8**	**62.2**	**44.5**	**69.7**	**62.4**	**70.0**	**65.1**	**55.2**	**62.1**	**75.4**
	Oracle	69.0	63.0	49.1	68.6	66.1	73.9	71.6	62.5	66.9	77.5
80%	Li *et al.*	59.8	50.4	33.0	62.8	51.9	62.7	56.9	40.3	54.2	61.2
	Ours	**67.3**	**58.9**	**40.8**	**68.5**	**57.3**	**70.9**	**67.8**	**53.7**	**61.4**	**73.1**
	Oracle	69.2	62.8	48.7	68.4	64.5	73.3	70.2	61.0	65.8	76.9
70%	Li *et al.*	53.9	43.5	26.8	59.2	44.0	50.2	44.0	29.5	44.7	52.6
	Ours	**65.0**	**55.0**	**35.7**	**66.5**	**55.1**	**64.1**	**59.2**	**44.8**	**56.9**	**70.8**
	Oracle	69.5	63.2	49.0	69.0	64.8	67.4	63.3	53.2	59.0	74.3

Qualitative Results