



UvA-DARE (Digital Academic Repository)

Online Action Detection

De Geest, R.; Gavves, E.; Ghodrati, A.; Li, Z.; Snoek, C.; Tuytelaars, T.

DOI

[10.1007/978-3-319-46454-1_17](https://doi.org/10.1007/978-3-319-46454-1_17)

Publication date

2016

Document Version

Other version

Published in

Computer Vision – ECCV 2016

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care>)

[Link to publication](#)

Citation for published version (APA):

De Geest, R., Gavves, E., Ghodrati, A., Li, Z., Snoek, C., & Tuytelaars, T. (2016). Online Action Detection. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016 : proceedings* (Vol. 5, pp. 269-284). (Lecture Notes in Computer Science; Vol. 9909). Springer. https://doi.org/10.1007/978-3-319-46454-1_17

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Online Action Detection - Supplementary Material

Roeland De Geest¹, Efstratios Gavves², Amir Ghodrati¹,
Zhenyang Li², Cees Snoek², Tinne Tuytelaars¹

¹KU Leuven - ESAT - PSI

{roeland.degeest,amir.ghodrati,tinne.tuytelaars}@esat.kuleuven.be

²University of Amsterdam - QUVA

{e.gavves,z.li2,c.g.m.snoek}@uva.nl

1 Dataset

The TVSeries dataset we collected consists of 30 action classes. The number of action instances of these classes, and their division over training, validation and test set, can be found in Figure 1. The dataset has videos from six TV series. The relative number of action instances from every series is shown in Figure 2. Some actions, in particular ‘smoke’, ‘fire weapon’ and ‘punch’, do not occur in every series. The distribution of the duration of the action instances can be found in Figure 3. The longer actions are ‘drive car’ and ‘kiss’; ‘punch’ and ‘answer phone’, on the other hand, are typically short. The attached file ‘TVSeries_dataset_example.avi’ contains some small parts of annotated episodes.

2 Further statistics

2.1 Offline action detection

The offline action detection results for all classes and six overlap ratios can be found in Figure 4 for FV, in Figure 5 for CNN and in Figure 6 for LSTM.

2.2 Online action detection

The calibrated average precision as a function of the duration of the action instances is shown in Table 1. We divided the action instances over four quartiles. Similarly, we calculated the calibrated average precision as a function of the motion (number of trajectories); these results can be found in Table 2.

This work was supported by the KU Leuven GOA project *CAMETRON*.

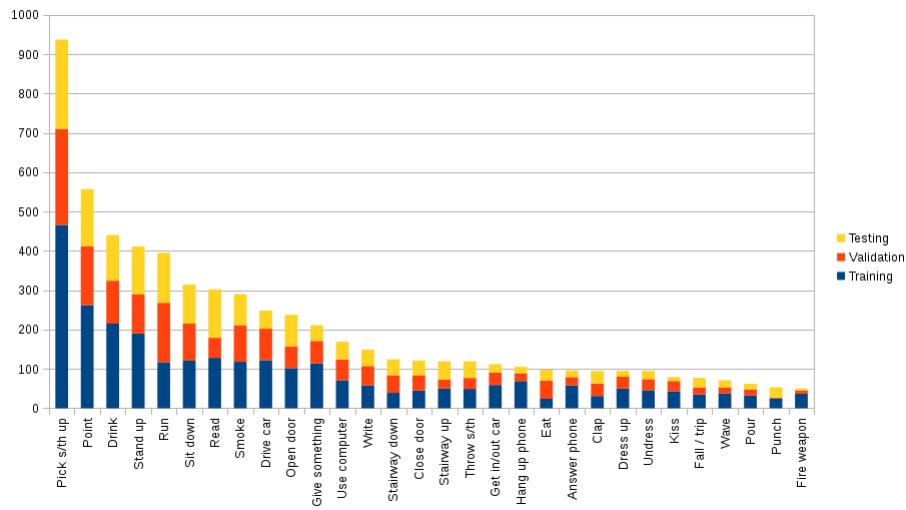


Fig. 1. The number of action instances in the TVSeries dataset and their division over training, validation and test set.

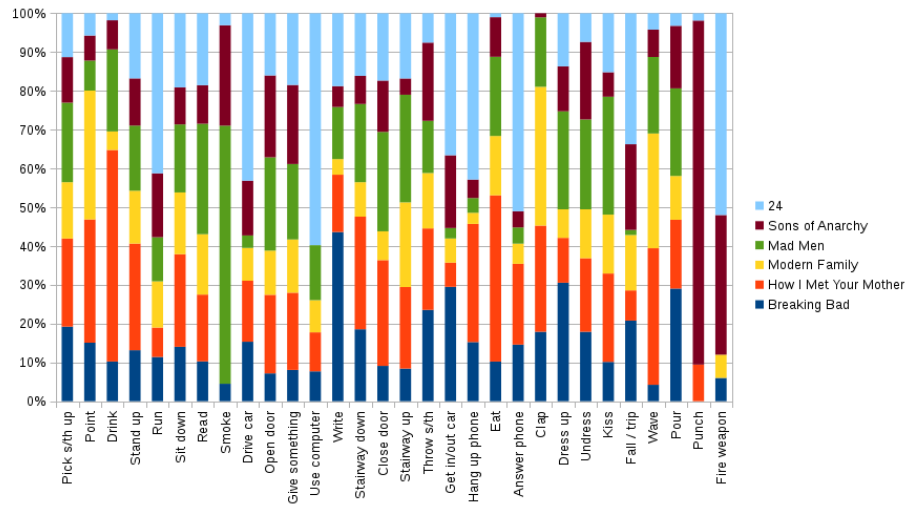


Fig. 2. Distribution of the series of the action instances of the TVSeries dataset.

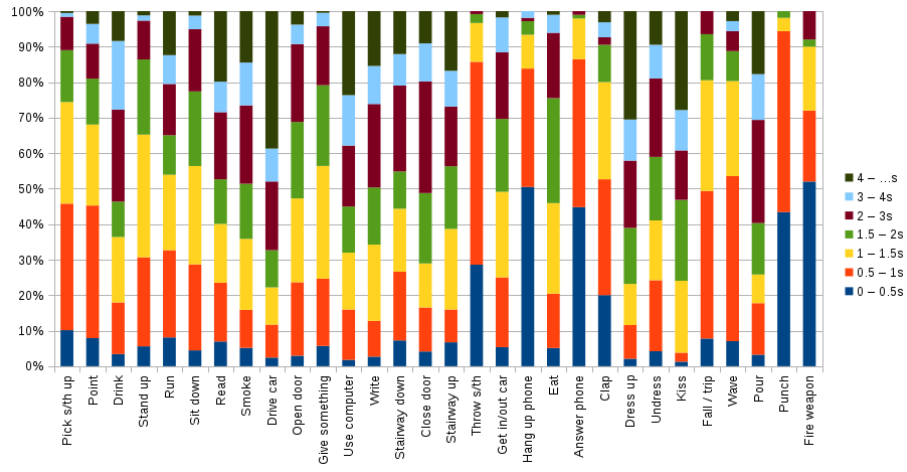


Fig. 3. Distribution of the duration of the action instances of the TVSeries dataset.

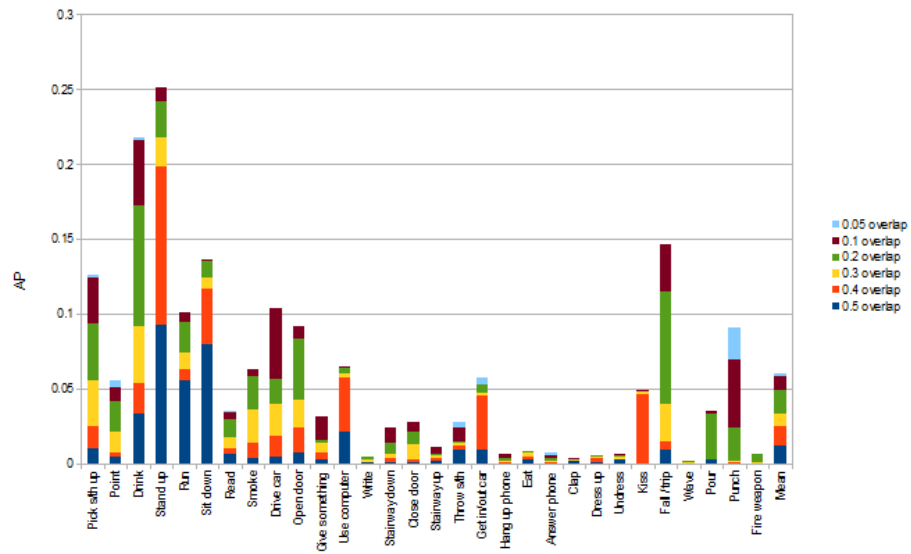


Fig. 4. Average precision for offline detection with FVs for multiple overlap ratios.

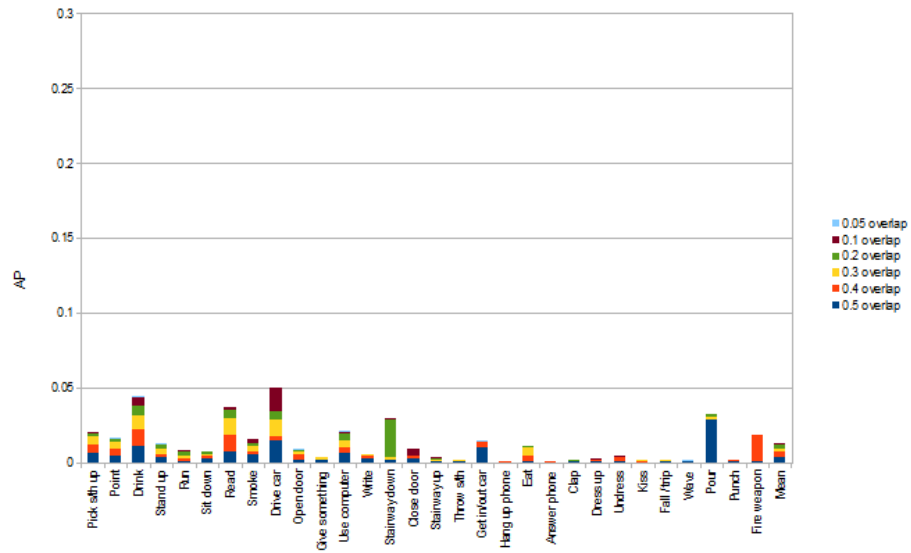


Fig. 5. Average precision for offline detection with CNN for multiple overlap ratios.

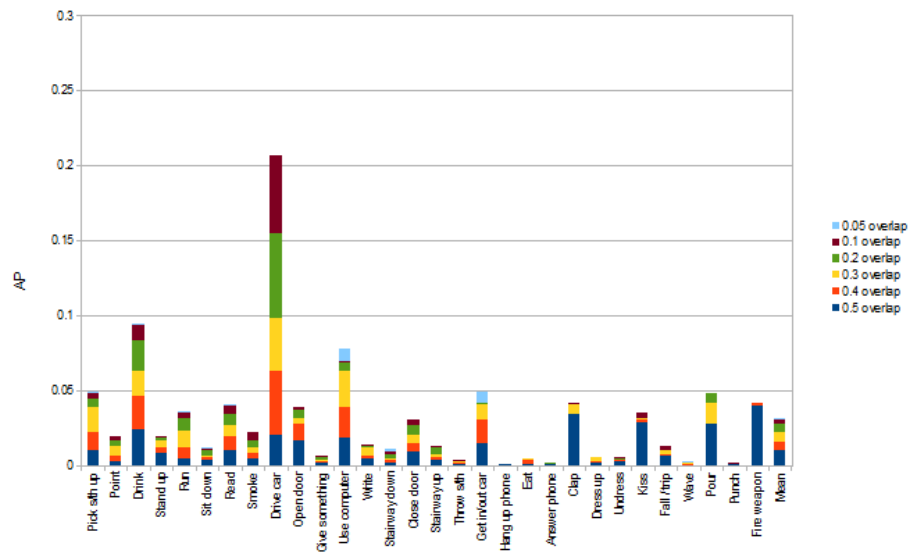


Fig. 6. Average precision for offline detection with LSTM for multiple overlap ratios.

cAP (%)	FV				CNN				LSTM			
	1st quartile	2nd quartile	3rd quartile	4th quartile	1st quartile	2nd quartile	3rd quartile	4th quartile	1st quartile	2nd quartile	3rd quartile	4th quartile
<i>Pick s/th up</i>	70.3	70.8	72.6	62.5	58.7	55.1	55.9	56.1	70.8	71.9	66.4	63.2
<i>Point</i>	66.9	65.2	72.9	66.4	64.7	61.2	60.5	46.9	58.1	60.3	56.9	48.3
<i>Drink</i>	83.5	82.2	83.1	90.0	74.1	71.4	66.8	75.2	82.7	77.1	76.5	80.8
<i>Stand up</i>	77.2	82.1	80.8	83.3	53.9	52.9	44.7	63.8	65.6	57.8	53.3	64.4
<i>Run</i>	80.4	82.1	91.3	90.0	49.7	48.9	58.2	62.9	66.1	70.0	80.4	72.6
<i>Sit down</i>	74.2	70.5	81.0	75.4	45.6	38.1	47.4	48.2	44.7	46.3	48.0	52.9
<i>Read</i>	66.6	68.5	62.9	64.7	69.6	68.5	73.2	55.6	62.0	81.5	71.4	55.8
<i>Smoke</i>	64.9	73.7	76.9	80.8	67.9	56.6	61.3	68.9	44.4	59.8	55.7	71.2
<i>Drive car</i>	95.2	88.7	86.4	91.6	73.8	81.9	84.2	80.2	98.6	81.4	83.0	90.7
<i>Open door</i>	79.5	80.4	76.9	83.2	58.0	52.5	52.3	49.8	79.7	79.2	73.2	68.7
<i>Give s/th</i>	72.4	65.2	78.0	76.6	51.9	52.9	50.2	56.4	71.8	57.3	51.3	43.8
<i>Use computer</i>	57.1	65.3	72.7	71.0	74.4	81.4	80.4	76.6	53.2	65.8	74.8	88.6
<i>Write</i>	53.6	53.1	56.1	43.7	45.8	60.6	55.0	42.4	71.1	62.5	79.6	54.5
<i>Stairway down</i>	73.1	56.0	74.9	87.3	49.5	56.7	53.4	68.8	64.1	69.7	51.9	71.7
<i>Close door</i>	64.0	82.0	82.8	82.0	56.2	45.2	51.3	56.9	72.6	80.8	73.5	72.6
<i>Stairway up</i>	62.6	61.7	55.5	72.2	49.6	50.4	54.7	54.9	55.3	57.6	63.8	76.1
<i>Throw s/th</i>	69.7	69.7	56.1		46.1	51.3	74.4		53.8	54.4	67.0	
<i>Get in/out car</i>	71.2	95.0	87.0	91.4	55.6	84.7	78.9	95.6	62.1	76.7	78.3	79.9
<i>Hang up phone</i>	62.0	49.0	67.3	41.8	54.3	58.7	45.9	42.2	53.6	36.4	47.4	49.9
<i>Eat</i>	65.2	67.8	55.2	69.6	54.6	88.5	60.9	56.1	42.5	49.0	52.9	53.4
<i>Answer phone</i>	50.6	79.9	64.0		56.5	32.0	33.3		51.6	40.7	39.1	
<i>Clap</i>	66.8	75.4	66.9	64.6	55.2	60.2	58.3	46.7	62.1	86.3	78.2	55.8
<i>Dress up</i>		51.8	66.6	64.8		40.9	47.7	44.9		95.2	53.6	55.2
<i>Undress</i>	57.9	67.5	60.1	73.5	66.6	71.5	61.1	68.3	59.3	64.1	45.3	56.6
<i>Kiss</i>		34.8	65.7	72.8		63.3	52.2	51.3		45.2	67.6	64.7
<i>Fall / trip</i>	91.6	96.3	83.6	68.2	50.7	56.5	45.8	45.6	79.6	68.9	59.5	68.2
<i>Wave</i>	58.2	63.4	52.7		61.9	70.8	53.3		47.2	50.6	64.5	
<i>Pour</i>	80.3	56.5	40.4	76.7	63.7	59.2	46.0	79.4	97.0	83.5	63.1	65.4
<i>Punch</i>	93.7				68.5				46.5			
<i>Fire weapon</i>	93.3	84.2	75.7		97.6	91.7	68.3		83.7	51.2	36.5	
Mean	71.5	70.3	70.6	73.8	59.8	60.8	57.8	59.8	64.3	64.9	62.5	65.0

Table 1. Calibrated average precision for online action detection. Analysis in function of the duration of the action instances.

cAP (%)	FV				CNN				LSTM			
	1st quartile	2nd quartile	3rd quartile	4th quartile	1st quartile	2nd quartile	3rd quartile	4th quartile	1st quartile	2nd quartile	3rd quartile	4th quartile
<i>Pick s/th up</i>	67.5	70.5	72.1	68.7	54.6	57.1	58.3	53.7	67.5	68.1	67.3	69.1
<i>Point</i>	66.5	67.2	62.4	70.5	48.1	59.6	62.8	49.1	44.8	53.9	58.3	58.6
<i>Drink</i>	90.4	90.3	83.5	76.5	77.8	75.1	67.4	62.1	79.8	82.3	73.2	82.0
<i>Stand up</i>	70.2	76.3	86.7	84.3	46.5	53.4	51.7	56.2	45.7	56.3	64.8	60.9
<i>Run</i>	85.3	87.0	89.0	88.7	66.4	52.5	57.8	60.1	79.6	67.4	64.3	79.2
<i>Sit down</i>	69.6	76.8	81.1	77.7	51.5	46.1	45.3	43.1	42.6	44.0	55.8	49.7
<i>Read</i>	62.6	61.1	67.5	70.7	53.9	51.7	71.5	71.9	57.9	58.3	69.8	64.6
<i>Smoke</i>	74.3	81.9	82.8	79.2	66.0	70.4	64.0	64.8	63.3	69.5	70.8	64.5
<i>Drive car</i>	83.6	87.4	95.5	90.8	94.9	87.6	74.5	69.1	93.8	91.3	89.2	89.1
<i>Open door</i>	68.0	82.2	80.8	86.0	47.1	45.7	50.6	58.9	70.2	76.7	71.4	72.3
<i>Give s/th</i>	62.9	74.9	77.8	76.1	38.8	53.1	62.7	50.1	52.7	47.9	55.4	52.0
<i>Use computer</i>	73.4	67.1	71.9	65.0	76.3	80.3	78.7	75.9	85.8	81.8	81.4	86.1
<i>Write</i>	52.3	46.4	41.9	45.9	37.9	50.4	50.2	54.0	66.9	64.6	57.7	54.9
<i>Stairway down</i>	87.6	77.7	79.3	78.6	75.4	55.0	53.3	52.7	74.5	69.5	66.8	52.7
<i>Close door</i>	76.4	86.3	81.2	83.9	48.9	58.1	59.0	53.8	78.6	79.0	66.9	62.0
<i>Stairway up</i>	75.9	66.3	62.8	61.2	62.6	50.6	47.5	57.5	63.8	79.2	63.1	68.7
<i>Throw s/th</i>	48.4	71.4	70.5	70.3	45.4	44.0	53.3	64.4	50.8	56.2	60.2	52.6
<i>Get in/out car</i>	74.5	87.3	93.2	94.0	72.7	86.2	88.5	84.8	53.5	78.6	80.0	85.3
<i>Hang up phone</i>	62.2	47.1	74.2	59.2	62.9	65.8	46.8	45.2	65.2	46.3	41.7	49.6
<i>Eat</i>	58.2	64.5	66.6	61.0	54.2	67.4	67.8	64.1	40.4	45.6	60.7	60.6
<i>Answer phone</i>	48.2	65.9	58.5	50.5	43.3	55.6	36.5	57.8	36.8	48.7	45.6	61.2
<i>Clap</i>	76.3	58.0	67.3	75.8	58.7	52.4	56.9	63.5	60.0	54.0	71.7	90.4
<i>Dress up</i>	52.3	64.9	69.5	66.2	42.2	46.2	46.8	44.8	42.2	49.8	57.7	66.4
<i>Undress</i>	52.5	72.5	66.5	71.1	69.6	78.8	59.5	62.9	39.7	57.7	48.2	64.1
<i>Kiss</i>	72.6	83.6	65.0	54.7	51.9	47.3	44.7	58.8	49.6	70.6	67.7	79.5
<i>Fall / trip</i>	70.5	75.9	93.4	90.6	43.9	48.9	53.7	50.1	44.7	74.8	74.7	68.0
<i>Wave</i>	58.0	72.3	38.8	58.2	63.7	73.0	49.6	60.1	38.1	53.9	49.9	70.4
<i>Pour</i>	73.3	84.3	79.7	62.0	47.7	79.8	80.8	81.8	48.8	86.8	75.4	56.6
<i>Punch</i>		98.4	92.2	93.9		71.7	72.0	67.8		49.6	53.0	44.9
<i>Fire weapon</i>	81.3	76.6	92.4	85.6	99.9	69.2	94.2	89.3	99.8	35.3	72.9	52.2
Mean	68.8	74.1	74.8	73.2	58.7	61.1	60.2	60.9	59.9	63.3	64.5	65.6

Table 2. Calibrated average precision for online action detection. Analysis in function of the amount of motion: quartiles selected based on average number of trajectories.