SEMBED: Semantic Embedding of Egocentric Action Videos Supplementary Material

Michael Wray*, Davide Moltisanti*, Walterio Mayol-Cuevas and Dima Damen

Department of Computer Science,
University of Bristol
<FirstName>.<LastName>@bristol.ac.uk

1 Verbs distribution in CMU-MMAC, GTEA+ and BEOID

In this Section we report the distributions of the annotated verbs of the CMU-MMAC [2], GTEA+ [3] and BEOID [1] datasets. Figures 1 to 3 illustrate the number of videos per annotation of the three datasets. As is clearly visible from the figures, all the datasets show an evident power-like trend in the distribution of videos per annotation. This trend becomes more remarkable as the number of classes increases from 12 (CMU-MMAC), to 25 (GTEA+) and 75 (BEOID).

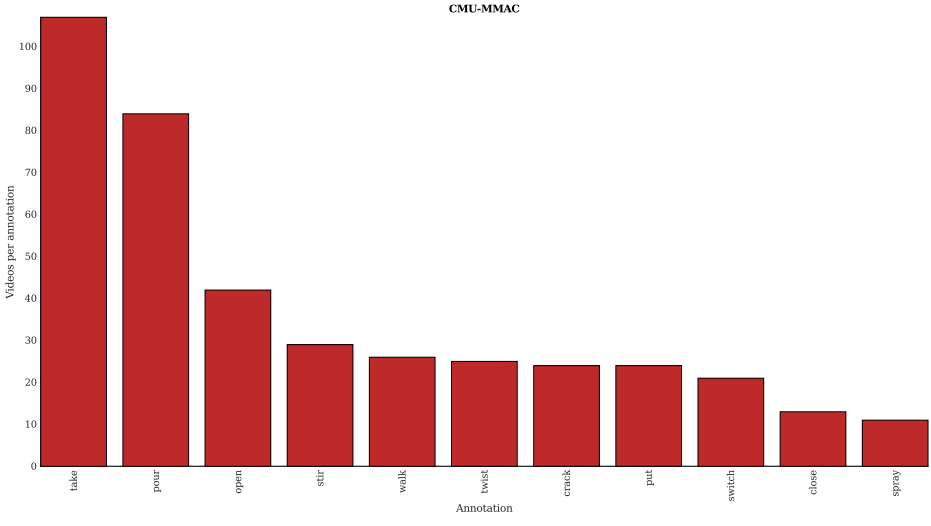

Fig. 1: Annotated verbs distribution for the CMU-MMAC dataset [2].

* Both authors contributed equally to this work

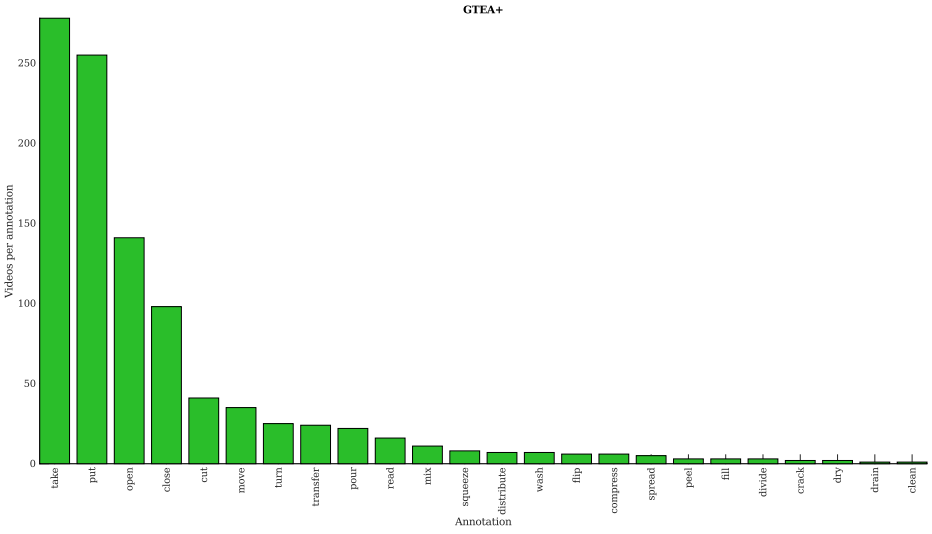

Fig. 2: Annotated verbs distribution for the GTEA+ dataset [3].

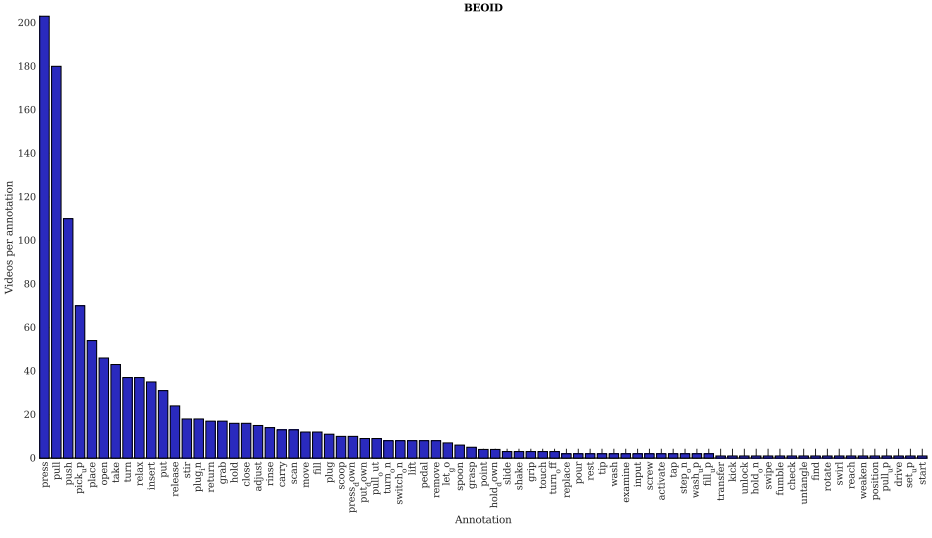

Fig. 3: Annotated verbs for the BEOID dataset [1].

2 Verbs-meanings distribution in BEOID

In this Section we report the distributions of the verbs obtained considering three different semantic relationships between the annotated verbs, which are Action Meaning (AM, Figure 4), Action Synset (AS, Figure 5) and Action Hyponym (AH, Figure 6). In this case the number of verbs were 108, 102 and 84 for AM,

AS and AH. As in the case where no verb meanings were not considered, all the distributions follow a power-like trend.

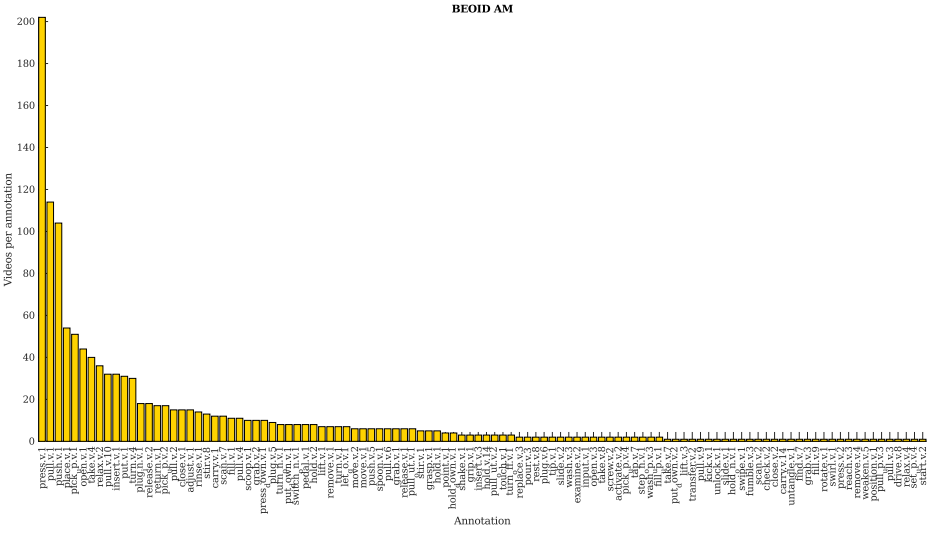

Fig. 4: Action Meaning distribution for the BEOID dataset [1].

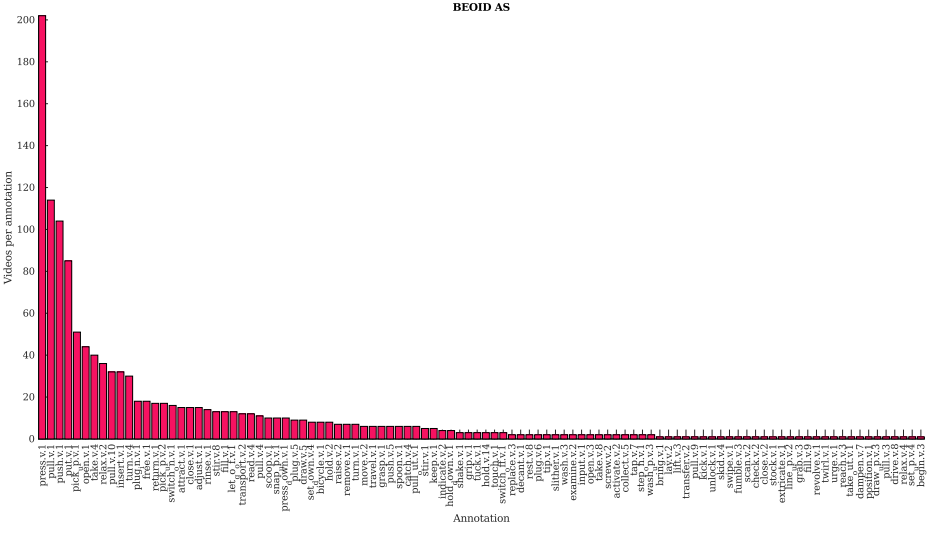

Fig. 5: Action Synset distribution for the BEOID dataset [1].

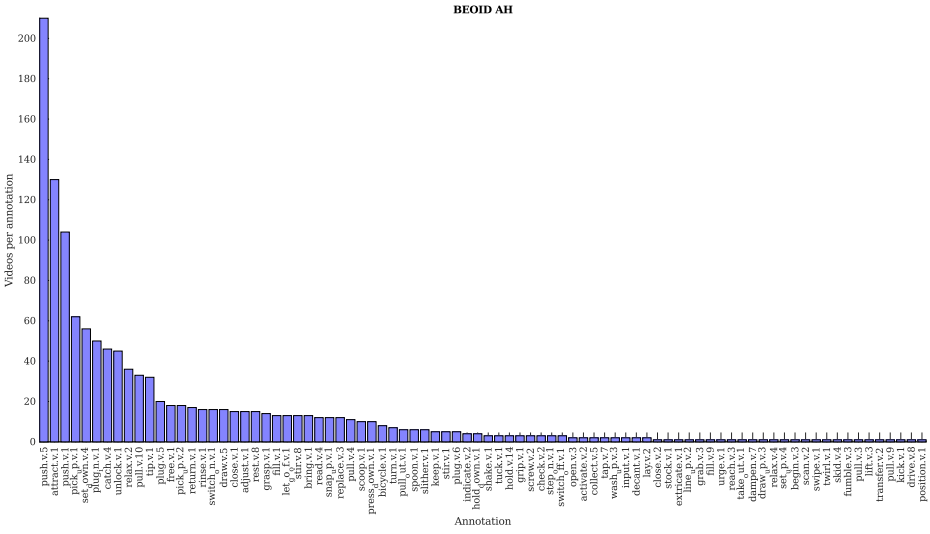

Fig. 6: Action Hyponym distribution for the BEOID dataset [1].

3 SEMBED parameters evaluation

Figures 7 to 9 show SEMBED results obtained on CMU-MMAC, GTEA+ and BEOID with different ⟨features,encoding⟩ combinations and z, t values ranging from 1 to 20. A cell colour in the figures corresponds to the accuracy obtained with the given z and t. The shown results were obtained with $m = 240$. Similarly, Figures 10 to 12 show z and t evaluation pictures for BEOID Action Meaning, Action Synset and Action Hyponym.

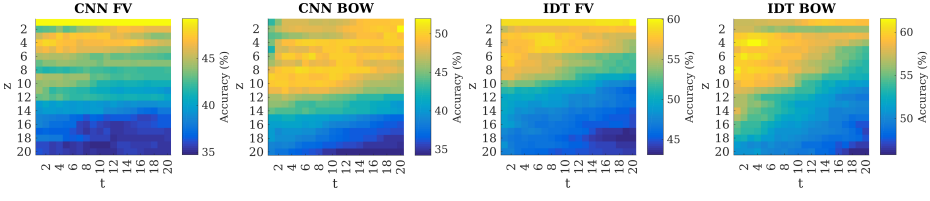

Fig. 7: CMU-MMAC [2] SEMBED z and t evaluation.

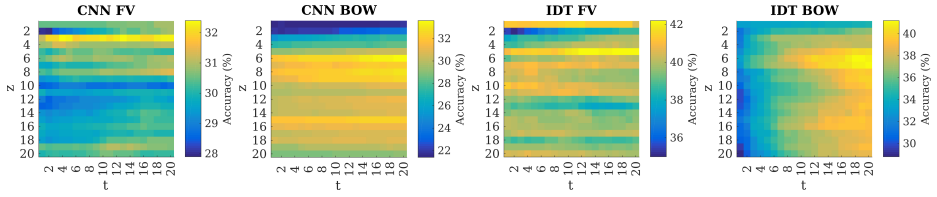

Fig. 8: GTEA+ [3] SEMBED z and t evaluation.

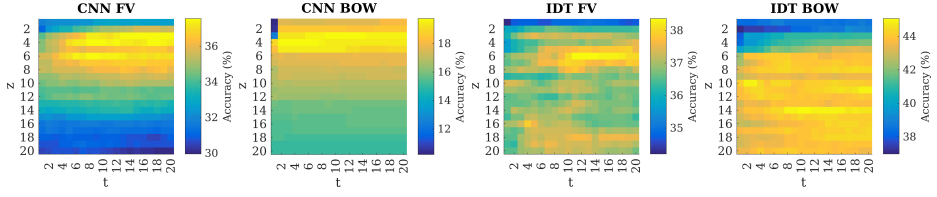

Fig. 9: BEOID (no meanings) [1] SEMBED z and t evaluation.

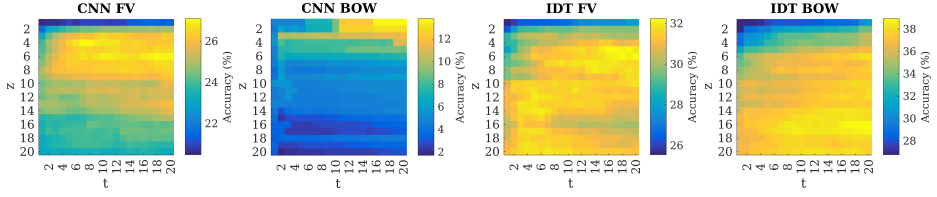

Fig. 10: BEOID Action Meaning [1] SEMBED z and t evaluation.

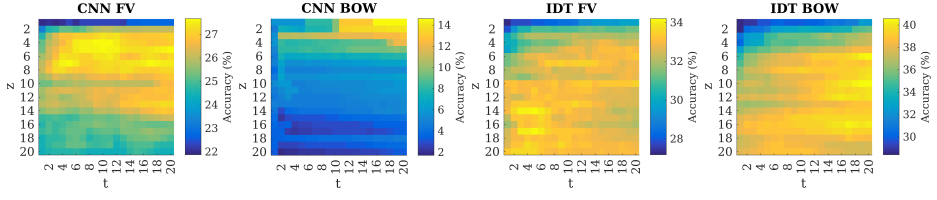

Fig. 11: BEOID Action Synset [1] SEMBED z and t evaluation.

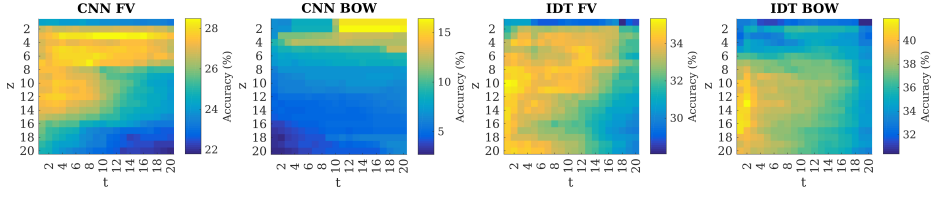

Fig. 12: BEOID Action Hyponym [1] SEMBED z and t evaluation.

References

1. Damen, D., Leelasawassuk, T., Haines, O., Calway, A., Mayol-Cuevas, W.W.: You-do, I-learn: Discovering task relevant objects and their modes of interaction from

multi-user egocentric video. In: BMVC (2014)
2. De La Torre, F., Hodgins, J., Bargteil, A., Martin, X., Macey, J., Collado, A., Beltran, P.: Guide to the Carnegie Mellon University Multimodal Activity (CMU-MMAC) database. Robotics Institute p. 135 (2008)
3. Fathi, A., Li, Y., Rehg, J.M.: Learning to recognize daily actions using gaze. In: Computer Vision–ECCV 2012. pp. 314–327. Springer (2012)