ChildLens: An Egocentric Video Dataset for Activity Analysis in Children

Nele-Pauline Suffo[1], Pierre-Etienne Martin[2], Daniel Haun[2], & Manuel Bohn[1]

[1] Institute of Psychology in Education, Leuphana University Lüneburg

[2] Max Planck Institute for Evolutionary Anthropology

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line. Enter author note here.

The authors made the following contributions. Nele-Pauline Suffo: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Manuel Bohn: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to Nele-Pauline Suffo, Universitätsallee 1, 21335 Lüneburg. E-mail: nele.suffo@leuphana.de

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words "**here we show**" or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

ChildLens: An Egocentric Video Dataset for Activity Analysis in Children

Introduction

Dataset Overview

Activity Classes.

- 14 classes in total
- some focus on person alone: like child talking
- some on person object interaciton: like drawing, playing with an object

Statistics: For every of the 14 activity classes we have a different number of clips ranging from *x-x* clips per class. The clips have different lengths based on the activities where audio-related actions like "child talking" can only last a few seconds up to activities like "reading a book" lasting several minutes. There is a total of xxx video files that are split into *xx-xx* videos per class for training, *xx** videos per class for validation and *xx* videos per class for testing. Table @ref(tab:table_train_val_test) provides the corresponding statistics. @ref(tab:table_train_val_test)

training	validation	testing
xx	xx	xx

How the Dataset was Built

Step 1: Generating a labeling strategy

Step 2: Manual labelling process

Discussion: Dataset bias

Benchmark Performance

Boundary-Matching Network

We utilize the BMN model (Lin, Liu, Li, Ding, & Wen, 2019) for temporal activity localization.

VTC

Implementation details

Conclusion

We used R (Version 4.4.1; **R-base?**) and the R-packages *bookdown* (Version 0.41; **R-bookdown?**), *papaja* (Version 0.1.3; **R-papaja?**) and *tinylabels* (Version 0.2.4; **R-tinylabels?**) for all our analyses.

Results

Discussion

References

Appendix

List of ChildLens Activity Classes

The dataset contains the following list of activities. The number of clips for each activity class is indicated by the number in brackets behind each class.

1. playing with object TBD
2. playing without object \textcolor{red}{TBD}
3. pretend play TBD
4. watching something TBD
5. reading book TBD
6. child talking TBD
7. other person talking TBD
8. overheard speech TBD
9. drawing TBD
10. crafting things TBD
11. singing / humming TBD
12. making music TBD
13. dancing TBD
14. listening to music / audiobook TBD

Lin, T., Liu, X., Li, X., Ding, E., & Wen, S. (2019). *BMN: Boundary-Matching Network for Temporal Action Proposal Generation.* arXiv. https://doi.org/10.48550/ARXIV.1907.09702