Leveraging Machine Learning and Wearable Cameras to Analyze Children's Social Interactions

Nele-Pauline Suffo
Leuphana University, Lüneburg, Germany

Pierre-Etienne Martin
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Anam Zahra
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Daniel Haun
Max Planck Institute for Evolutionary Anthropology, Leipzig, Saxony, Germany

Manuel Bohn
Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Abstract

Direct insights into children's daily experiences are limited despite their importance for development (Rogoff et al., 2018). Traditional methods, like laboratory play sessions, fail to capture naturalistic interactions. Wearable recording devices provide richer data, but their sheer volume challenges traditional coding. We introduce a machine-learning approach to analyze children's everyday interactions. Sixty-four children (ages 3–5) in Leipzig, Germany, wore vests with small cameras, recording 224 hours of video since 03/2020. Our analysis focuses on social interaction cues: person presence, face, gaze, and voice. Using YOLO11, we achieved 80% accuracy in person detection and a 0.9 F1 score for face detection. Preliminary results indicate children often spend time alone or with one person, with face presence in only 17.75% of frames. We will next integrate gaze and voice detection to assess child-directed speech. Our machine-learning approach provides novel insights into children's natural social environments, advancing research on early development.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*