

Detecting social information in a dense database of infants natural visual experience

Bria Long

Stanford University, Stanford, California, United States

George Kachergis

Stanford University, Stanford, California, United States

Ketan Agrawal

Stanford University, Stanford, California, United States

Michael Frank

Stanford University, Stanford, California, United States

Abstract

The faces and hands of caregivers and other social partners offer a rich source of social and causal information that may be critical for infants cognitive and linguistic development. Previous work using manual annotation strategies and cross-sectional data has found systematic changes in the proportion of faces and hands in the egocentric perspective of young infants. Here, we examine the prevalence of faces and hands in a longitudinal collection of nearly 1700 headcam videos collected from three children along a span of 6 to 32 months of age the SAYCam dataset (Sullivan, Mei, Perfors, Wojcik, & Frank, under review). To analyze these naturalistic infant egocentric videos, we first validated the use of a modern convolutional neural network of pose detection (OpenPose) for the detection of faces and hands. We then applied this model to the entire dataset, and found a higher proportion of hands in view than previous reported and a moderate decrease the proportion of faces in childrens view across age. In addition, we found variability in the proportion of faces/hands viewed by different children in different locations (e.g., living room vs. kitchen), suggesting that individual activity contexts may shape the social information that infants experience.