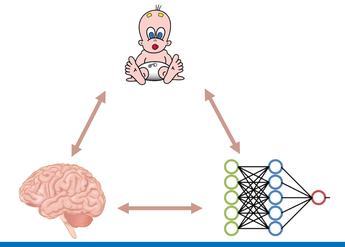
Objects or Context? Learning From Temporal Regularities in Continuous Visual Experience With an Infant-inspired DNN

Cliona O'Doherty, Rhodri Cusack
School of Psychology and Trinity College Institute of Neuroscience, Trinity College Dublin, Ireland

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin



Overview

Infant developmental science is a rich source of inspiration for the next generation of DNN models (Zaadnoordijk, Besold & Cusack, 2022).

We implemented an infant-inspired learning mechanism into a self-supervised DNN, by using contrastive learning to find commonalities in naturalistic videos across various timescales.

We test if longer timescales reflect more slowly changing scene-context, with shorter timescales reflecting object-centric features.

Find the paper here.

Motivation

Infants learn to recognise and understand the things they see by observing the environment with minimal supervision.

They learn by leveraging hidden statistical signals (Saffran & Aslin, 1996), through co-occurrence patterns (Unger, Vales & Fisher, 2021) and by comparing objects across time (Oakes & Ribar, 2005).

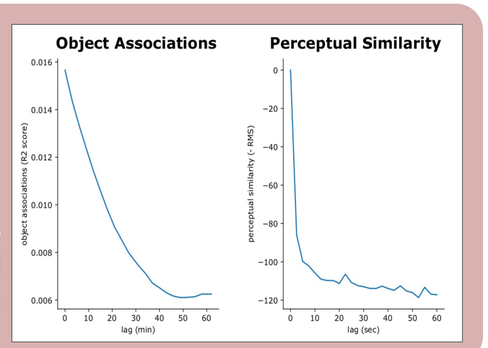
What will a DNN learn if it is trained in this infant-inspired manner?

Model and Training

- Generated a new dataset to test learning: 154 hr live action movies, 573,000 images (sampling interval = 1 sec).
- Trained self-supervised CMC (Tian et al., 2019) to build a representation by contrastive learning. The model relates two images separated by a specified lag.
- The naturalistic, persistent temporal co-occurrence patterns in the movies provide a signal for learning wider object/context associations.

Objects or Contexts Over Time?

- Autoregression revealed perceptual similarity drops off quicker (40 s) than object co-occurrences (40 min).
- Longer timescales may be informative for contextual understanding built from associations.



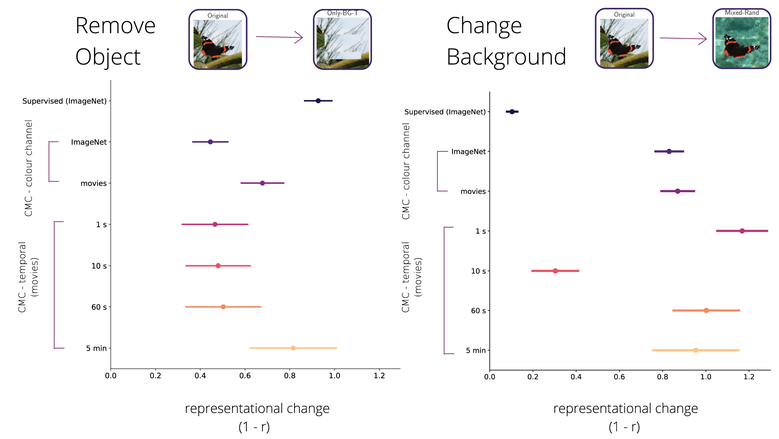
We used this difference in object co-occurrences over various timescales to train CMC using our movie dataset.

Representational Similarity Analysis

- Using the Imagenet-9 categories in the Madry background challenge (Xiao et al., 2020) we assessed what each network learned when given access to different timescales.

- Permuted Pearson correlations from RDMs of network activations were used to quantify the extent of representational change with object removal or background change.

Learning differently depending on timescale



While the supervised network was sensitive to object removal as expected, the temporal networks were sensitive to backgrounds.

- The 5 min network was sensitive to both objects and contexts. Could longer or intermediate timescales enable more holistic representations to be learned?
- Future work will validate network performances, and extend the range of tests explored here to expand upon preliminary results.

References

Zaadnoordijk, L., Besold, T. R., & Cusack, R. (2022). Lessons from infant learning for unsupervised machine learning. Nature Machine Intelligence, 4(6), 510-520.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274(5294), 1926-1928.
Unger, L., Vales, C., & Fisher, A. V. (2020). The Role of Co-Occurrence Statistics in Developing Semantic Knowledge. Cognitive Science, 44(9), e12894.
Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. Infancy, 7(1), 85-98.
Tian, Y., Krishnan, D., & Isola, P. (2019). Contrastive multiview coding. arXiv preprint arXiv:1906.05849.
Xiao, K., Engstrom, L., Ilyas, A., & Madry, A. (2020). Noise or signal: The role of image backgrounds in object recognition. arXiv preprint arXiv:2006.09994.

Acknowledgements

Funded by the Irish Research Council under grant number GOIPG/2021/223 and ERC Advanced Grant ERC-2017-ADG, FOUNDCOG, 787981.

Authors declare no conflict of interest.