Unsupervised Pretraining in Biological Neural Networks

Lin Zhong, Scott Baptista, Rachel Gattoni, Jon Arnold, Daniel Flickinger, Carsen Stringer, Marius Pachitariu

Janelia Research Campus, Howard Hughes Medical Institute, USA

Email: zhongl@hhmi.janelia.org

Representation learning in neural networks may be implemented with supervised or unsupervised algorithms, distinguished by the presence or absence of reward feedback. Both types of learning are highly effective in artificial neural networks. In biological systems, task learning has been shown to modify sensory neural representations, but it is not known if these changes are due to supervised or unsupervised learning. Here we recorded populations of up to 70,000 neurons simultaneously from primary visual cortex (V1) and higher visual areas (HVA), while mice learned multiple tasks as well as during unrewarded exposure to the same stimuli. We found that most neural changes in task mice were replicated in the mice with unrewarded exposure. These changes were concentrated in the medial HVAs after mice learned to discriminate visual textures from two different classes. In contrast, the changes were widespread across visual areas after mice learned to discriminate between two exemplars of the same visual class. In both tasks, neural representations of the most recently learned exemplar generalized to new exemplars of the same visual category and the behavior of the mice generalized according to the same rule. These specific neural changes were replicated in mice with unrewarded exposure. In task mice only, we found a neural population in anterior HVAs encoding a ramping reward prediction signal, potentially involved in the supervised learning. Our neural results predict that unsupervised pretraining may accelerate subsequent task learning, a prediction which we validated with behavioral training experiments.