Title: **An artificial neural network based method to decode visual representations from primary sensory cortex activity in awake mice**

Authors: Antonio Fernández, Melina Timplalexi, Adam Ranson.

In this study, our aim is to develop a data-driven method for decoding internal visual representations from sensory cortex activity in large populations of single neurons in the primary visual cortex of mice. The broader objective of the study is to apply this decoder to reconstruct visual representations from neural activity measured during sleep and dreaming.

Instead of making assumptions about the receptive field properties of recorded neurons (such as approximating them with a Gabor function, etc.), the convolutional artificial neural-based decoder learns the visual stimulus → neural activity mapping directly from the data. This approach allows for the incorporation of additional variables, including pupil dilation (an index of arousal state), eye position, and locomotion speed as inputs to the model. It also enables the discovery of more complex non-linear visual receptive field properties.

We present results from both simulated neural responses to images and an in vivo neural dataset recorded using 2-photon calcium imaging. In the simulated dataset, we demonstrate the feasibility of this assumption-free decoding approach and examine the impact of various parameters, including neural receptive field size, response variability, number of neurons used to train the model, number of training images, and other model parameters on decoding performance.

Additionally, we use the simulated dataset to develop and validate a method to estimate the size of a training set required to approach optimal cross-validated decoding performance. We then use the in vivo dataset to gain insight into the characteristics the visual stimulus should have to optimally train such a model.

We observe 'habituation' of average population neural activity over a period of 30 seconds of continuous visual stimulation with a video of a natural scene, which could be highly detrimental to decoder model training. To test if this habituation is due to a non-specific loss of neural responsiveness over time, or alternatively due to a gradual sparsification of representation (while retaining signal content), we calculate sample entropy as a function of stimulus duration.

Finally, we present the results of training a decoder model on neural activity elicited by natural or artificial stimuli in vivo, assess the benefit of incorporating behavioral parameters into the model (eye position, dilation, and locomotion state), and attempt to estimate the parameters required for a more comprehensive training set based on the model learning rate.