Fabio Anselmi, PhD
Assistant Professor
Center for Neuroscience and Artificial Intelligence

Neurally Plausible Mechanisms for Learning Selective and Invariant Representations

Fabio is an assistant professor at the Center for Neuroscience and Artificial Intelligence at Baylor College of Medicine where he has been faculty member since 2020. He is also affiliated to the Centre for Brains Minds and Machines at MIT.
Fabio completed his PhD in quantum mechanics at Hertforshire University (Uk) and he graduated in Physics at Padova University in Italy.

His work lies at the interface between computational neuroscience and machine learning with focus on development of biologically grounded machine learning algorithms with application to visual cortex.

Coding for visual stimuli in the ventral stream is known to be invariant to object identity preserving nuisance transformations. Indeed, much recent theoretical and experimental work suggests that the main challenge for the visual cortex is to build up such nuisance invariant representations. Recently, artificial convolutional networks have succeeded in both learning such invariant properties and, surprisingly, predicting cortical responses in macaque and mouse visual cortex with unprecedented accuracy. However, some of the key ingredients that enable such success—supervised learning and the backpropagation algorithm—are neurally implausible. This makes it difficult to relate advances in understanding convolutional networks to the brain. In contrast, many of the existing neurally plausible theories of invariant representations in the brain involve unsupervised learning, and have been strongly tied to specific plasticity rules. To close this gap, we study an instantiation of simple-complex cell model and show, for a broad class of unsupervised learning rules (including Hebbian learning), that we can learn object representations that are invariant to nuisance transformations belonging to a finite orthogonal group. These findings may have implications for developing neurally plausible theories and models of how the visual cortex or artificial neural networks build selectivity for discriminating objects and invariance to real-world nuisance transformations.