Architecture matters: How Well Neural Networks Explain It Representation Does Not Depend on Depth and Performance Alone

Submission ID 3000233

Submission Type Poster

Topic Neuroscience

Status Submitted

Submitter Katherine Storrs

Affiliation MRC Cognition and Brain Sciences Unit, Cambridge, UK

SUBMISSION DETAILS

Presentation Type Either Poster or Oral Presentation

Presentation Abstract Summary Humans are able to classify complex visual objects with extremely high accuracy. Recently, deep convolutional neural network (DCNN) models have reached and even surpassed human performance at this task. Among recent networks, the deeper the architecture, the better the performance. Although loosely inspired by biological brains, it remains unclear whether models reaching human-level accuracy also perform computations similar to those in the human brain. In earlier studies using shallower architectures with poorer object classification accuracy, greater depth and higher task performance were associated with improved explanation of inferior temporal cortex (IT) (Khaligh-Razavi & Kriegeskorte, 2014; Yamins et al., 2014). Our results show that this is not the case for state-of-the-art deep architectures that near or surpass human performance; the deepest, best-performing models are not best at explaining representations in human IT. In particular, deep residual networks (ResNets) are a relatively poor match to the brain, despite their very high classification performance. These findings open the door to detailed explorations of the architectures that best account for the representational transformations, and thus computations, performed in the ventral visual stream.

Co-author Information

* Presenting Author

First Name	Last Name	Affiliation	E-mail
Katherine *	Storrs *	MRC Cognition and Brain Sciences Unit, Cambridge, UK	katherine.storrs@mrc-cbu.cam.ac.uk
Johannes	Mehrer	MRC Cognition and Brain Sciences Unit, Cambridge, UK	johannes.mehrer@mrc-cbu.cam.ac.uk

Alexander	Walther	YouGov, London	awalthermail@gmail.com
Nikolaus	Kriegeskorte	MRC Cognition and Brain Sciences Unit, Cambridge	nikokriegeskorte@gmail.com

Keywords

Keywords
object recognition
visual cortex
Deep Neural Networks
fMRI
representational similarity analysis