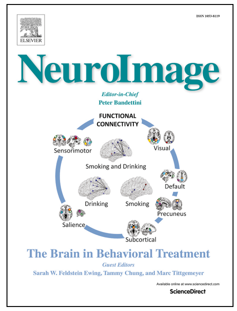

Convolutional neural network-based encoding and decoding of visual object recognition in space and time

K. Seeliger, M. Fritsche, U. Güçlü, S. Schoenmakers, J.-M. Schoffelen, S.E. Bosch, M.A.J. van Gerven

PII: S1053-8119(17)30586-4

DOI: 10.1016/j.neuroimage.2017.07.018

Reference: YNIMG 14183

To appear in: *NeuroImage*

Received Date: 18 March 2017

Revised Date: 1053-8119 1053-8119

Accepted Date: 10 July 2017

Please cite this article as: Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J.-M., Bosch, S.E., van Gerven, M.A.J., Convolutional neural network-based encoding and decoding of visual object recognition in space and time, *NeuroImage* (2017), doi: 10.1016/j.neuroimage.2017.07.018.

Convolutional Neural Network-based Encoding and Decoding of Visual Object Recognition in Space and Time

K. Seeliger[*], M. Fritsche, U. Güçlü, S. Schoenmakers,
J.-M. Schoffelen, S. E. Bosch, M. A. J. van Gerven

Radboud University, Donders Institute for Brain, Cognition and Behaviour
Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Abstract

Representations learned by deep convolutional neural networks (CNNs) for object recognition are a widely investigated model of the processing hierarchy in the human visual system. Using functional magnetic resonance imaging, CNN representations of visual stimuli have previously been shown to correspond to processing stages in the ventral and dorsal streams of the visual system. Whether this correspondence between models and brain signals also holds for activity acquired at high temporal resolution has been explored less exhaustively. Here, we addressed this question by combining CNN-based encoding models with magnetoencephalography (MEG). Human participants passively viewed 1000 images of objects while MEG signals were acquired. We modelled their high temporal resolution source-reconstructed cortical activity with CNNs, and observed a feed-forward sweep across the visual hierarchy between 75–200 ms after stimulus onset. This spatiotemporal cascade was captured by the network layer representations, where the increasingly abstract stimulus representation in the hierarchical network model was reflected in different parts of the visual cortex, following the visual ventral stream. We further validated the accuracy of our encoding model by decoding stimulus identity in a left-out validation set of viewed objects, achieving state-of-the-art decoding accuracy.

Keywords

Visual neuroscience, deep learning, encoding, decoding, magnetoencephalography

1 Introduction

Automatic object recognition has been a long-standing and difficult research problem in computer vision. A few years ago, driven by the availability of large-scale computing and training data resources, automated object recognition has reached and surpassed human-level performance. The most successful object recognition models by far are deep feed-forward convolutional neural networks (CNNs), which can learn statistical properties of structured data such as natural images well (Schmidhuber, 2014).

CNNs can be seen as an abstraction of rate-based coding in biological neural circuits (Dayan and Abbott, 2005) and are inspired by the anatomical wiring of the visual system as a hierarchy of processing stages (Felleman and Van Essen, 1991). Receptive field properties become increasingly complex higher up in this visual hierarchy. That is, receptive fields in striate cortex respond to oriented bars in the visual input (Hubel and Wiesel, 1959), whereas receptive fields in inferior temporal cortex respond specifically to complex object properties. This is very similar to CNNs, which learn simple edge detectors in early layers and more abstract object features in higher layers. CNN-like architectures have been motivated from a neuroscientific point of view with the

[*]Corresponding author: ✉ kseeliger@posteo.jp