DeepLight: A Structured Framework For The Analysis of Neuroimaging Data Through Recurrent Deep Learning Models

Armin W. Thomas (athms.research@gmail.com)
Machine Learning Group, Technische Universität Berlin, Marchstr. 23, Berlin 10587, Germany

Hauke R. Heekeren, Klaus-Robert Müller, Wojciech Samek

We propose a structured framework for the application of recurrent deep learning (DL) models to the analysis of fMRI data. To identify an association between cognitive state and brain activity, DeepLight utilizes the layer-wise relevance propagation method. Thereby, decomposing the DL model's decoding decisions into the contributions of the single input voxels to these decisions. Importantly, DeepLight is able to identify this association on multiple levels of data granularity, from the level of the group down to single subjects, trials and time points.

Keywords: neuroimaging; decoding; deep learning; interpretability

Neuroimaging researchers have started collecting large and structured corpora of experimental functional Magnetic Resonance Imaging (fMRI) data. At first sight, the analysis of fMRI data thereby seems ideally suited for the application of deep learning (DL) methods. Yet, two major challenges have so far prevented broad DL usage:

1. fMRI data are high dimensional, while containing comparably few samples.
2. DL models are often considered as nonlinear *black boxes*, disguising any relationship between input data (i.e., brain activity) and decoding decision (i.e., cognitive state).

Here, we approach these challenges by proposing the DeepLight framework (see Figure 1):

1. To decode a cognitive state from a whole-brain fMRI volume, DeepLight first separates this volume into a sequence of 2D axial slices. These slices are then processed by a convolutional neural network, resulting in a sequence of higher-level slice representations. This sequence is fed to a bi-directional long short-term memory unit, modeling the spatial dependencies of brain activity within and across axial brain slices. Thereby, DeepLight drastically reduces the dimensionality of the input (as well as the number of required model parameters), while still fully accounting for the wide-spread spatial dependencies of whole-brain activity.
2. To maintain an interpretability of the fMRI data, DeepLight utilizes the layer-wise relevance propagation (LRP) method. LRP decomposes the decoding decision of the DL model into the contributions of the single input voxels to the decision.

To demonstrate the versatility of DeepLight, we apply it to an openly available fMRI dataset of the Human Connectome Project, in which 100 subjects viewed images of either body parts, faces, places or tools. We evaluate the performance of DeepLight in decoding these images from the fMRI data as well as identifying the brain regions associated with each image class. To this end, we compare DeepLights performance to that of three conventional uni- and multivariate means of neuroimaging analysis, namely, the General Linear Model, searchlight analysis, and whole-brain Least Absolute Shrinkage Regression.

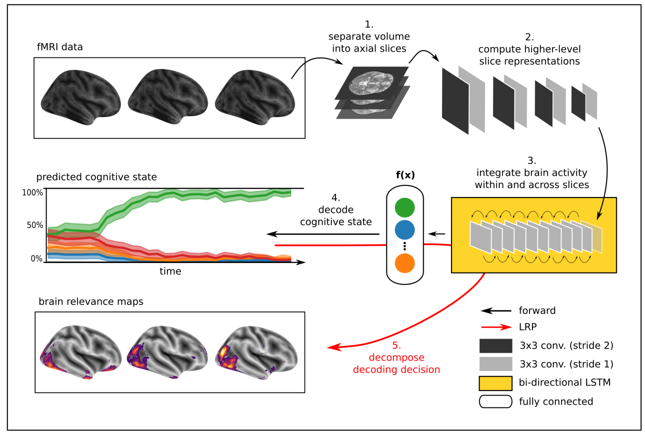

Figure 1: The DeepLight framework.

We find that DeepLight (1) decodes the cognitive states underlying the fMRI data more accurately than these other approaches, (2) better improves its decoding performance with growing datasets, (3) accurately identifies the physiologically appropriate associations between cognitive states and brain activity, and (4) identifies these associations on multiple levels of data granularity (the level of the group, subject, trial and time point). Interestingly, DeepLight also accurately captures the temporo-spatial distribution of brain activity over sequences of single fMRI samples. Thereby, enabling new means to study this distribution wihin single trials.

Acknowledgments

This work was supported by the German Federal Ministry for Education and Research (Berlin Big Data Centre; 01IS14013A), the Berlin Center for Machine Learning (01IS18037I), the TraMeExCo project (01IS18056A). KRM is also supported by the Korea government (2017-0-00451).