End-to-end reconstruction of natural images from multi-unit recordings with Brain2Pix

Lynn Le (lynn.le@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands

Paolo Papale (p.papale@nin.knaw.nl)
Netherlands Institute for Neuroscience
Amsterdam, Netherlands

Antonio Lozano (a.lozano@nin.knaw.nl)
Netherlands Institute for Neuroscience
Amsterdam, Netherlands

Thirza Dado (thirza.dado@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands

Feng Wang (f.wang@nin.knaw.nl)
Netherlands Institute for Neuroscience
Amsterdam, Netherlands

Marcel van Gerven (marcel.vangerven@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands

Pieter Roelfsema (p.roelfsema@nin.knaw.nl)
Netherlands Institute for Neuroscience
Department of Integrative Neurophysiology, Centre for Neurogenomics and Cognitive Research, Vrije Universiteit
Department of Psychiatry, Amsterdam UMC, University of Amsterdam
Amsterdam, Netherlands
Laboratory of Visual Brain Therapy
Paris, France.

Yağmur Güçlütürk (y.gucluturk@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands

Umut Güçlü (u.guclu@donders.ru.nl)
Donders Institute for Brain, Cognition and Behaviour
Radboud University
Nijmegen, Netherlands

Abstract

Reconstructing naturalistic images from brain signals has been a challenging task for scientists, with successful results largely limited to large human fMRI datasets. In this study, we apply the brain2pix reconstruction model to multi-unit activity (MUA) data from the macaque brain, providing a novel extension of the model. This approach allows for investigation of information representation in different brain regions and time windows, with greater spatial and temporal precision. Our results offer insights into the neural basis of visual perception, showing that V1 neurons represent texture and color, V4 neurons exhibit symmetric representations, and IT neurons reveal concept-like features. We also demonstrate that the model can be used to decode features at different layers of a neural network, with V1 more strongly correlated with initial layers and V4 and IT with deeper layers. Overall, our approach provides a valuable tool for studying brain representations in high temporal and spatial detail.

Keywords: Neural Decoding; Naturalistic Images; Multi-unit Activity; Macaque

Introduction

Considerable efforts have been made to build accurate reconstruction models for decoding naturalistic images from brain activity (Nishimoto et al., 2011; Naselaris, Prenger, Kay, Oliver, & Gallant, 2009; Shen, Horikawa, Majima, & Kamitani, 2019; Le et al., 2022). Currently, research in this area is concentrated on training the reconstruction models mostly on fMRI signals. On the other hand, a neural decoding model trained on micro-electrode array (MEA) data with high temporal and spatial resolution can bring us closer to understanding the information that is carried by single neurons and neuronal populations.

Methods

Here, we trained and analysed an end-to-end convolutional reconstruction model on a dataset of 22,348 naturalistic images and corresponding MEA recordings from a macaque brain with 576 micro-electrodes across V1, V4 and IT.

Convolutional neural networks (CNNs) are ideal for processing complex topographic multi-dimensional data like naturalistic images. They can exploit the topography by applying convolutions along the topographic dimensions (LeCun et al., 1989; LeCun, Bottou, Bengio, & Haffner, 1998). Considering that each micro-electrode can be precisely mapped to a local region in the visual field as the location of its receptive field, we can use this mapping to transform the MEA responses (to a stimulus image) into a receptive field image which has the same topology as the stimulus image. We can thus make use of a CNN-based image-to-image decoder to process signals from multiple region of interests (ROIs) of the brain as receptive field images, with each ROI signal treated as a separate channel in the data. Similarly, signals from various time-points can be represented as a series of receptive field images over time, with each time point treated as a separate channel in the data. The brain2pix approach was introduced to provide a similar CNN based solution for fMRI data, where voxel responses are first mapped to the visual space based on the retinotopic mappings, resulting in grid-like topological data structures called RFSimages, which become the input of the decoder (Le et al., 2022). Here, we adapt the brain2pix approach to MEA data and make it more spatially and temporally precise, resulting in high-quality reconstructions of natural images that can be space- and time-dependent (i.e., dependent on the brain regions and time windows being analyzed).

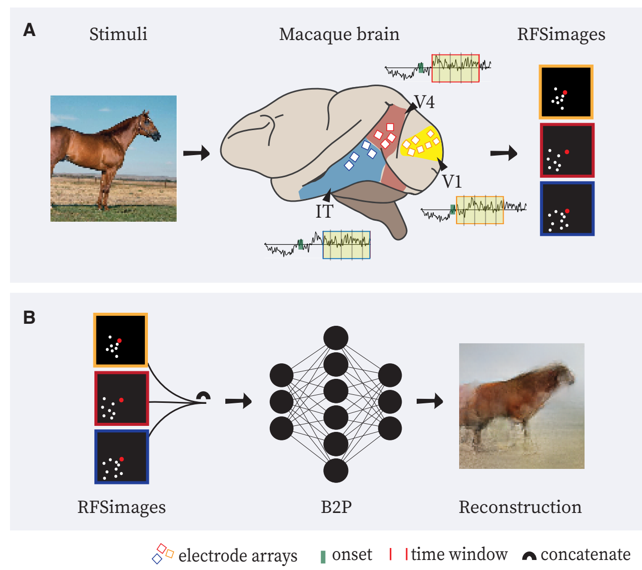

Figure 1: Overview of the reconstruction method. (a) MUA signals are recorded while the monkey is presented with naturalistic images. The brain signals are split into micro-electrode channels and time-windows and then mapped into 2D space according to the retinotopic mapping, to become RFSimages. (b) Then the B2P model, designed to take in the RFSimages is trained resulting in reconstructions.

The training consisted of spatial-based (S-B) and temporal-based (T-B) models. The S-B model partitioned the data into 15 channels based on electrodes in V1, V4, and IT and selected time windows based on their corresponding regions. The T-B model divided the original time windows into five smaller time windows and trained the model on 15 of these shorter time windows, organizing them by region of interest. The S-B approach emphasized the spatial location of the electrodes and the corresponding regions of the brain, while the T-B approach focused on the temporal dynamics of the brain activity by breaking down the time window into smaller segments. In both the S-B and T-B cases, the brain2pix architecture (Le et al., 2022) was used. Figure 1 provides a visual summary of the pipeline.

Results

Both the T-B and S-B models generate reconstructions that precisely capture spatial information, colors, textures, shapes,

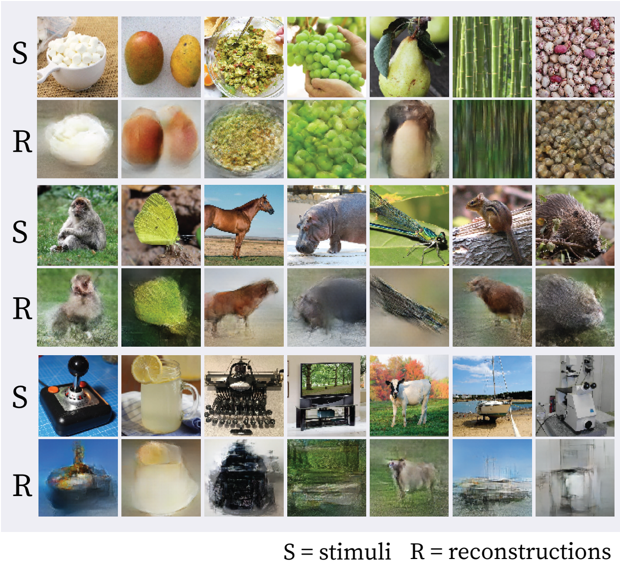

Figure 2: Examples of stimuli and their corresponding reconstructions using our method are shown here. The reconstructions are obtained by averaging the outputs of both the temporal-based (T-B) and spatial-based (S-B) models.

and other features that correspond to the observed stimuli by the monkey. Figure 2 illustrates examples of these combined reconstructions. The reconstructions from each model are available in the GitHub repository[1].

Spatial-based

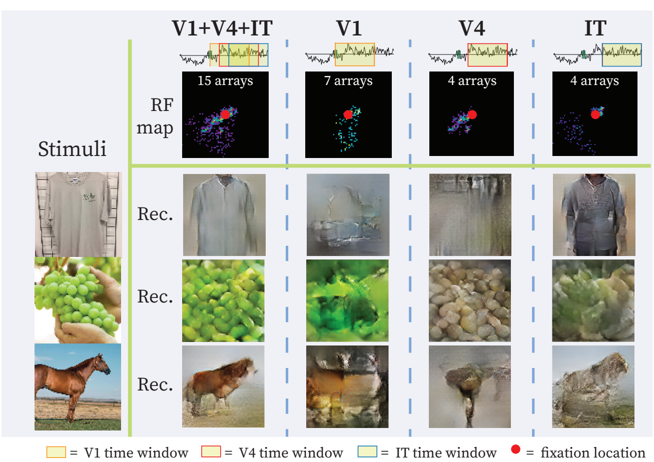

Figure 3: Results from the S-B experiment described in Spatial-based subsection.

The S-B model is used for forward propagation, with input values of electrodes in specific ROIs being kept active and the rest set to 0. The resulting reconstructions show distinct features associated with different brain regions, such as texture and color information in V1 neurons, symmetric representations in V4 neurons, and concept-like features in IT neurons, like a person wearing a jacket (see fig 3, col. 4, row 1).

Table 1 shows Pearson correlation values for AlexNet features. Underlined values indicate the best results among all four conditions, while bold values represent the best results for V1, V4, and IT. The observed values suggest that V1 is highly correlated with the first two neural network layers, while V4 and IT show stronger correlation with deeper layers, consistent with prior research linking image processing in V1 with initial neural network layers (Güçlü & van Gerven, 2015).

Table 1: S-B experiment

	V1+V4+IT	V1	V4	IT
pool1	0.166	**0.132**	0.090	0.092
pool2	0.095	**0.057**	0.051	0.051
pool3	0.069	0.039	0.038	**0.042**
pool4	0.050	0.023	**0.028**	0.027
pool5	0.029	0.011	**0.013**	0.011

Temporal-based

The T-B model is used to perform forward propagation by keeping the input values that belong to the time windows of interest active and setting the rest to 0. The notation TW-15 indicates all time windows, while TW-$3i+n$ refers to the time windows starting at initial milliseconds with 100ms, 133ms, and 166ms corresponding to V1, V4, and IT, respectively (i) plus the shift in milliseconds (n) indicated for the time window. The highest correlation value for the first AlexNet layer was observed in the fourth time-window, while for the rest of the layers, the initial time window ($3i$) showed the highest values.

Table 2: T-B experiment

TW-	15	3_i	3_{i+26}	3_{i+53}	3_{i+78}	3_{i+104}
pool1	0.180	0.097	0.043	0.107	**0.108**	0.082
pool2	0.094	**0.043**	0.043	0.042	0.030	0.030
pool3	0.072	**0.037**	0.036	0.036	0.029	0.029
pool4	0.052	**0.024**	0.023	0.019	0.016	0.016
pool5	0.033	**0.015**	0.013	0.012	0.013	0.010

Discussion

Our study demonstrates that our model can effectively reconstruct intricate stimuli from MUA signals and appears to organize image features into channels associated with different ROIs. Our results highlight the importance of incorporating temporal information in brain decoding models for more accurate reconstructions of naturalistic stimuli. The results suggest that the temporal-based model is more effective than the spatial-based model, possibly due to the presence of relevant temporal information in the signal.

[1] https://github.com/neuralcodinglab/MonkeySee

References

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, *35*(27), 10005–10014.

Le, L., Ambrogioni, L., Seeliger, K., Güçlütürk, Y., Van Gerven, M., & Güçlü, U. (2022). Brain2pix: Fully convolutional naturalistic video frame reconstruction from brain activity. *Frontiers in Neuroscience*, *16*.

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Naselaris, T., Prenger, R. J., Kay, K. N., Oliver, M., & Gallant, J. L. (2009). Bayesian reconstruction of natural images from human brain activity. *Neuron*, *63*(6), 902–915.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology*, *21*(19), 1641–1646.

Shen, G., Horikawa, T., Majima, K., & Kamitani, Y. (2019). Deep image reconstruction from human brain activity. *PLoS computational biology*, *15*(1), e1006633.