The Algonauts Project: A Platform for Communication between the Sciences of Biological and Artificial Intelligence

Radoslaw Martin Cichy[1,*] (radoslaw.cichy@fu-berlin.de), Gemma Roig[2] (gemmar@mit.edu), Alex Andonian[3] (aandonia@mit.edu), Kshitij Dwivedi[2] (kshitij_dwivedi@mymail.sutd.edu.sg), Benjamin Lahner[3] (blahner@mit.edu), Alex Lascelles[3] (alexlasc@mit.edu), Yalda Mohsenzadeh[3] (yalda@mit.edu), Kandan Ramakrishnan[3] (krama@mit.edu), Aude Oliva[3] (oliva@mit.edu)

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

[2]Information Systems Technology and Design, Singapore University Technology and Design, Singapore

[3]Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, USA

[*] corresponding author

Abstract

In the last decade, artificial intelligence (AI) models inspired by the brain have made unprecedented progress in performing real-world perceptual tasks like object classification and speech recognition. Recently, researchers of natural intelligence have begun using those AI models to explore how the brain performs such tasks. These developments suggest that future progress will benefit from increased interaction between disciplines. Here we introduce the Algonauts Project as a structured and quantitative communication channel for interdisciplinary interaction between natural and artificial intelligence researchers. The project's core is an open challenge with a quantitative benchmark whose goal is to account for brain data through computational models. This project has the potential to provide better models of natural intelligence and to gather findings that advance AI. The 2019 Algonauts Project focuses on benchmarking computational models predicting human brain activity when people look at pictures of objects. The 2019 edition of the Algonauts Project is available online: `http://algonauts.csail.mit.edu/`.

Keywords: human neuroscience; vision; object recognition; prediction; challenge; competition; benchmark

Introduction

The quest to understand the nature of human intelligence and engineer advanced forms of artificial intelligence (AI) are increasingly intertwined (Hassabis, Kumaran, Summerfield, & Botvinick, 2017; Kriegeskorte, 2015; Yamins & DiCarlo, 2016). To explain human intelligence, we require computational models that can handle the complexity of real-world tasks. To engineer artificial intelligence, biological systems can provide inspiration and guidance of how to solve the task efficiently.

With this algorithmic exploration paradigm for explaining the brain, it is becoming essential to have standardized benchmarks for comparing how well different algorithms account for neural data. Open challenges are a particular form of standardized benchmark that foster fast-paced advance in a collaborative and transparent manner.

Open challenges have helped science to thrive in many times and fields. As early as 1900, Hilbert proposed 23 problems as challenges in mathematics to be solved. More recently, benchmarks for open competition have emerged in other disciplines such as robotics (e.g. the DARPA robotics challenge) and computer science on a diverse sets of topics including visual recognition (Everingham et al., 2015; Russakovsky et al., 2015; Zhou et al., 2019)), reasoning (Johnson et al., 2017) and natural language understanding (Wang et al., 2018). Those challenges are well accepted in the their scientific communities and suggest standardized benchmarks as fruitful platforms for collaboration.

Inspired by these approaches, we propose a challenge platform with standardized benchmarks for the artificial and biological sciences. At the core of the platform is an open competition with the goal of accounting for brain activity through computational models and algorithms. We coin the platform the Algonauts Project. Inspired by the astronauts (i.e. sailors of the stars) who launched into space to explore a new frontier, the algonauts (i.e. sailors of algorithms) set out to relate brains and computer algorithms in an exploratory way.

We believe that the Algonauts Project will facilitate the interaction between biological and artificial intelligence researchers, allowing the communities to exchange ideas and advance both fields rapidly and in a transparent way.

The 2019 Edition of the Algonauts Project: Explaining the Human Visual Brain

The 2019 edition is the first edition of the Algonauts Project's challenge and workshop. It is titled "Explaining the Human Visual Brain", and its specific target is to determine which computational model best accounts for human visual brain activity.

We focus on visual object recognition as it is an essential cognitive capacity of systems embedded in the real world. Visual object recognition has long fascinated neuroscientists and computer scientists alike, and it is here that the recent advances in AI and their adoption into neurosciences have taken place most prominently. Currently, particular deep neural networks trained with the engineering goal to recognize objects in images do best in accounting for brain activity during visual object recognition (Schrimpf et al., 2018; Bashivan, Kar, & DiCarlo, 2019). However, a large portion of the signal measured in the brain remains unexplained. This is so because we do not have models that capture the mechanisms of the human

brain well enough. Thus, what is needed are advances in computational modelling to better explain brain activity.

Related challenges in neuroscience. The 2019 edition "Explaining the Human Visual Brain" relates to initiatives such as the The neural prediction challenge (`http://neuralprediction.berkeley.edu/`) and brain-score (`http://www.brain-score.org/`) (Schrimpf et al., 2018) that provide benchmarks and leaderboards. The Algonauts Project emphasizes human brain data, and an automated submission procedure with immediate assessment. It couples neural prediction benchmarks to a challenge limited in time, and adds educational and collaborative components through the accompanying workshop.

Materials and Methods

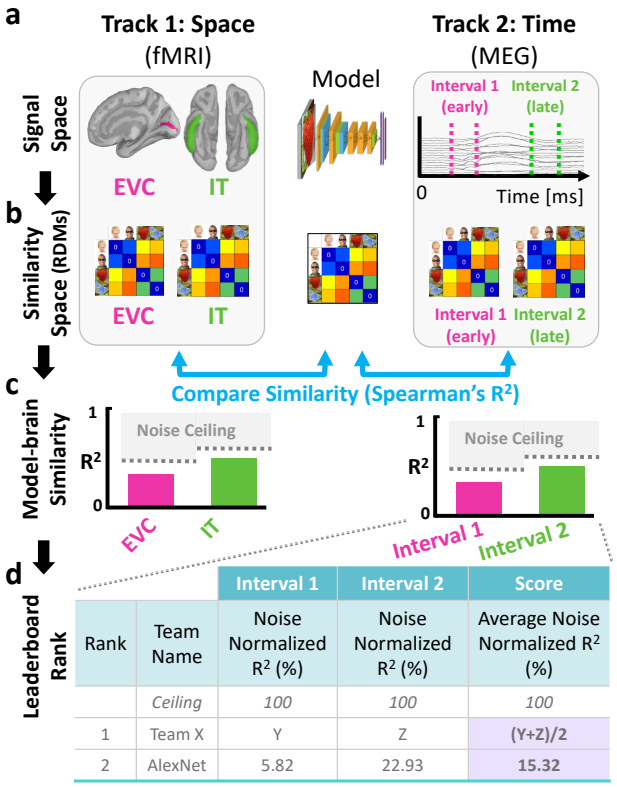
The target of the 2019 challenge is to account for activity in the human visual brain responsible for object recognition. This is the so-called ventral visual stream (Grill-Spector & Malach, 2004), a hierarchically ordered set of brain regions in which neural activity unfolds across regions in space and time when human beings see an object. It starts with early visual cortex (EVC) and continues in inferior temporal (IT) cortex. Neurons in EVC respond preferentially to simple visual features such as oriented edges, whereas neurons in IT respond to more complex and larger features such as object parts. Consistent with their position in the processing hierarchy, neurons in EVC have been found to respond to visual stimulation earlier in time than neurons in IT. Stages of brain processing can thus be identified both in space (different regions) and in time (early and late). Correspondingly we have two challenge tracks.

Track 1 aims to account for brain data in space, providing data from the start and later point of the ventral visual stream: early visual cortex (EVC) and inferior temporal cortex (IT), respectively (Fig. 1a). We provide brain data measured with functional magnetic resonance imaging (fMRI[1]), a technique with high spatial resolution (millimeters) that measures blood flow changes associated with neural activity.

Track 2 aims to account for brain data in time, providing data recorded early and late in visual processing (Fig. 1a). For this we provide brain data measured with magnetoencephalography (MEG) at time points identified to correspond to processing in EVC and IT. MEG is a technique with very high temporal resolution (milliseconds) that measures the magnetic fields accompanying electrical activity in the brain.

Comparison metric from brain activity and models to challenge score. Comparing human brains and models is challenging because of the numerous differences between them (e.g. in-silico vs. biological, number of units). Different approaches have been proposed (Diedrichsen & Kriegeskorte, 2017; Wu, David, & Gallant, 2006), and here we make use of a technique called representational similarity analysis (RSA) (Kriegeskorte, 2008; Kriegeskorte & Kievit, 2013). RSA

[1]For more details on MEG and fMRI see `http://algonauts.csail.mit.edu/fmri_and_meg.html`



Rank	Team Name	Interval 1 Noise Normalized R^2 (%)	Interval 2 Noise Normalized R^2 (%)	Score Average Noise Normalized R^2 (%)
	Ceiling	*100*	*100*	*100*
1	Team X	Y	Z	(Y+Z)/2
2	AlexNet	5.82	22.93	**15.32**

Figure 1: *Procedure of the Algonauts 2019 edition challenge.* **a)** In two tracks, the goal is to account for human brain activity measured during object perception in space and time. **b)** RSA makes models and brain activity comparable, yielding **c)** percent variance explained relative to the noisiness of the data. **d)** Models are ranked in a leaderboard (i.e. for Track 2).

has low computational demands and is straightforward to implement. The idea behind RSA is that models and brains are similar if they treat the same images as similar (or equivalently dissimilar). RSA is a two-step procedure. In a first step (Fig. 1b), we abstract from the incommensurate signal spaces into similarity space by calculating pairwise dissimilarities between signals for all conditions (images) and order them in so-called representational dissimilarity matrices (RDMs) indexed in rows and columns by the conditions compared. RDMs for the different signal spaces have the same dimensions and are directly comparable. We relate RDMs in a second step (Fig. 1c) by calculating their similarity (Spearman R). Finally, we square the result to R^2 to indicate the amount of variance explained, and display results in the leaderboard (Fig. 1d).

Noise ceiling. The noise ceiling is the expected RDM correlation achieved by the (unknown) ideal model, given the noise in the data. The noise ceiling is computed by the assumption that the subject-averaged RDM is the best estimate of the ideal model RDM, i.e. by averaging the correlation of each subject's RDM with the subject-averaged RDM. We use the

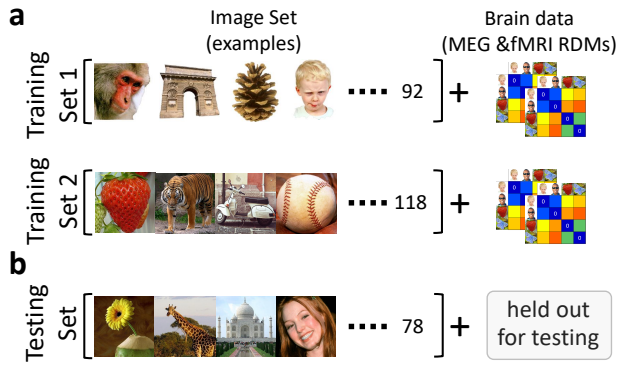


Figure 2: *Training and Testing Material.* **a)** There are two sets of training data, each consisting of an image set and brain activity in RDM format (for fMRI and MEG). Training set 1 has 92 silhouette object images, and training set 2 has 118 object images with natural backgrounds. **b)** Testing data consists of 78 images of objects on natural backgrounds. Associated brain data is held back and used to evaluate models online for the leaderboard.

noise ceiling to normalize R^2 values to noise-normalized variance explained. Thus, any model can explain from 0 to 100% of the explainable variance.

Training Data. Participants can submit their models out of the box to determine how well they predict brain activity in each track. We also provide training data that can help optimizing models for predicting brain data. We provide two sets of training data published previously (Fig. 2a) (Cichy, Pantazis, & Oliva, 2014; Cichy, Khosla, Pantazis, Torralba, & Oliva, 2016). Each set consists of a set of images (92 silhouette object images and 118 images of objects on natural background), and brain data recorded with fMRI (EVC and IT) and MEG (early and late in time) in response to viewing those images (by 15 participants). Participants differ across training sets but are the same across imaging modalities (MEG and fMRI).

Testing Data and Procedure. The testing set consists of 78 images and the respective brain activity recorded with fMRI and MEG (Fig. 2b). Participants in the challenge receive only the images, and the brain data is held back. On the basis of the image test set participants calculate model RDMs as predictions of human brain activity. Participants submit the RDMs which are compared against the held-out brain data using RSA as described above. This results in a challenge score and determines the relative place in the leaderboard.

Rules. To encourage broad participation the challenge consists of a simple submission process. Participants can use any model trained on any type of data, however we explicitly forbid the use of human brain responses to the test image set. We request participants to submit a short report to a preprint server describing their final submitted model.

Development Kit. The development kit contains the aforementioned training and testing data. In addition, we provide example extraction code (matlab and python) to extract activation values from models into RDMs and evaluation code that compares model RDMs with brain RDMs, calculating the noise-normalized score for a model.

Baseline Model. Deep neural networks trained on object classification are currently the model class best performing in predicting visual brain activity. We used *AlexNet* (Krizhevsky, Sutskever, & Hinton, 2012) as an example often used in neuroscientific studies as baseline model. *AlexNet* is a feedforward deep neural network, trained on object categorization, with 5 convolutional and 3 fully connected layers. In Track 1 (fMRI), *AlexNet* accounts for 6.58% (layer 2) and 8.22% (layer 8) of noise-normalized variance in EVC and IT. In track 2 (MEG), it accounts for 5.82% (layer 2) and 22.93% (layer 4) noise-normalized variance in early and late visual processing.

Discussion

Challenges as scientific instruments in cognitive science. Open challenges at the intersection of natural and artificial intelligence sciences hold promise for both sides. The natural intelligence sciences, in particular neuroscience and psychology, might benefit in two ways. For one, open challenges provide the incentive structure to promote and ensure transparency and openness. These are values recognized to promote replicability of results (Nosek et al., 2015; Poldrack et al., 2017). Second, challenges provide a clear and quantitatively concise metric for success. They can thus play an important role in guiding research by differentiating between theories: predictive success is a necessary property of a good explanatory model (Kriegeskorte, 2015). The sciences creating artificial intelligence in turn might benefit, too, in several ways. Biological systems can provide insight into how a cognitive problem might be solved mechanistically. More specifically, neuroscience can provide constraints on the infinite number of free parameters when engineering a model from scratch.

Prediction vs. explanation. Challenges like the Algonauts Project provide one measure of success: predictive power. Having an artifact that even perfectly predicts a phenomenon does not by itself explain the phenomenon. However, prediction and explanation are related goals (Cichy & Kaiser, 2019). For one, successful explanations ultimately must also provide successful predictions (Breiman, 2001; Yarkoni & Westfall, 2017). Second, the ordering of models on a challenge benchmark can help scientist to concentrate future research efforts in creating explanations based on the most successful models. Further, bringing success rate in connection with the models' properties can reveal what it is about those models that is responsible for the success. It can thus generate hypotheses and guide the next engineering steps.

Limitations of the current approach. Constitutive for a challenge are the choice of a particular data set and analysis steps. We readily assert that we could have structured the

challenge differently (e.g. which data to provide, in which format, how to relate brain data and models). The choices we made were motivated by providing a low threshold to participation and a low computational load. Future challenges that make use of other data sets (e.g. large-scale) will invite a different type of data format and analytic treatment. We will invite an open discussion on those issues during the workshop.

The future of the project. We hope that the 2019 edition of the Algonauts Project will inspire other researchers to initiate open challenges and collaborate with the Algonauts Project. We see potential in tackling problems that become increasingly interesting to both natural and artificial intelligence communities. In the context of perception, future challenges might put the focus on action recognition or involve other sensory modalities such as audition or the tactile sense, or focus on other cognitive functions such as learning and memory.

Acknowledgments

This research was funded by DFG (CI-241/1-1 CI-241/1-3) and an ERC grant (ERC-2018-StG 803370) to R.M.C; NSF award (1532591) in Neural and Cognitive Systems and the Vannevar Bush Faculty Fellowship program funded by the ONR (N00014-16-1-3116) to A.O. We thank our sponsors: the MIT Quest for Intelligence and the MIT-IBM Watson AI Lab.

References

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. *Science*, *364*(6439), eaav9436.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3), 199-231.

Cichy, R. M., & Kaiser, D. (2019). Deep neural networks as scientific models. trends in cognitive sciences. *Trends in Cognitive Sciences*, *23*(4), 305-317.

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*, 27755.

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455-462.

Diedrichsen, J., & Kriegeskorte, N. (2017). Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology*, *13*(4), e1005508.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*(1), 98–136.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649–677.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258.

Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Zitnick, C. L., & Girshick, R. (2017). CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *CVPR*.

Kriegeskorte, N. (2008). Representational similarity analysis connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4.

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, *1*(1), 417-446.

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*, 401-412.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 25* (p. 1097-1105).

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422-1425.

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., ... Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115-126.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, *115*(3), 211-252.

Schrimpf, M., Kubilius, J., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., ... DiCarlo, J. J. (2018). Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, 407007.

Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. *CoRR*, *abs/1804.07461*.

Wu, M. C.-K., David, S. V., & Gallant, J. L. (2006). Complete functional characterization of sensory neurons by system identification. *Annual Review of Neuroscience*, *29*, 477–505.

Yamins, D. L. K., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*, 356–365.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100-1122.

Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2019). Semantic Understanding of Scenes Through the ADE20K Dataset. *International Journal of Computer Vision (IJCV)*, *127*(3), 302–321.