Digital Twins for the Brain

Digital Twins for the Brain

F. H. Sinz, M. Bashiri, S. Shrinivasan, K. Lurz, K. Willeke, P. Elumalai, C. Blessing, P. Pierzchlewicz, A. Nix*, Campus Institute Data Science, University Göttingen*

In Short

- Despite decades of seminal research, very little is known about how neural activity in our brains contributes to understanding sensory signals and plan appropriate actions in a complex environment.
- Recent progress in experimental technology to record large number of neurons as well as major progress in deep machine learning have lead to the development of "digital twins" of large populations of neurons.
- These digital twins are machine learning models that can be learned from observations of brain activity and can faithfully predict neural activity to novel sensory inputs.
- This allows computational neuroscientists to generate new hypotheses about brain function by performing analyses and experiments on the digital twins in the computer. These hypotheses can later be verified in real experiments.
- Several recent publications have demonstrated that new insights into brain function can be gained from these models. However, the development of these model is still at the beginning.
- The proposed project pursues several avenues to extend and improve digital twins of the brain to improve our functional understanding of the brain.
- The sub-projects are devoted to developing models for neural responses to video, capturing meaningful variability in neural responses, transferring properties between brain networks and artificial networks, developing predictive models for higher visual areas under free viewing, and developing novel graph based models for brain areas that are involved in sensor-motor integration and motor execution.

How does our brain represent the complex world around us with neural activity and how does it use neural activity to plan and execute actions interacting with the environment? Even though the brain has been studied extensively for more than 60 years, these questions remain mostly elusive. However, recent advances in neurophysiological recording techniques have yielded a wealth of data, in particular large population recordings of neuronal responses to natural stimuli or during free behavior. At the same time artificial intelligence – in particular deep learning – has enabled a quantum leap in our ability to analyze neural dynamics and behavior at scale. In particular, these methods allow us to learn faithful machine learning models jointly on sensory input, behavior, and neural activity to predict neural activity in novel situations. We call these models *digital twins*. These digital twins have already yielded multiple new insights into the visual system of mice and monkeys [1–3] such as novel complex selectivities to visual stimuli, image based neural control, and novel connections between rapid eye movements and color selectivity of neurons in visual cortex. These findings impressively demonstrate the potential of learning digital twins from neural recordings and analyzing them generate specific testable predictions about the brain.

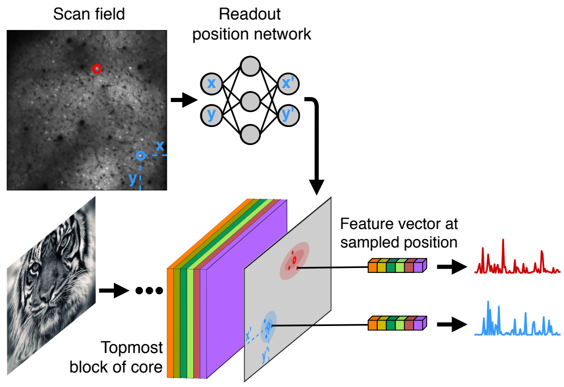

Figure 1: Typical model architecture for a deep learning based digital twin model (for vision). *A common feature set (bottom middle) is computed from the input (here image) and provides the basis for predicting single neurons (bottom right). The readout for each neuron can be informed by additional sources of information, such as the location within the brain (top left).*

Nevertheless, research into data-driven digital twins for neural activity is just at the beginning and offers several new research directions and questions.

Specifically

- Most digital twin models for the visual system currently work on images. However, the visual system operates on a continuous stream of visual input. This creates the need for video based models posing new technical and conceptual challenges.
- Current digital twin models make a single predic-

tion for a given sensory/behavioral input. However, neural activity is variable, i.e. neurons do not respond the same way every time even to the same input. In particular, statistical dependencies between neural activity are important fingerprints of perceptual decision making and neural coding. This creates the need for digital twin models that predict entire distributions for population responses instead of single values.

- In contrast to many computer vision models, the visual system is amazingly robust against distortions in the sensory input. However, to date it is unclear what features of neural activity are responsible for this robustness. One approach to understand this phenomenon is to transfer robustness properties between brains and neural networks. Early proofs of concept for this approach have recently been published [4,5]. However, the reliability of the transfer of robustness as well as its interpretability is still in its early stages.
- Most digital twin models for viewing focus on early visual areas during artificial viewing conditions (i.e. gaze fixation). Models for higher visual areas of the brain during natural (free) viewing conditions are still missing.
- All digital twin models for the brain currently focus on sensory systems, modeling the relation between sensory input and neural responses. Models for brain areas that integrate sensory input and motor planning are lacking.

The current project develops better digital twin models for the brain based on deep learning and large scale neuronal recordings. Insights from these models will yield a better understanding how the computations in our brains help us to interpret the outside world.

WWW

https://sinzlab.org

More Information

[1] Pouya Bashivan, Kohitij Kar, and James J DiCarlo, *Neural population control via deep image synthesis*, Science, **364**, no. 6439, 2019.

[2] E Y Walker, **Sinz, F H**, E Cobos, T Muhammad, E Froudarakis, P G Fahey, A S Ecker, J Reimer, X Pitkow, and A S Tolias, *Inception loops discover what excites neurons most using deep predictive models*, Nat. Neurosci., 2019.

[3] Katrin Franke, **Konstantin F. Willeke**, Kayla Ponder, Mario Galdamez, Na Zhou, Taliah Muhammad, Saumil Patel, Emmanouil Froudarakis, Jacob Reimer, **Fabian H. Sinz**, and Andreas S. Tolias, *State-dependent pupil dilation rapidly shifts visual feature selectivity*, Nature, **610**, no. 7930, 128–134, Sept. 2022.

[4] Zhe Li, Wieland Brendel, Edgar Walker, Erick Cobos, Taliah Muhammad, Jacob Reimer, Matthias Bethge, **Fabian Sinz**, Zachary Pitkow, and Andreas Tolias, *Learning from brains how to regularize machines*, Advances in neural information processing systems, **32**, 2019.

[5] **Shahd Safarani**, **Arne Nix**, **Konstantin Willeke**, Santiago Cadena, Kelli Restivo, George Denfield, Andreas Tolias, and **Fabian Sinz**, *Towards robust vision by multi-task learning on monkey visual cortex*, Adv. Neural Inf. Process. Syst., **34**, 739–751, Dec. 2021.

[6] **K Willeke**, P Fahey, **M Bashiri**, L Pede, M Burg, **C Blessing**, S Cadena, Z Ding, **K Lurz**, K Ponder, T Muhammad, S Patel, A Ecker, A Tolias, and **F Sinz**, *The Sensorium competition on predicting large-scale mouse primary visual cortex activity*, arXiv preprint arXiv:2206.08666, 2022.

[7] **Mohammad Bashiri**, Edgar Walker, **Konstantin-Klemens Lurz**, Akshay Jagadish, Taliah Muhammad, Zhiwei Ding, Zhuokun Ding, Andreas Tolias, and **Fabian Sinz**, *A flow-based latent state generative model of neural population responses to natural images*, Advances in Neural Information Processing Systems, **34**, 15801–15815, 2021.

[8] **K Lurz**, **M Bashiri**, **K Willeke**, A Jagadish, E Wang, E Walker, S Cadena, T Muhammad, E Cobos, A Tolias, and **F Sinz**, *Generalization in data-driven models of primary visual cortex*, ICLR, 2021.

[9] Santiago A. Cadena, **Konstantin F. Willeke**, Kelli Restivo, George Denfield, **Fabian H. Sinz**, Matthias Bethge, Andreas S. Tolias, and Alexander S. Ecker, *Diverse task-driven modeling of macaque V4 reveals functional specialization towards semantic tasks*, May 2022.

Funding

intramural, BMBF FKZ 01GQ2107, DFG CRC 1456

DFG Subject Area

206-03