Network-Level Examination of Correspondence between Human-Brain and ANN

Subjects: Biology

Contributor: Trung Quang Pham , Teppei Matsui , Junichi Chikazoe

Artificial neural networks (ANNs) that are heavily inspired by the human brain now achieve human-level performance across multiple task domains. ANNs have thus drawn attention in neuroscience, raising the possibility of providing a framework for understanding the information encoded in the human brain. However, the correspondence between ANNs and the brain cannot be measured directly. They differ in outputs and substrates, neurons vastly outnumber their ANN analogs (i.e., nodes), and the key algorithm responsible for most of modern ANN training (i.e., backpropagation) is likely absent from the brain. Neuroscientists have thus taken a variety of approaches to examine the similarity between the brain and ANNs at multiple levels of their information hierarchy.

human brain | artificial neural networks | correspondence | network-level | sequential alignment | gradient

1. Introduction

Understanding the information processing of the human brain is one of the biggest challenges for neuroscientists. In recent years, the artificial neural network (ANN) has become a powerful tool, performing at or better than human levels in several domains, including image classification (AlexNet [1]), conversation (ChatGPT [2], LaMDA [3]), games (Go [4], Starcraft II [5], and biological science (i.e., protein folding [6][7]). Growing interest has thus emerged as to the degree to which the information processing of ANNs can inform that which occurs in the brain.

Studies have shown that the processing of human perceptions is hierarchically distributed over the brain ([8][9][10][11][12][13]). In the visual domain, for instance, the V2 neuron appears to be sensitive to naturalistic texture stimuli [14], V4 neurons increase selectivity for the conjunction of features representing the surface shape (i.e., non-Cartesian gratings [15]), and IT neurons show stimulus selectivity, sensitive to the specific combinations of features, i.e., the face [16]. A similar hierarchy can be found in language processing [17][18][19], music processing [20][21][22], and tactile processing [23][24][25]. Taking a broader perspective, converging evidence alerts us to the brain's global hierarchical organization beyond a collection of independent sensory hierarchies. At the cellular level, Murray et al. [11] found different decay rates of a single-unit activity in early sensory areas and distant cortical regions. Whereas sensory areas may need to respond more rapidly to environmental changes to reflect faster decay rates, regions involved in more complex, integrative cognitive tasks exhibit longer decay rates, suggesting a hierarchical ordering of a measure as intrinsic as single-neuron spiking timescales. Neuroimaging evidence of a global, sensorimotor-to-

transmodal gradient supported this hierarchy of temporal dynamics, as well as other converging evidence such as increasing intracortical myelination, functional connectivity, and semantic processing along the gradient [26][27].

Given the intrinsic hierarchical architecture of ANNs, it becomes natural to wonder if they can capture the information processing that occurs in the human brain, thus serving as a framework for its understanding [28][29]. While both the human brain and the ANN are “black-boxes”, the latter is easier to customize and analyze. ANNs may provide a useful model for understanding the former, akin to how the atomic model can usefully convey the interaction between protons and electrons. As statistician George Box once said, “All the models are wrong, but some are useful”.

Relationships between the human brain and modern ANNs have been found since the early stages of ANN development. Studies have revealed the similarity between cognitive processing, such as vision and audition, and the hidden layers of ANNs [28][30][31][32][33][34][35][36]. The similarity was not limited to well-known “supervised” learning, but also “unsupervised” learning and “self-supervised” learning [37][38]. The growing number of studies in this area offers promise toward improving our understanding of the brain, as ANNs rapidly grow in sophistication and performance across problem domains. However, comparing ANNs to the brain to arrive at meaningful references is not a straightforward process. ANNs are inspired by the brain but are not replicas. Not only do they differ in substrate, but there are also vastly fewer ANN nodes than neurons. The principal algorithm that discovers the hierarchical circuitry of most modern ANNs is unlikely to exist in the brain [39].

Conventionally, evaluation of the similarity between the ANN and the human brain has been based on their performance in “intelligent” tasks (e.g., object detection, object classification, text generation, image generation, game playing, etc.). However, just this high-level comparison is inadequate for determining whether the ANN under the hood is undergoing comparable information processing to the brain. Neuroscientists have thus taken a variety of indirect approaches to evaluate the correspondence between ANNs and the brain.

Network-level correspondence examines the overall information flow inside an ANN to a comparable network in the brain, such as the hierarchical representations across a single modality, or the multimodal integrative network across the whole brain. In relation to the layer-level correspondence, a straightforward approach is to quantify the alignment between the sequence of ANN layers and sequential processing expected in the brain. For example, one can compute the correlation between the two and count the nodes of layer that are most associated with each ROI in order to test if there is a shifting of distribution from low-level to high-level cortices. Given the intrinsic feed-forward characteristics of the ANN, a sequential alignment between the brain and the ANN would indicate a hierarchical network-level correspondence.

2. Sequential Alignment Approach

Early examinations of network-level correspondence have been conducted for sensory networks (visual network, auditory network) due to their interpretability [40]. For the visual network, the distribution of a model-explained variance of neural activity from Yamins et al. [30] shows a clear shift from V1 to IT as the layers changed from the

first to the top layer. A similar correspondence across layers was found for information extracted along the visual ventral pathway [41] as well as the dorsal pathway [42]. A recent study from Mineault et al. [43] confirmed a similar correspondence between the ANN and the visual dorsal pathway in non-human primates. Using the ANN decoding approach, Horikawa and Kamitani showed that dreaming recruits visual feature representations that correlated hierarchically across the visual system [33]. For the auditory network, Kell et al. [36] found that an ANN trained on speech and music correlated with the auditory processing hierarchy in the brain with different layers processing different aspects of sound. In another study using ANNs trained to classify music genres, Guclu et al. [44] showed a representational gradient along the superior temporal gyrus, where anterior regions were associated with shallower layers and posterior regions with deeper layers.

Evaluating large-scale networks, such as across modalities or the global brain hierarchy, poses an additional problem. For instance, the actual hierarchical correspondence between the human auditory system and visual ANNs remains unclear, as other studies have raised the suggestion of parallel organization [45]. Spatial locations like brain coordinates may provide an intuitive correspondence but not concrete evidence of the brain's structural–functional organization. For instance, not all of the many functional networks of the brain may adhere to a clear posterior-to-anterior hierarchy.

3. Gradient-Based Approach

Brain–ANN correspondence at the network level should also account for the sequence of chosen ROIs and the design of the ANN, such as the features that each node processes, whether they are processed sequentially or in parallel, how multiple modalities are integrated, and so on. A promising approach here is to use the principal gradient (PG) [26] as a reference. The PG is a global axis of brain organization that accounts for the highest variability in human resting-state functional connectivity. Its arrangement begins with multiple satellites of unimodal sensory information that converge transmodally and integrate with the default mode network (DMN). A meta-analysis using the NeuroSynth database [46] has reinforced the relationship between cognitive function and position along the PG, with sensory perception and motion exhibiting lower positions, and higher-order, abstract processes such as emotion and social recognition exhibiting higher positions [26]. The implications of PG on the hierarchical organization of functionality are further supported by clinical evidence, such as the compression of the principal motor-to-supramodal gradient in patients with schizophrenia (96 patients with schizophrenia vs. 120 healthy controls) [47] and the decrease in PG values in a neurodegenerative condition like Alzheimer's disease [48].

For evaluating correspondence at the global brain level, the PG provides an independent, quantifiable metric of its hierarchy, anywhere from sensorimotor and transmodal to higher cognitive and affective information processing. Examining how subjective value emerges in the brain, ANNs individually trained to output subjective value from visual input have been shown to hierarchically correspond to the PG in the brains of those same individuals experiencing a similar value during fMRI [49], whereas Nonaka et al. [31] showed that most ANNs tend to have similar representations to the lower portion of the higher visual cortex (divided by the PG), but not the middle and higher ones, suggesting that findings of detailed correspondence in local areas could be more complex than simply an adherence to a global hierarchy.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the NIPS'12: 25th International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.
2. OpenAI. GPT-4 Technical Report. arXiv 2023, arXiv:2303.08774.
3. Thoppilan, R.; Freitas, D.D.; Hall, J.; Shazeer, N.; Kulshreshtha, A.; Cheng, H.T.; Jin, A.; Bos, T.; Baker, L.; Du, Y.; et al. LaMDA: Language Models for Dialog Applications. arXiv 2022, arXiv:2201.08239.
4. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016, 529, 484–489.
5. Vinyals, O.; Babuschkin, I.; Czarnecki, W.M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D.H.; Powell, R.; Ewalds, T.; Georgiev, P.; et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. Nature 2019, 575, 350–354.
6. Sudha, P.; Ramyachitra, D.; Manikandan, P. Enhanced Artificial Neural Network for Protein Fold Recognition and Structural Class Prediction. Gene Rep. 2018, 12, 261–275.
7. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. Nature 2021, 596, 583–589.
8. Kiebel, S.J.; Daunizeau, J.; Friston, K.J. A Hierarchy of Time-Scales and the Brain. PLoS Comput. Biol. 2018, 4, e1000209.
9. Hasson, U.; Chen, J.; Honey, C.J. Hierarchical process memory: Memory as an integral component of information processing. Trends Cogn. Sci. 2015, 19, 304–313.
10. Hasson, U.; Yang, E.; Vallines, I.; Heeger, D.J.; Rubin, N. A Hierarchy of Temporal Receptive Windows in Human Cortex. J. Neurosci. 2008, 28, 2539–2550.
11. Murray, J.D.; Bernacchia, A.; Freedman, D.J.; Romo, R.; Wallis, J.D.; Cai, X.; Padoa-Schioppa, C.; Pasternak, T.; Seo, H.; Lee, D.; et al. A hierarchy of intrinsic timescales across primate cortex. Nat. Neurosci. 2014, 17, 1661–1663.
12. Burt, J.B.; Demirtas, M.; Eckner, W.J.; Navejar, N.M.; Ji, J.L.; Martin, W.J.; Bernacchia, A.; Anticevic, A.; Murray, J.D. Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. Nat. Neurosci. 2018, 21, 1251–1259.

13. Demirtaş, M.; Burt, J.B.; Helmer, M.; Ji, J.L.; Adkinson, B.D.; Glasser, M.F.; Van Essen, D.C.; Sotiropoulos, S.N.; Anticevic, A.; Murray, J.D. Hierarchical Heterogeneity across Human Cortex Shapes Large-Scale Neural Dynamics. Neuron 2019, 101, 1181–1194.e13.
14. Freeman, J.; Ziemba, C.M.; Heeger, D.J.; Simoncelli, E.P.; Movshon, J.A. A functional and perceptual signature of the second visual area in primates. Nat. Neurosci. 2013, 16, 974–981.
15. Gallant, J.L.; Connor, C.E.; Rakshit, S.; Lewis, J.W.; Van Essen, D.C. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. J. Neurophysiol. 1996, 76, 2718–2739.
16. Gross, C.G.; Rocha-Miranda, C.E.; Bender, D.B. Visual properties of neurons in inferotemporal cortex of the Macaque. J. Neurophysiol. 1972, 35, 96–111.
17. Moro, A. Dynamic Antisymmetry; MIT Press: Cambridge, MA, USA, 2000.
18. Friederici, A.D. The brain basis of language processing: From structure to function. Physiol. Rev. 2011, 91, 1357–1392.
19. Friederici, A.D.; Gierhan, S.M. The language network. Curr. Opin. Neurobiol. 2013, 23, 250–254.
20. Thompson-Schill, S.; Hagoort, P.; Dominey, P.F.; Honing, H.; Koelsch, S.; Ladd, D.R.; Lerdahl, F.; Levinson, S.C.; Steedman, M. Multiple Levels of Structure in Language and Music. In Language, Music, and the Brain: A Mysterious Relationship; MIT Press: Cambridge, MA, USA, 2013.
21. Patel, A.D. Language, music, syntax and the brain. Nat. Neurosci. 2003, 6, 674–681.
22. Koelsch, S.; Rohrmeier, M.; Torrecuso, R.; Jentschke, S. Processing of hierarchical syntactic structure in music. Proc. Natl. Acad. Sci. USA 2013, 110, 15443–15448.
23. Bodegård, A.; Geyer, S.; Grefkes, C.; Zilles, K.; Roland, P.E. Hierarchical Processing of Tactile Shape in the Human Brain. Neuron 2001, 31, 317–328.
24. Sathian, K. Analysis of haptic information in the cerebral cortex. J. Neurophysiol. 2016, 116, 1795–1806.
25. Bola, Ł.; Matuszewski, J.; Szczepanik, M.; Droździel, D.; Sliwinska, M.W.; Paplińska, M.; Jednoróg, K.; Szwed, M.; Marchewka, A. Functional hierarchy for tactile processing in the visual cortex of sighted adults. NeuroImage 2019, 202, 116084.
26. Margulies, D.S.; Ghosh, S.S.; Goulas, A.; Falkiewicz, M.; Huntenburg, J.M.; Langs, G.; Bezgin, G.; Eickhoff, S.B.; Castellanos, F.X.; Petrides, M.; et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. Proc. Natl. Acad. Sci. USA 2016, 113, 12574–12579.
27. Huntenburg, J.M.; Bazin, P.L.; Margulies, D.S. Large-Scale Gradients in Human Cortical Organization. Trends Cogn. Sci. 2017, 22, 21–31.

28. Yamins, D.L.K.; DiCarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci. 2016, 19, 356–365.
29. Richards, B.; Lillicrap, T.; Beaudoin, P.; Bengio, Y.; Bogacz, R.; Christensen, A.; Clopath, C.; Costa, R.P.; de Berker, A.; Ganguli, S.; et al. A deep learning framework for neuroscience. Nat. Neurosci. 2019, 22, 1761–1770.
30. Yamins, D.L.K.; Hong, H.; Cadieu, C.F.; Solomon, E.A.; Seibert, D.; DiCarlo, J.J. Perfomance-optimized hierarchical models predict neural responses in higher visual cortex. Proc. Natl. Acad. Sci. USA 2014, 111, 8619–8624.
31. Nonaka, S.; Majima, K.; Aoki, S.C.; Kamitani, Y. Brain hierarchy score: Which deep neural networks are hierarchically brain-like? iScience 2021, 24, 103013.
32. Bashivan, P.; Kar, K.; DiCarlo, J.J. Neural population control via deep image synthesis. Science 2019, 364, e453.
33. Horikawa, T.; Tamaki, M.; Miyawaki, Y.; Kamitani, Y. Neural Decoding of Visual Imagery During Sleep. Science 2013, 340, 639–642.
34. Horikawa, T.; Kamitani, Y. Hierarchical Neural Representation of Dreamed Objects Revealed by Brain Decoding with Deep Neural Network Features. Front. Comput. Neurosci. 2017, 11, e4.
35. Horikawa, T.; Kamitani, Y. Generic decoding of seen and imagined objects using hierarchical visual features. Nat. Commun. 2017, 8, 15037.
36. Kell, A.J.E.; Yamins, D.L.K.; Shook, E.N.; Norman-Haignere, S.V.; McDermott, J.H. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. Neuron 2018, 98, 630–644.e16.
37. Zhuang, C.; Yan, S.; Nayebi, A.; Schrimpf, M.; Frank, M.C.; DiCarlo, J.J.; Yamins, D.L. Unsupervised neural network models of the ventral visual stream. Proc. Natl. Acad. Sci. USA 2023, 118, e2014196118.
38. Konkle, T.; Alvarez, G.A. A self-supervised domain-general learning framework for human ventral stream representation. Nat. Commun. 2022, 13, 491.
39. Lillicrap, T.P.; Santoro, A.; Marris, L.; Akerman, C.J.; Hinton, G. Backpropagation and the brain. Nat. Rev. Neurosci. 2020, 21, 335–346.
40. Kell, A.J.; McDermott, J.H. Deep neural network models of sensory systems: Windows onto the role of task constraints. Curr. Opin. Neurobiol. 2019, 55, 121–132.
41. Güçlü, U.; van Gerven, M.A.J. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J. Neurosci. 2015, 35, 10005–10014.

42. Güçlü, U.; van Gerven, M.A.J. Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. NeuroImage 2017, 145, 329–336.
43. Mineault, P.; Bakhtiari, S.; Richards, B.; Pack, C. Your head is there to move you around: Goal-driven models of the primate dorsal pathway. bioRxiv 2021.
44. Güçlü, U.; Thielen, J.; Hanke, M.; van Gerven, M.A.J. Brains on Beats. arXiv 2016, arXiv:1606.02627.
45. Hamilton, L.S.; Oganian, Y.; Hall, J.; Chang, E.F. Parallel and distributed encoding of speech across human auditory cortex. Cell 2021, 184, 4626–4639.e13.
46. Yarkoni, T.; Poldrack, R.A.; Nichols, T.E.; Van Essen, D.C.; Wager, T.D. Large-scale automated synthesis of human functional neuroimaging data. Nat. Methods. 2011, 8, 665–670.
47. Dong, D.; Luo, C.; Guell, X.; Wang, Y.; He, H.; Duan, M.; Eickhoff, S.B.; Yao, D. Compression of Cerebellar Functional Gradients in Schizophrenia. Schizophr. Bull. 2020, 46, 1282–1295.
48. Hu, Q.; Li, Y.; Wu, Y.; Lin, X.; Zhao, X. Brain network hierarchy reorganization in Alzheimer's disease: A resting-state functional magnetic resonance imaging study. Hum. Brain Mapp. 2022, 43, 3498–3507.
49. Pham, T.Q.; Yoshimoto, T.; Niwa, H.; Takahashi, H.K.; Uchiyama, R.; Matsui, T.; Anderson, A.K.; Sadato, N.; Chikazoe, J. Vision-to-value transformations in artificial neural networks and human brain. bioRxiv 2021.