Relating Representational Geometry to Cortical Geometry in the Visual Cortex

Francisco Acosta
Department of Physics
UC Santa Barbara
facosta@ucsb.edu

Colin Conwell
Department of Cognitive Science
Johns Hopkins University

Sophia Sanborn
Electrical & Computer Engineering
UC Santa Barbara

David Klindt
SLAC
Stanford University

Nina Miolane
Electrical & Computer Engineering
UC Santa Barbara

Abstract

A fundamental principle of neural representation is to *minimize wiring length* by spatially organizing neurons according to the frequency of their communication [Sterling and Laughlin, 2015]. A consequence is that nearby regions of the brain tend to represent similar content. This has been explored in the context of the visual cortex in recent works [Doshi and Konkle, 2023, Tong et al., 2023]. Here, we use the notion of *cortical distance* as a baseline to ground, evaluate, and interpret measures of *representational distance*. We compare several popular methods—both second-order methods (Representational Similarity Analysis, Centered Kernel Alignment) and first-order methods (Shape Metrics)—and calculate how well the representational distance reflects 2D anatomical distance along the visual cortex (the *anatomical stress score*). We evaluate these metrics on a large-scale fMRI dataset of human ventral visual cortex [Allen et al., 2022b], and observe that the 3 types of Shape Metrics produce representational-anatomical stress scores with the smallest variance across subjects, (Z score = -1.5), which suggests that first-order representational scores quantify the relationship between representational and cortical geometry in a way that is more invariant across different subjects. Our work establishes a criterion with which to compare methods for quantifying representational similarity with implications for studying the anatomical organization of high-level ventral visual cortex.

1 Introduction

A neural representation comprises the collective activity of a population of neural units, such as the firing frequencies of individual neurons, haemodynamic responses of fMRI voxels, or the activations of nodes in an artificial neural network (ANN); a representation can therefore be viewed as a set of vectors in a high-dimensional neural state space. There is much interest in investigating representations through the lens of geometry. *Representational geometry* [Chung and Abbott, 2021] offers a framework to quantify the operational dissimilarities between different neural systems performing related tasks. The central idea is to compare across brain networks or computational models of brain function like ANNs by comparing the geometry of their neural representations. In

Preprint.

deep learning there is much interest in understanding what makes a good representation in ANNs [Higgins et al., 2022, 2018]. In the biological domain, information is encoded in a population of neurons as patterns of electrical activity. The ability to compare different representations in the brain may yield insights into phenomena like *representational drift* [Schoonover et al., 2021, Driscoll et al., 2022] and allow researchers to quantify how representations form and change during learning [Guo et al., 2020]. It is also widely hypothesized that ANNs and biological neural networks form similar representations from data in order to solve similar problems [Schrimpf et al., 2018, Conwell et al.], or that making them more similar could help in building more robust artificial systems [McClure and Kriegeskorte, 2016, Li et al., 2019]. Moreover, comparing representations and finding the common core has also been linked to disentanglement Duan et al. [2019]. This motivates the development of methods to compare between representations in AI models and in the brain.

A line of recent work lends credence to the idea that there is a direct correspondence between the geometry of a neural representation and the structure of the task variables it encodes [Chaudhuri et al., 2019, Gardner et al., 2022, Nieh et al., 2021]—what has been called a first-order isomorphism between task variables and their representations. Alternatively, we may seek to find a "second-order isomorphism" between the task variables and their representations, i.e. a correspondence between the relations among task variables and the relations among their representations Shepard and Chipman [1970]. Kriegeskorte [2008] operationalize this idea by introducing Representational Similarity Analysis (RSA). Kornblith et al. [2019] introduced Centered Kernel Alignment (CKA), a related method that also exploits the second-order isomorphism idea to quantify dissimilarity between neural network representations. On the other hand, Williams et al. [2021] introduced Generalized Shape Metrics (GSM), which quantify the distance between representations based on their shape in neural state space, and can therefore be related to the first-order isomorphism approach.

In this paper, we evaluate these three methods for quantifying dissimilarities ("distance") between neural representations across several functionally-defined regions of interest (ROIs) in human visual cortex, in a large-scale human fMRI dataset Allen et al. [2022a]. Here, we use 2D anatomical *cortical distance* as a ground-truth measure of representational distance, under the hypothesis that the cortex is spatially organized such that similar information is represented in nearby brain regions [Sterling and Laughlin, 2015, Doshi and Konkle, 2023]. Our approach provides a method for grounding, evaluating, and interpreting different metrics of representational distance. Although anatomical distance provides only a very coarse measure of expected representational distance, we propose it can nonetheless provide insight into the differences between approaches.

2 Background: Approaches in Representational Geometry

Consider a population of n neural units in a neural network and a set of m stimuli: for example, a group of fMRI voxels in a region of the human visual cortex and a set of m natural images. The i^{th} stimulus elicits a neural population response $x_i \in \mathbb{R}^n$, which can be viewed as a point in the n-dimensional *neural state space*. The neural representation of a task in a network is the collection of responses elicited by all possible task-relevant stimuli across all neural units.

2.1 Neural Representations

Neural Representation Matrix X In practice we are restricted to sampling a finite subset of possible stimuli and record from a limited number of neural units in a population. Without making any assumptions about an underlying manifold, we can simply consider the set of points in neural state space and organize the m neural response vectors $x_1, \ldots, x_m$ as rows of a neural representation matrix $X \in \mathbb{R}^{m \times n}$. We call methods that directly use the neural representation matrix to measure representational similarity *first-order* methods, because they compare representations a and b by directly comparing $X^{(a)}$ and $X^{(b)}$. These methods are therefore sensitive to more fine-grained details of the geometric structure of neural responses in the neural state space. Williams et al. [2021] is one such first-order method that compares representations based on their *shape* as described in section 2.2.3.

The Gramian G There are other representational similarity methods like RSA and CKA (described in sections 2.2.1 and 2.2.2 respectively) that instead can be classified as *second-order* methods. These methods compare representations a and b by comparing second-order matrices built from the

representation matrices. One such object is the *Gramian* normalized by number of neural units n, hereinafter simply the Gramian G, computed as:

$$G = \frac{1}{n} X X^T.$$

Second-order methods compare representations a and b by comparing $G^{(a)}$ and $G^{(b)}$.

2.2 Methods for Dissimilarity between Neural Representation

2.2.1 Representational Similarity Analysis (RSA)

RSA is a broad framework to compare representations, and a large number of dissimilarity scores have been proposed Kriegeskorte [2008], Kriegeskorte and Kievit [2013], Diedrichsen and Kriegeskorte [2017], Diedrichsen et al. [2021], Schütt et al. [2023]. All variants of RSA can be broken down into 2 steps: (1) Construction of Representational Dissimilarity Matrices (RDMs), and (2) Comparison of RDMs. Several types of RDMs can be derived directly from the Gramian G. For example, we can compute a Pearson RDM whose ij entry is one minus the Pearson correlation ρ between x_i, x_j from G:

$$\text{RDM}_{ij}^{\text{Pearson}} = 1 - \rho(x_i, x_j) = 1 - \frac{\text{cov}(x_i, x_j)}{\text{std}(x_i)\text{std}(x_j)} = 1 - \frac{G_{ij}}{\sqrt{G_{ii} G_{jj}}}.$$

Once we have computed $\text{RDM}^{(a)}$ and $\text{RDM}^{(b)}$ for two networks a and b, the remaining step is to quantify how different their RDMs are. Because RDMs are symmetric and their diagonals are zero (assuming the neural activity vectors are centered), they are often compared by computing the Pearson correlation or the Spearman rank correlation between their vectorized upper triangular entries. Thus, when one chooses to compare the Pearson RDMs for a and b using Pearson correlation, the dissimilarity score is computed as

$$d_{ab}^{\text{RSA, Pearson-Pearson}} = 1 - \text{Pearson}(\text{vec}(\text{RDM}^{\text{Pearson},(a)}), \text{vec}(\text{RDM}^{(b),\text{Pearson}})),$$

where $\text{vec}(\cdot)$ returns the vectorized upper triangular of each matrix.

2.2.2 Centered Kernel Alignment (CKA)

Kornblith et al. [2019] propose to use centered kernel alignment (CKA), another second-order method to measure dissimilarity between neural representations. It is based on the Hilbert-Schmidt Independence Criterion (HSIC) [Gretton et al., 2005]. Consider representations $X^{(a)}$ and $X^{(b)}$. Let $H = \mathbb{1} - \frac{1}{m}\mathbf{1}\mathbf{1}^T$ be the $m \times m$ centering matrix, $k^{(a)}$ and $k^{(b)}$ be kernels, and $K_{ij}^{(a)} = k^{(a)}(x_i^{(a)}, x_j^{(a)})$. HSIC is given by $\text{HSIC}(K^{(a)}, K^{(b)}) = \frac{1}{(m-1)^2}\text{Tr}(K^{(a)} H K^{(b)} H)$. CKA is computed from HSIC as

$$d_{ab}^{\text{CKA}} = \frac{\text{HSIC}(K^{(a)}, K^{(b)})}{\sqrt{\text{HSIC}(K^{(a)}, K^{(a)})\text{HSIC}(K^{(b)}, K^{(b)})}}.$$

2.2.3 Generalized Shape Metrics (GSM)

Williams et al. [2021] introduce a family of metrics based on the *shape* of representations in neural state space that can be used to compare representations while explicitly accounting for desired symmetries, such as permutation invariance across neurons. This framework defines a *metric space* over neural representation matrices. Crucially, unlike RSA and CKA, the shape metrics are proper metrics and are therefore guaranteed to satisfy the triangle inequality, which ensures the performance of several geometric analyses like k-means clustering[Yianilos, 1970]. Given a symmetry group $\mathfrak{G}$ and a "preprocessing function" ϕ, the distance between networks a and b is given by

$$d_{ab}^{\text{GSM}} = \min_{T \in \mathfrak{G}} \|\phi^a(X^{(a)}) - \phi^b(X^{(b)})T\|_F.$$

One shortcoming with this approach is that it requires all networks being compared to have the same dimensionality. Williams et al. [2021] deal with this by performing PCA projections of all networks to a common dimension.

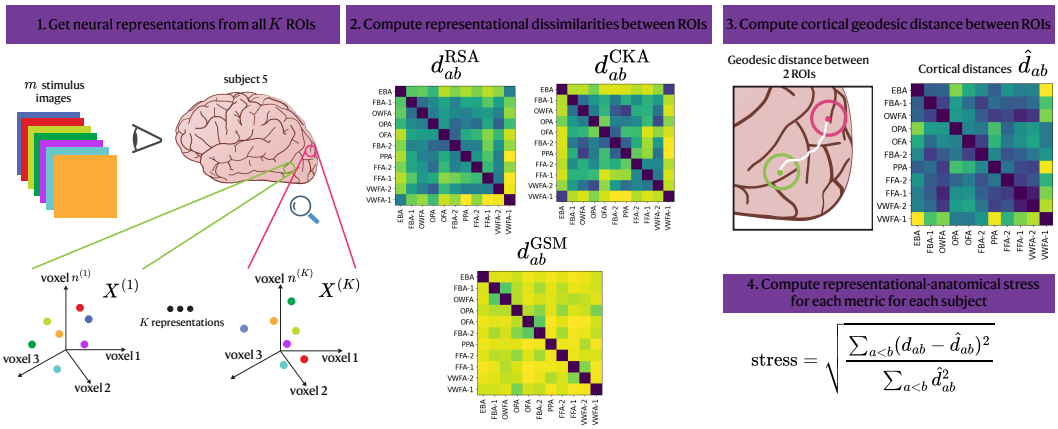

Figure 1: **Methods Overview. (1)** In one subject, we obtain the K neural representations corresponding to K ROIs in the visual cortex. Within a ROI, a stimulus image elicits a particular response in every fMRI voxel; the representation of m stimulus images can be seen as m points in a high dimensional vector space, where each voxel specifies a dimension. **(2)** We compute the representational dissimilarity between every pair of ROIs using the various metrics (RSA, CKA, GSM). This yields a pairwise distance matrix d_{ab} for each metric. **(3)** We compute the geodesic distance along the cortical surface between the centroids of each pair of ROIs, to obtain the anatomical pairwise distance matrix $\hat{d}_{ab}$. **(4)** We calculate the representational-anatomical stress score for each representational metric, which quantifies how well the metric captures the cortical geometry of the visual cortex.

3 Methods & Results: Comparing Approaches to Representational Geometry

3.1 Natural Scenes Dataset

The Natural Scenes Dataset [Allen et al., 2022b] is a large-scale, high-resolution fMRI dataset (7T field strength, 1.6-s TR, $1.8mm^3$ voxel size) that measures the brain responses of 8 subjects to a large sampling of stimuli from the Microsoft Common Objects in Context (COCO) dataset [Lin et al., 2014] across 40 scanning sessions. Subjects in the scanner were tasked only to judge whether or not they had seen the presented images before. In this analysis, we focus on the brain responses to a 1000-image subset of this dataset (the NSD 'Shared1000') that each of 4 subjects (subjects 01, 02, 05, 07) saw for all 3 of the experimental design's intended repetitions. The 3 image repetitions were averaged to yield single voxel-level response values for each stimuli.

3.2 Comparing Representational Geometry to Anatomical Geometry

We implement 14 variants of RSA, where each variant is specified by the choice of one of Euclidean distance, Pearson correlation, Spearman correlation, Mahalanobis distance, and Concordance correlation to construct the RDM and the choice of one of Pearson, Spearman, and Concordance to compare the RDMs.We implement 2 variants of CKA (linear and kernel). For GSM, we implement 3 different metrics corresponding to 3 α values (0,0.5,1). For every selected variant of RSA, CKA, and Shape Metrics we obtain a pairwise representational dissimilarity matrix among 11 ROIs: EBA, OPA, FFA-1, FFA-2, FBA-1, FBA-2, VWFA-1, VWFA-2, OFA, OWFA, and PPA. This matrix encapsulates the "geometry" of these ROIs based on their neural representations. We then seek to compare the result of each method to the anatomical geometry of these ROIs, encapulated by a pairwise anatomical distance matrix. This approach is summarized in Fig. 1

3.2.1 Anatomical Geometry of Human Visual Cortex

For each subject, we have the 3D coordinates associated with every voxel, which we project to the cortical surface using the Python package `pycortex` [Gao et al., 2015]. In this setup, the cortical surface is modeled as a 2D mesh composed of *vertices*; all voxels are projected to corresponding vertices. Because the human cortex is highly folded and curved, with intricate patterns of gyri (ridges) and sulci (grooves), computing distances between different vertices on the cortex is non-trivial. We

must compute distances *along* the cortical surface, which can be modeled as 2-dimensional curved manifold in $\mathbb{R}^3$, and therefore requires finding geodesics between vertices.

3.2.2 Fréchet mean of ROI vertices

To find the distances between ROIs we must first find the centroid of each ROI which, as a notion of the average of points on a curved manifold, is given by the Fréchet mean of all the vertices in the ROI. Take the cortical surface to be a curved manifold $\mathcal{M}$ embedded in $\mathbb{R}^3$; the Fréchet mean of a set of vertices $p_1, p_2, ..., p_N$ is the point $\mu \in \mathcal{M}$ such that

$$\mu = \arg\min_{p \in \mathcal{M}} \sum_{i=1}^{N} d^2(p, p_i) \tag{1}$$

where $d^2(p, p_i)$ is the squared geodesic distance between p and p_i. Once we have found the Fréchet mean or centroid of every functional ROI, we compute the pairwise geodesic distance between all 11 ROI centroids, to obtain a 11×11 anatomical pairwise distance matrix.

3.2.3 Stress score

We can now calculate a measure of "stress" between each representational geometry pairwise-dissimilarity matrix and the "ground-truth" anatomical geometry pairwise-distance matrix:

$$\text{stress} = \sqrt{\frac{\sum_{a<b}(d_{ab} - \hat{d}_{ab})^2}{\sum_{a<b} \hat{d}_{ab}^2}}.$$

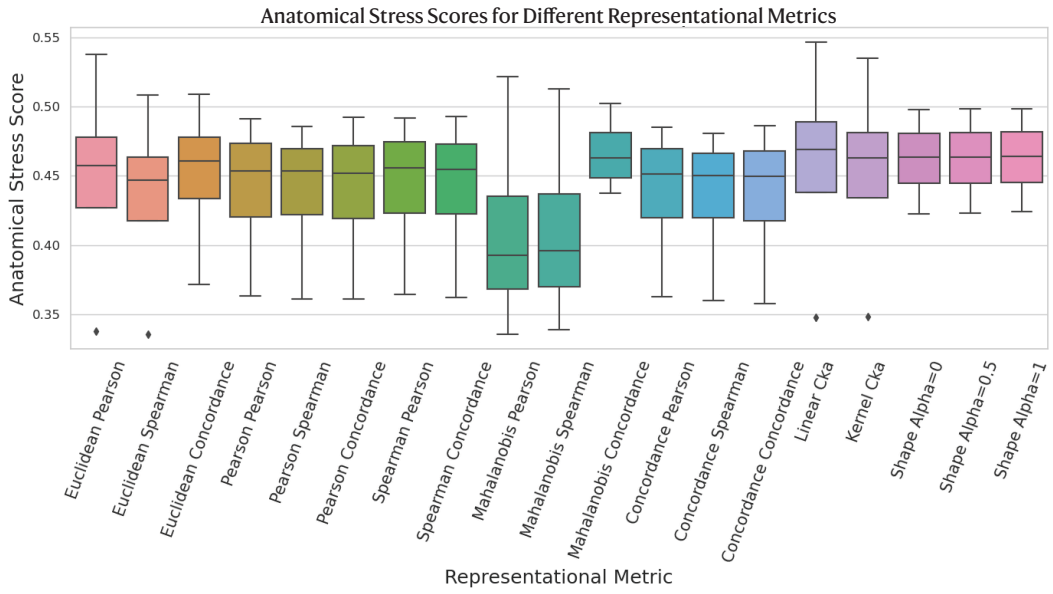

Figure 2: **Anatomical geometry stress scores for various representational dissimilarity metrics.** We implement 14 variants of RSA (defined by the choice of RDM construction and RDM comparison), 2 variants of CKA (linear and kernel), and 3 variants of GSM (3 different values for α, which parameterizes a transformation on the raw representations).

We observe in Fig. 2 that the 3 types of Shape Metrics produce representational-anatomical stress scores with the smallest variance across subjects, (Z score = -1.5), which suggests that first-order representational scores quantify the relationship between representational and cortical geometry in a way that is more invariant across different subjects. Our preliminary results motivate more extensive analysis to understand the relationship between neural representations and the cortical organization of the visual stream, as well as the development of further benchmarks to evaluate the performance of different representational metrics. Understanding the unique advantages of different representational metrics will yield insights into the role of representational geometry in computations in the brain.

Acknowledgments and Disclosure of Funding

We thank Mathilde Papillon and other members of the Geometric Intelligence Lab for feedback on the manuscript. We also thank Jacob Prince for helping to navigate the Natural Scenes Dataset. F. Acosta acknowledges funding from the National Science Foundation, grant 2134241. N. Miolane acknowledges funding from the National Science Foundation, grant 2313150.

References

Emily J. Allen, Ghislain St-Yves, Yihan Wu, Jesse L. Breedlove, Jacob S. Prince, Logan T. Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, J. Benjamin Hutchinson, Thomas Naselaris, and Kendrick Kay. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1):116–126, January 2022a. ISSN 1546-1726. doi: 10.1038/s41593-021-00962-x.

Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, and others. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature neuroscience*, 25(1):116–126, 2022b. doi: 10.1038/s41593-021-00962-x. Publisher: Nature Publishing Group US New York.

Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep. *Nature Neuroscience*, 22(9):1512–1520, September 2019. ISSN 1546-1726. doi: 10.1038/s41593-019-0460-x.

SueYeon Chung and L. F. Abbott. Neural population geometry: An approach for understanding biological and artificial neural networks. *Current Opinion in Neurobiology*, 70:137–144, October 2021. ISSN 0959-4388. doi: 10.1016/j.conb.2021.10.010.

Colin Conwell, Jacob S Prince, George A Alvarez, and Talia Konkle. What can 5.17 billion regression fits tell us about artificial models of the human visual system?

Jörn Diedrichsen and Nikolaus Kriegeskorte. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. *PLOS Computational Biology*, 13(4):e1005508, April 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005508.

Jörn Diedrichsen, Eva Berlot, Marieke Mur, Heiko H. Schütt, Mahdiyar Shahbazi, and Nikolaus Kriegeskorte. Comparing representational geometries using whitened unbiased-distance-matrix similarity, August 2021.

Fenil R. Doshi and Talia Konkle. Cortical topographic motifs emerge in a self-organized map of object space. *Science Advances*, 9(25):eade8187, June 2023. ISSN 2375-2548. doi: 10.1126/sciadv.ade8187.

Laura N. Driscoll, Lea Duncker, and Christopher D. Harvey. Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, 76:102609, October 2022. ISSN 0959-4388. doi: 10.1016/j.conb.2022.102609.

Sunny Duan, Loic Matthey, Andre Saraiva, Nicholas Watters, Christopher P Burgess, Alexander Lerchner, and Irina Higgins. Unsupervised model selection for variational disentangled representation learning. *arXiv preprint arXiv:1905.12614*, 2019.

James S. Gao, Alexander G. Huth, Mark D. Lescroart, and Jack L. Gallant. Pycortex: An interactive surface visualizer for fmri. *Frontiers in Neuroinformatics*, 9, 2015. doi: 10.3389/fninf.2015.00023.

Richard J. Gardner, Erik Hermansen, Marius Pachitariu, Yoram Burak, Nils A. Baas, Benjamin A. Dunn, May-Britt Moser, and Edvard I. Moser. Toroidal topology of population activity in grid cells. *Nature*, 602(7895):123–128, February 2022. ISSN 1476-4687. doi: 10.1038/s41586-021-04268-7.

Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring Statistical Dependence with Hilbert-Schmidt Norms. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Sanjay Jain, Hans Ulrich Simon, and Etsuji Tomita, editors, *Algorithmic Learning Theory*, volume 3734, pages 63–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-29242-5 978-3-540-31696-1. doi: 10.1007/11564089_7.

Wei Guo, Jie J. Zhang, Jonathan P. Newman, and Matthew A. Wilson. Latent learning drives sleep-dependent plasticity in distinct CA1 subpopulations. Preprint, Neuroscience, February 2020.

Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a Definition of Disentangled Representations. *arXiv:1812.02230 [cs, stat]*, December 2018.

Irina Higgins, Sébastien Racanière, and Danilo Rezende. Symmetry-Based Representations for Artificial and Biological General Intelligence, March 2022.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey Hinton. Similarity of Neural Network Representations Revisited, July 2019.

Nikolaus Kriegeskorte. Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2008. ISSN 16625137. doi: 10.3389/neuro.06.004.2008.

Nikolaus Kriegeskorte and Rogier A. Kievit. Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, 17(8):401–412, August 2013. ISSN 1364-6613. doi: 10.1016/j.tics.2013.06.007.

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.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

Patrick McClure and Nikolaus Kriegeskorte. Representational distance learning for deep neural networks. *Frontiers in computational neuroscience*, 10:131, 2016.

Edward H. Nieh, Manuel Schottdorf, Nicolas W. Freeman, Ryan J. Low, Sam Lewallen, Sue Ann Koay, Lucas Pinto, Jeffrey L. Gauthier, Carlos D. Brody, and David W. Tank. Geometry of abstract learned knowledge in the hippocampus. *Nature*, 595(7865):80–84, July 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03652-7.

Carl E. Schoonover, Sarah N. Ohashi, Richard Axel, and Andrew J. P. Fink. Representational drift in primary olfactory cortex. *Nature*, 594(7864):541–546, June 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03628-7.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-score: Which artificial neural network for object recognition is most brain-like? *bioRxiv preprint*, 2018. URL https://www.biorxiv.org/content/10.1101/407007v2.

Heiko H Schütt, Alexander D Kipnis, Jörn Diedrichsen, and Nikolaus Kriegeskorte. Statistical inference on representational geometries. *eLife*, 12:e82566, August 2023. ISSN 2050-084X. doi: 10.7554/eLife.82566.

Roger N. Shepard and Susan Chipman. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1(1):1–17, 1970. ISSN 1095-5623. doi: 10.1016/0010-0285(70)90002-2.

Peter Sterling and Simon Laughlin. *Principles of neural design*. MIT press, 2015.

Rudi Tong, Ronan da Silva, Dongyan Lin, Arna Ghosh, James Wilsenach, Erica Cianfarano, Pouya Bashivan, Blake Richards, and Stuart Trenholm. The feature landscape of visual cortex. *bioRxiv*, pages 2023–11, 2023.

Alex H. Williams, Erin Kunz, Simon Kornblith, and Scott W. Linderman. Generalized shape metrics on neural representations. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Peter Yianilos. Data structures and algorithms for nearest neighbor search in general metric spaces. *Proceedings of the Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, 02 1970.