Why all roads don't lead to Rome: Representation geometry varies across the human visual cortical hierarchy

Arna Ghosh[*] [1] [2] [6] ARNA.GHOSH@MAIL.MCGILL.CA

Zahraa Chorghay[*] ZAHRAA.CHORGHAY@UTORONTO.CA

Shahab Bakhtiari [†] [1] [3] SHAHAB.BAKHTIARI@UMONTREAL.CA

Blake A. Richards [†] [1] [2] [4] [5] [6] BLAKE.RICHARDS@MCGILL.CA

Editors: List of editors' names

Abstract

Biological and artificial intelligence systems navigate the fundamental efficiency-robustness tradeoff for optimal encoding, i.e., they must efficiently encode numerous attributes of the input space while also being robust to noise. This challenge is particularly evident in hierarchical processing systems like the human brain. With a view towards understanding how systems navigate the efficiency-robustness tradeoff, we turned to a population geometry framework for analyzing representations in the human visual cortex alongside artificial neural networks (ANNs). In the ventral visual stream, we found general-purpose, scale-free representations characterized by a power law-decaying eigenspectrum in most areas. However, in certain higher-order visual areas did not have scale-free representations, indicating that scale-free geometry is not a universal property of the brain. In parallel, ANNs trained with a self-supervised learning objective also exhibited scale-free geometry, but not after fine-tuning on a specific task. Based on these empirical results and our analytical insights, we posit that a system's representation geometry is not a universal property and instead depends upon the computational objective.

Keywords: Representation Geometry, Eigenspectrum, Human Visual Cortex, fMRI, Self-supervised Learning, Representation Learning, Neural Code.

1. Introduction

A fundamental challenge for both biological and artificial systems is navigating the trade-off between robustness and efficiency. Efficiency, in the context of the "efficient coding" hypothesis in neuroscience, refers to reducing information redundancy by eliminating correlations in the system's input space (Barlow et al., 1961; Atick and Redlich, 1990; Olshausen and Field, 1996; Simoncelli and Olshausen, 2001). This coding strategy results in a high-dimensional, sparse neural code, requiring only relatively simple downstream networks to

[*] Co-first authors [†] Co-senior authors
[1] Mila - Quebec AI Institute, Montreal, QC, Canada
[2] Computer Science, McGill University, Montreal, QC, Canada
[3] Department of Psychology, Universite de Montreal, Montreal, QC, Canada
[4] Montreal Neurological Institute, Montreal, QC, Canada
[5] CIFAR Learning in Machines & Brains Program, Toronto, ON, Canada
[6] Google , Paradigms of Intelligence

read out complex features. Overall, an efficient neural code allows the system to capture numerous attributes of the input space and better utilizes its total representation capacity. On the other hand, a low-dimensional, correlated, and redundant coding strategy confers robustness in the presence of noise at the cost of reduced representation capacity (Shadlen and Newsome, 1998; Reich et al., 2001). Since information processing requires representing increasingly complex features and abstractions of inputs we turned to population geometry approaches in the human visual cortex and in artificial neural networks (ANNs) to understand the coding strategies used by different intelligent systems.

While there exist various measures of population geometry, Stringer et al. (2019) showed that the neural responses in the mouse primary visual cortex (V1) have a unique signature: the eigenspectrum of the neural activity covariance obeyed a power law $n^{-\alpha}$, where $\alpha \sim 1$. These "scale-free" representations are high-dimensional yet smooth, such that the coefficient of the power law, α, reflects the fraction of neural variance that is devoted to representing coarse versus fine stimulus features. Examples of coarse distinctions are animate versus inanimate objects and outdoor versus indoor scenes, while fine distinctions could include differentiating individuals of the same species (e.g., faces of people, types of birds, etc.) or even different views of the same object. By having representations that are high-dimensional yet smooth, i.e., differentiable such that fine differences between stimuli can be represented while preserving coarse large-scale stimulus features, the neural code can navigate the tradeoff between efficiency and robustness.

Although scale-free representations have been observed in other species (Kong et al., 2022; Gauthaman et al., 2024) and in artificial neural networks (ANNs) (Agrawal et al., 2022), it remains an open question whether this geometric signature is a universal property or linked to a system's specific computational objective. To address this question, we studied the representation geometry by measuring the eigenspectrum in both the human ventral visual cortex and ANNs, and present an analytical framework that links a system's computational objective to its representation geometry. We find that representations in most visual areas are scale-free but in certain higher-order areas, representations lie in a subspace that does not have scale-free properties (i.e., a finite number of orthogonal directions are sufficient for capturing most of the stimulus-related information). This observation demonstrates that representations in the human brain are not universally scale-free. In parallel, ANNs trained with self-supervised losses exhibited scale-free representations. When we finetuned these ANNs on a specific task, they exhibited representations without scale-free properties. Together, these observations lead us to postulate that the geometric shift that we observe in certain higher-order cortical areas and task-finetuned ANNs could be a hallmark of functional specialization, which is a key aspect of hierarchical information processing. Together, our results provide a parallel between biological and artificial systems, suggesting that the shift in representation geometry likely indicates a shift in the computational objectives in the respective information processing stages.

2. Results

2.1. Representation geometry variation in the ventral visual cortical hierarchy

To investigate whether the human ventral visual cortex universally exhibits a power law code, we characterized the eigenspectrum of the activity covariance of individual brain areas

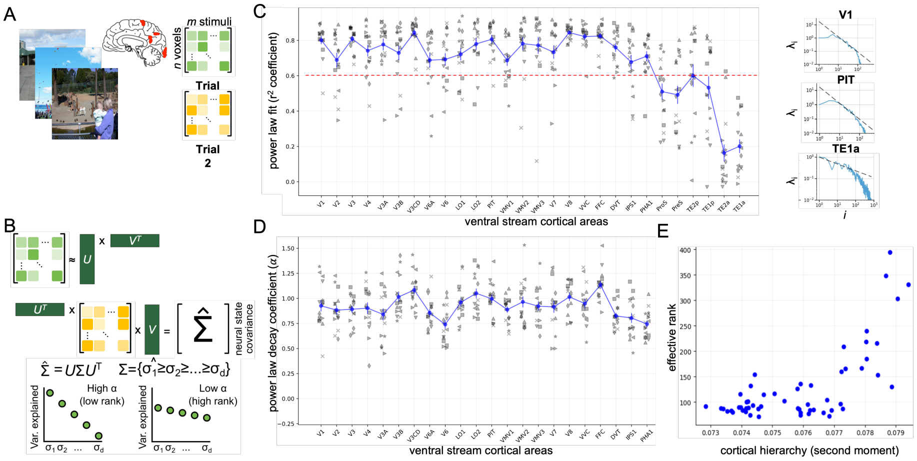

Figure 1: **Representation geometry across the human cortical hierarchy.** Early and intermediate areas, but not higher-order areas, exhibit scale-free representations. **(A)** Natural Scenes Dataset (NSD) and **(B)** cross-validation Principal Component Analysis (cvPCA) schematics. **(C)** Goodness of fit of the power law fit to the neural activity eigenspectrum for ventral visual cortical areas. **(D)** Decay coefficient (α) for areas with a good power law fit ($r^2 > 0.60$). **(E)** Effective rank of neural activity in ventral visual cortical areas increases with increase in cortical hierarchy metric ($\rho = 0.69$, $p < 1e^{-6}$).($n =$ left and right hemispheres for all 8 participants of NSD $= 16$ per area.)

using the Natural Scenes Dataset (NSD) (Figure 1a). Similar to Stringer et al. (2019), we used cross-validation principal component analysis (cvPCA) to compute the eigenspectrum of the neural responses to natural images of the early and ventral visual stream cortex as defined by the HCP-MMP parcellation atlas ((Glasser et al., 2016))(Figure 1b). Next, we fitted a power law to the eigenspectrum decay, and characterized its fit using the r^2 coefficient (Figure 1c). Most cortical areas exhibited a power law fit ($r^2 > 0.6$), but some areas showed a poor power law fit ($r^2 < 0.6$), including the presubiculum (PreS), prostriate area (ProS), and the anterior and posterior areas of TE1 and TE2. Strikingly, in areas with a power law fit ($r^2 > 0.60$), α was close to 1 (Figure 1d) in line with existing literature (Stringer et al., 2019; Kong et al., 2022; Gauthaman et al., 2024). In summary, while most areas of the ventral stream showed a scale-free geometry with $\alpha \sim 1$, certain higher-order areas did not, revealing that scale-free geometry is not a universal property across the brain.

To further validate our findings, we turned to another useful metric for characterizing population geometry that has been used in the self-supervised learning (SSL) literature to quantify the expressivity of learned representations (Roy and Vetterli, 2007; Garrido et al., 2023). This metric, called effective rank, is a distribution-agnostic metric of information geometry in high-dimensional spaces. We compared effective rank along the visual cortical hierarchy, with the hierarchy quantified by moment-based parameterization of staining intensity profiles ("second moment"; as per Paquola et al. (2021)). We observed that the representations in higher-order cortical areas exhibit a higher effective rank (Figure 1e). Along with Figure 1c, the effective rank indicates that higher-order cortical areas have high-dimensional representations that appear to lie in a subspace that does not have scale-free properties, i.e., most of the stimulus-related information is captured by a definite set of orthogonal directions.

Having observed differences in population geometry across the ventral visual hierarchy, we wondered what computational advantages this variation in representation geometry may confer? To address this question, we turned to artificial neural networks (ANNs). In contrast to the human brain, ANNs provide a more controllable experimental platform to examine the role of these two distinct population geometry coding strategies in supporting the computations performed by their respective systems.

2.1.1. An analytical viewpoint

Let us define a representation learning system that maps inputs from a domain $\mathcal{X} \subset \mathbb{R}^d$ to a learned feature space, $\mathbf{f} \subset \mathbb{R}^{d'}$. We adopt a theoretically ideal model to simplify the problem and postulate that each input $x \in \mathcal{X}$ is generated by a continuous and injective function $\mathbf{g} : \mathbb{S}^{D-1} \to \mathbb{R}^d$, where $\mathbb{S}^{D-1}$ is a sphere in D-dimensional latent space of latent factors z. These latent factors z are assumed to encode the fundamental attributes, or a combination thereof, present in the stimulus. For instance, in the context of the natural image space $\mathcal{X}$, z would encapsulate information regarding objects and their properties such as orientation or position. The function $\mathbf{g}(\cdot)$ then transforms these latent factors to synthesize the observed input x that is subsequently processed by the representation learning system.

The goal of any representation learning system is to reliably recover (some or all) latent factors, z, from given input data x. When these latent factors of interest are known *a priori*, the system aims to extract these specific factors while disregarding information about other factors. To the astute reader, this scenario may be reminiscent of the supervised learning setup, where the latent factors of interest are analogous to the predefined class identity of an image. Conversely, when latent factors of interest are not known *a priori*, the representation learning system aims to identify as many latent factors as possible while also upholding specific symmetries. This situation is akin to the self-supervised learning (SSL) paradigm, in which symmetries are explicitly defined by incorporating specific invariance properties in the learned representation space, typically achieved through the application of data augmentations. These data augmentations are crucial for identifying which latent factors will be learned, as those that are preserved under the chosen set of data augmentations will be considered as task-relevant "signal"; augmentations that are not preserved will be considered "noise," and rejected or suppressed in the representation space.

Despite these differences, both supervised and self-supervised learning leverage similar loss functions to train ANNs to learn effective representations. Supervised learning fundamentally relies on the cross-entropy loss to train ANNs, mapping input data to a probability distribution over predefined class identities. Recently, Reizinger et al. (2025) demonstrated a crucial theoretical insight: minimizing the cross-entropy loss was sufficient for learning representations that are a linear transformation of the underlying latent factor, z. This finding is particularly interesting if considering the connection to SSL, wherein there appears to be a fundamental equivalence between a diverse array of proposed loss functions (Zhai et al., 2024; Ghosh et al., 2024) and the cross-entropy loss commonly used in contrastive

SSL, such as SimCLR (Chen et al., 2020). This SSL cross-entropy loss operates like an instance classification setup (Balestriero and LeCun, 2024), where all symmetry-preserving augmented versions of a given image are required to be classified as belonging to the same instance. This SSL cross-entropy loss is also sufficient for learning a representation space that is a linear transformation of the latent factors preserved by the defined augmentations (Reizinger et al., 2025). Therefore, a unifying theoretical argument emerges: both supervised and self-supervised learning setups inherently promote the learning of representations that are linear transformations of the true factors. This fundamental link paves the way for a unified lens through which representations learned in both paradigms can be analyzed.

$$\mathbf{f}(x) = \mathbf{f} \circ \mathbf{g}(z) = \mathcal{O}\tilde{z} \tag{1}$$

where $\mathcal{O}$ represents an orthogonal linear transformation, and $\tilde{z}$ denotes the task-relevant subset of latent factors.

This theoretical finding relating the learned representation space to the latent factor space does not make specific assumptions about the ANN architecture or expressivity. Instead, it assumes that the network has sufficient capacity to learn a linear mapping and the training has converged to a minimum of the loss. When an eigenspectrum decomposition is applied to the learned representation space, it reveals the variance in the orthogonal directions within this space. Given that the learned representation space is a linear transformation of the task-relevant latent factor space, the representation eigenspectrum is indicative of the variance associated with the independent dimensions within the latent factors. Consequently, in a supervised learning setup, the representation eigenspectrum of ANNs will exhibit substantial variance corresponding to the latent factors essential for performing the task, followed by a sharp decline, signifying the suppression of other irrelevant latent factors. In contrast, the SSL representation eigenspectrum will demonstrate variance across all latent factors that the network successfully captured. The scale-free nature of this variance profile is very likely a consequence of the inherent structure of the latent factor space, a phenomenon further supported by recent studies on SSL learning dynamics (Simon et al., 2023; Ghosh et al., 2024). We will now empirically validate our analytical insights by studying the population geometry of representations learned in supervised and SSL setups.

2.1.2. Empirical evidence

To empirically test our core theoretical claim that the loss function dictates the geometry of the learned representation space, we compared ANNs trained with supervised and self-supervised learning (SSL) loss functions. We hypothesized that ANNs with general representations would exhibit a power law decay of the eigenspectrum, whereas specialized representations would not.

Here, we pretrained a Resnet-18 network using the BarlowTwins objective (Zbontar et al., 2021) on the Imagenet-100 dataset (Figure 2a). As expected, these SSL-pretrained networks learned generic representations of naturalistic images and exhibited a strong power law decay in their representation covariance eigenspectrum (Figure 2b). To generate task-specialized versions of these networks, we finetuned the pretrained network using a cross-entropy loss to perform image recognition, resulting in an improved accuracy of $\geq 5\%$ on the downstream image recognition task (as compared to linear evaluation alone). While

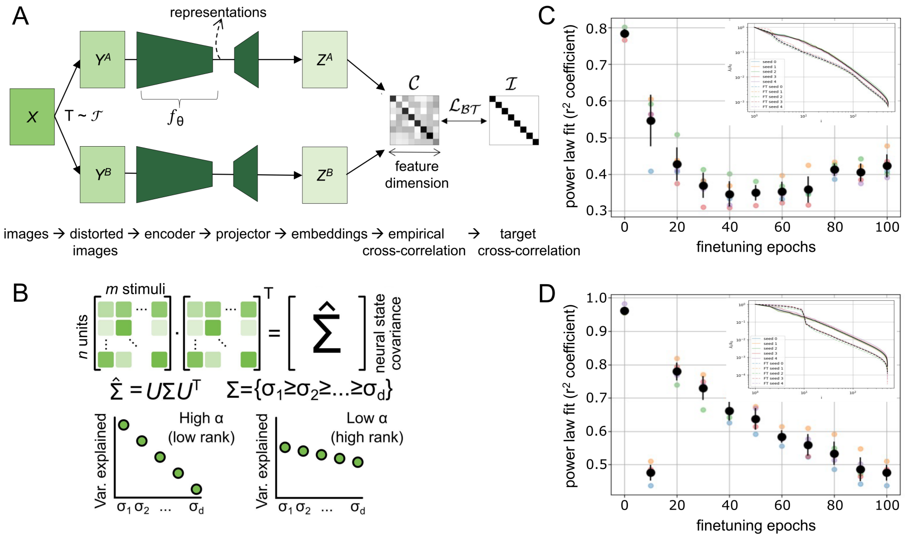

Figure 2: **Representation geometry changes as ANNs become specialized to a task.** General representations are scale-free but finetuned representations are not. **(A)** BarlowTwins training and **(B)** eigenspectrum computation schematics. Goodness of power law fit to the activity eigenspectrum across finetuning epochs on **(C)** ImageNet100 and **(D)** CIFAR-10.

accuracy improved over finetuning epochs, these specialized networks no longer exhibited a good power law fit (Figure 2c) of their eigenspectra (Figure 2c inset). To further validate these findings, we also finetuned the BarlowTwins-pretrained networks to perform image recognition on a different dataset, CIFAR-10. We saw an impressive 20% improved accuracy with finetuning, compared to linear evaluation alone. Following finetuning, these task-specialized networks also no longer demonstrated a good power law fit (Figure 2d) of their eigenspectra (Figure 2d inset).

These findings demonstrate that representations shift away from being scale-free as a network specializes for a given task, supporting our theoretical claims about the link between a system's representation geometry and respective computation objective function.

3. Discussion & Conclusion

In this work, we applied a population geometry framework to uncover fundamental computational principles of hierarchical information processing in both biological and artificial neural systems. We find that the representations are high-dimensional throughout the human ventral visual cortical hierarchy, as measured by the eigenspectrum decay and effective rank. Furthermore, we demonstrate a striking parallel between the ventral visual stream and artificial neural networks (ANNs): general-purpose representations, from most visual areas or from SSL pretrained networks, exhibit a scale-free geometry characterized by a power law decaying eigenspectrum. In contrast, specialized representations found in certain higher-order cortical areas and in ANNs finetuned to specific tasks lack this scale-free property. Together, our core finding is that the computational objective of a system is a key determinant of its representation geometry.

We posit that scale-free geometry is a hallmark of robust, generic representation learning, analogous to SSL-pretrained ANNs. By capturing features at multiple scales, this coding strategy provides an inherent robustness to noise. It also allows representing fine stimulus details while preserving coarse, large-scale stimulus features (Stringer et al., 2019; Agrawal et al., 2022). Such scale-free representations, whereby α is close to 1, are seen in ventral visual areas in the human brain (Stringer et al., 2019; Kong et al., 2022; Gauthaman et al., 2024) and in ANN models trained for generic representation learning (Agrawal et al., 2022). Overall, high-dimensional, scale-free representations are suited for supporting a wide variety of downstream behaviors.

On the other hand, some higher-order visual areas did not show a power law decay of their eigenspectrum, demonstrating for the first time (to our knowledge) that scale-free representations are not universal in the human brain. In our ANN experiments, finetuning a pretrained network to perform a specific image recognition task also eliminated the power law decay of the representation covariance eigenspectrum. Notably, such specialized representations (i.e., high-dimensional and not scale-free) have been shown to be computationally advantageous for tasks like few-shot learning (Sorscher et al., 2022), whereby a system must quickly generalize to novel concepts with limited examples.

These computational advantages of respective coding strategies (scale-free versus not) suggest that they can be applied by intelligent systems for distinct computational objectives. In other words, we speculate that the lack of power law exhibited in areas like PreS, ProS, TE1, and TE2 is linked to their different computational objective than that of early visual cortical areas. This matters in a neuroscience context, as our findings could provide a computational perspective of how initial sensory areas "solve" for robust, generic representations, while higher-order unimodal and multimodal areas operate on this robust code to achieve increased functional specialization through the cortical hierarchy.

Our results provide a unified framework for understanding the link between a system's computational objective and its representation geometry. A critical next step will be to establish causal links between the computational objective and representation geometry, for instance, by investigating the specific training recipes and architectural constraints needed for the formation of a scale-free eigenspectrum. Furthermore, our findings suggest that the hierarchical structure of the brain may be an elegant solution to the dual challenge of achieving both generic, robust representations and efficient, task-specialized encoding. Our work hopes to inspire future research on a more mechanistic understanding of how hierarchical organization allows the brain to navigate the efficiency-robustness tradeoff, and how this biological solution might inform the design of better artificial intelligence systems.

4. Methods

4.1. Neural dataset

Experiment Design To study the representation geometry across the visual hierarchy, we used the Natural Scenes Dataset (NSD). The NSD consists of high-resolution (7T) functional magnetic resonance imaging (fMRI) responses acquired from 8 participants performing a continuous recognition memory task on $\sim$10,000 natural scenes viewed over 30–40 scan sessions over one year (Allen et al., 2022). The stimuli were obtained from the Microsoft COCO dataset (Lin et al., 2014), containing complex everyday scenes with common objects. Images were shown three times over the course of the experiment, and trial-level response estimates were obtained for $\sim 30,000$ stimulus presentations.

Preprocessing We used the 1 mm volume preparation of the NSD and version 3 of the NSD single-trial betas (betas_fithrf_GLMdenoise_RR). In this version, the haemodynamic response function was estimated for each voxel, the GLMdenoise technique was applied for denoising, and ridge regression was used to estimate the single-trial stimulus-evoked activity. The Human Connectome Project Multi-Modal Parcellation (HCP-MMP) atlas segment the brain into different cortical areas. Based on known literature, we identified the ventral visual areas and analyzed the data for these areas: visual areas 1-4 and 7 (V1, V2, V3, V4, V7); posterior inferotemporal complex (PIT); and occipital-temporal areas, including lateral temporal posterior areas 2 and 1 (TE2p, TE1p) and lateral temporal anterior areas 2 and 1 (TE2a, TE1a). For each area, similar to Stringer et al. (2019), to preserve only task-relevant activity for each voxel, we performed spontaneous data cleaning by removing correlated activity across voxels that is shared during resting state (restingbetas_fithrf) and task sessions.

Eigenspectrum Computation To estimate the eigenspectrum of neural activity, we performed cross-validation Principal Component Analysis (cvPCA) as per Stringer et al. (2019). Briefly, we split the repeated responses for each unique stimuli into one of two groups, 'Trial 1' and 'Trial 2'. This split ensured that both groups had neural responses to the same stimuli set. Thereafter, we performed singular value decomposition of the 'Trial 1' response matrix (X_1) yielding $X_1 = U\Sigma_1 V^T$. Next, we computed the neural state covariance, $\hat{\Sigma} = U^T X_2 V$, where X_2 denotes the 'Trial 2' response matrix. The diagonal elements of $\hat{\Sigma}$ were used as the neural covariance eigenspectrum. This process was repeated 5 times, each time with a different split of the repeats of stimuli presentation.

Cortical hierarchy metric To quantify cortical hierarchy, for each area obtained from the BigBrain dataset (Amunts et al., 2013) (n=1 subject), we calculated the second moment of the cortical staining intensity profile as per Paquola et al. (2021). The cortical intensity profile indicates the relative density and depth of pyramidal neurons at each voxel, thereby providing an estimate of the cytoarchitecture-based laminar differentiation. A higher second moment indicates a higher-order cortical area.

4.2. Spectral metrics

Powerlaw decay coefficient, α To characterize the decay of the eigenspectrum, we tested whether it exhibits a heavy-tailed distribution, specifically, a power law decay, i.e., whether eigenvalues follow a distribution $\lambda_i \sim i^{-\alpha}$ (Stringer et al., 2019; Agrawal et al., 2022). We used a weighted linear regression in log space from rank 10 to $\sim$100, with weights as the inverse of log of the rank. The 10 to $\sim$100 range was used as the eigenspectrum becomes more sensitive to measurement noise beyond this range. We used the regression r^2 coefficient to estimate the goodness of power law fit. For areas with a good fit, we used the slope of the regression fit as the eigenspectrum's power law decay exponent, α.

Effective rank The effective rank provides a quantitative measure of the effective number of dimensions that capture most of the variance. We measured the effective rank of the representation space as per Roy and Vetterli (2007):

$$\text{Effective\_Rank} := exp\left(-\sum_k p_k \log(p_k)\right) \quad , \quad p_k = \frac{\lambda_k}{\sum_i \lambda_i} + \in \tag{2}$$

4.3. ANN experiments

SSL pretraining We trained a ResNet-18 on Imagenet-100 (Deng et al., 2009) with the BarlowTwins loss function (Zbontar et al., 2021) for 100 epochs using the Adam optimizer (Kingma and Ba, 2014), saving intermediate checkpoints at every 10 epochs. The pretraining loss function was as follows:

$$\begin{aligned}
\mathcal{L}_{BT} &= \sum_i (C_{ii} - 1)^2 + \beta \sum_i \sum_{j \neq i} C_{ij}^2 \\
C &= \frac{1}{n-1} \sum_{i=1}^{n} (f(\mathbf{x}_i) - \overline{f(\mathbf{x})})(f(\tilde{\mathbf{x}}_i) - \overline{f(\tilde{\mathbf{x}})})^T \\
\overline{f(\mathbf{x})} &= \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x}_i) \quad , \quad \overline{f(\tilde{\mathbf{x}})} = \frac{1}{n} \sum_{i=1}^{n} f(\tilde{\mathbf{x}}_i)
\end{aligned} \tag{3}$$

For each model checkpoint, we computed α and effective rank of the final layer representations. We found that the eigenspectrum of representation covariance matrix at each checkpoint had a good power law fit ($r^2 > 0.9$).

Linear evaluation We appended a linear layer that mapped the learned features to class logits and trained this linear layer on the classification loss. The parameters of the rest of the network were frozen, thereby keeping the representation space unchanged. For the Imagenet-100 dataset, the output of the linear layer was 100-dimensional, whereas it was 10-dimensional for CIFAR-10 finetuning. In both cases, the linear layer was trained for 200 epochs with Adam optimizer.

Supervised finetuning We appended a linear layer that mapped the learned features to class logits and finetuned the BarlowTwins-pretrained network along with the linear layer on the classification loss. As with linear evaluation, the output of the linear layer was 100-dimensional for the Imagenet-100 dataset, whereas it was 10-dimensional for CIFAR-10 finetuning. In both cases, we finetuned the network for 200 epochs with Adam optimizer.

Eigenspectrum computation In each case, the representations were extracted from the final layer of ResNet-18 — before the projector network (used during BarlowTwins pretraining) or the linear classification layer (used during finetuning). We computed the representation covariance matrix for 10,000 stimuli and performed eigenspectrum decomposition of this covariance matrix. As for the neural covariance eigenspectrum, the spectral metrics were computed using this representation covariance eigenspectrum.

Acknowledgments

We would like to thank Sethu Kovendhan Boopathy Jegathamabal for facilitating access to the BigBrain dataset, and Casey Paquola for her insights on cortical hierarchy metrics.

Funding Sources A.G. was supported by Vanier Canada Graduate Scholarship. S.B. was supported by NSERC (Discovery Grant RGPIN-2023-03875) and Canada Excellence Research Chair (CERC). B.A.R. was supported by NSERC (Discovery Grant RGPIN-2020-05105; Discovery Accelerator Supplement RGPAS-2020-00031) and CIFAR (Canada AI Chair; Learning in Machines and Brains Fellowship). We also acknowledge the material support of NVIDIA in the form of computational resources, as well as the compute resources, software, and technical help provided by Mila (mila.quebec).

References

Kumar K Agrawal, Arnab Kumar Mondal, Arna Ghosh, and Blake Richards. α-req: Assessing representation quality in self-supervised learning by measuring eigenspectrum decay. *Advances in Neural Information Processing Systems*, 35:17626–17638, 2022.

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, et al. A massive 7t fmri dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1):116–126, 2022.

Katrin Amunts, Claude Lepage, Louis Borgeat, Hartmut Mohlberg, Timo Dickscheid, Marc-Étienne Rousseau, Sebastian Bludau, Pierre-Louis Bazin, Lindsay B Lewis, Ana-Maria Oros-Peusquens, et al. Bigbrain: an ultrahigh-resolution 3d human brain model. *Science*, 340(6139):1472–1475, 2013.

Joseph J Atick and A Norman Redlich. Towards a theory of early visual processing. *Neural Computation*, 2(3):308–320, 1990.

Randall Balestriero and Yann LeCun. The birth of self supervised learning: A supervised theory. In *NeurIPS 2024 Workshop: Self-Supervised Learning-Theory and Practice*, 2024.

Horace B Barlow et al. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01):217–233, 1961.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Quentin Garrido, Randall Balestriero, Laurent Najman, and Yann Lecun. Rankme: Assessing the downstream performance of pretrained self-supervised representations by their rank. In *International conference on machine learning*, pages 10929–10974. PMLR, 2023.

Raj Magesh Gauthaman, Brice Ménard, and Michael F Bonner. Universal scale-free representations in human visual cortex. *arXiv preprint arXiv:2409.06843*, 2024.

Arna Ghosh, Kumar Krishna Agrawal, Shagun Sodhani, Adam Oberman, and Blake Richards. Harnessing small projectors and multiple views for efficient vision pretraining. *Advances in Neural Information Processing Systems*, 37:39837–39868, 2024.

Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Nathan CL Kong, Eshed Margalit, Justin L Gardner, and Anthony M Norcia. Increasing neural network robustness improves match to macaque v1 eigenspectrum, spatial frequency preference and predictivity. *PLOS Computational Biology*, 18(1):e1009739, 2022.

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.

Bruno A Olshausen and David J Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

Casey Paquola, Jessica Royer, Lindsay B Lewis, Claude Lepage, Tristan Glatard, Konrad Wagstyl, Jordan DeKraker, Paule-J Toussaint, Sofie L Valk, Louis Collins, et al. The bigbrainwarp toolbox for integration of bigbrain 3d histology with multimodal neuroimaging. *Elife*, 10:e70119, 2021.

Daniel S Reich, Ferenc Mechler, and Jonathan D Victor. Independent and redundant information in nearby cortical neurons. *Science*, 294(5551):2566–2568, 2001.

Patrik Reizinger, Alice Bizeul, Attila Juhos, Julia E Vogt, Randall Balestriero, Wieland Brendel, and David Klindt. Cross-entropy is all you need to invert the data generating process. In *The thirteenth international conference on learning representations*, 2025.

Olivier Roy and Martin Vetterli. The effective rank: A measure of effective dimensionality. In *2007 15th European signal processing conference*, pages 606–610. IEEE, 2007.

Michael N Shadlen and William T Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998.

James B Simon, Dhruva Karkada, Nikhil Ghosh, and Mikhail Belkin. More is better in modern machine learning: when infinite overparameterization is optimal and overfitting is obligatory. *arXiv preprint arXiv:2311.14646*, 2023.

Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24(1):1193–1216, 2001.

Ben Sorscher, Surya Ganguli, and Haim Sompolinsky. Neural representational geometry underlies few-shot concept learning. *Proceedings of the National Academy of Sciences*, 119(43):e2200800119, 2022.

Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. High-dimensional geometry of population responses in visual cortex. *Nature*, 571 (7765):361–365, 2019.

Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In *International conference on machine learning*, pages 12310–12320. PMLR, 2021.

Runtian Zhai, Bingbin Liu, Andrej Risteski, J Zico Kolter, and Pradeep Kumar Ravikumar. Understanding augmentation-based self-supervised representation learning via rkhs approximation and regression. In *The twelfth international conference on learning representations*, 2024.

Appendix A. Additional spectral metric results

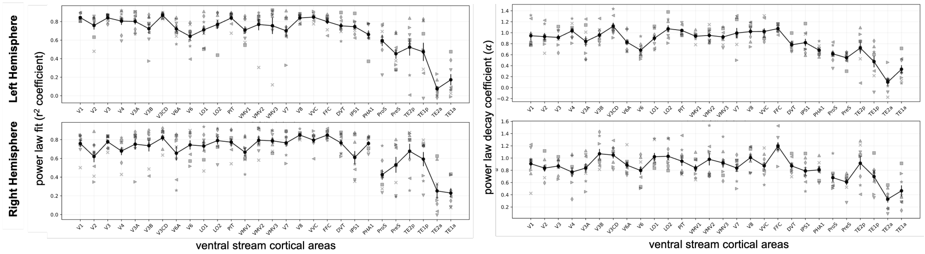

Figure 3: Power law goodness of fit, measured using regression r^2 coefficient, and decay coefficient (α) for different areas for each hemisphere.

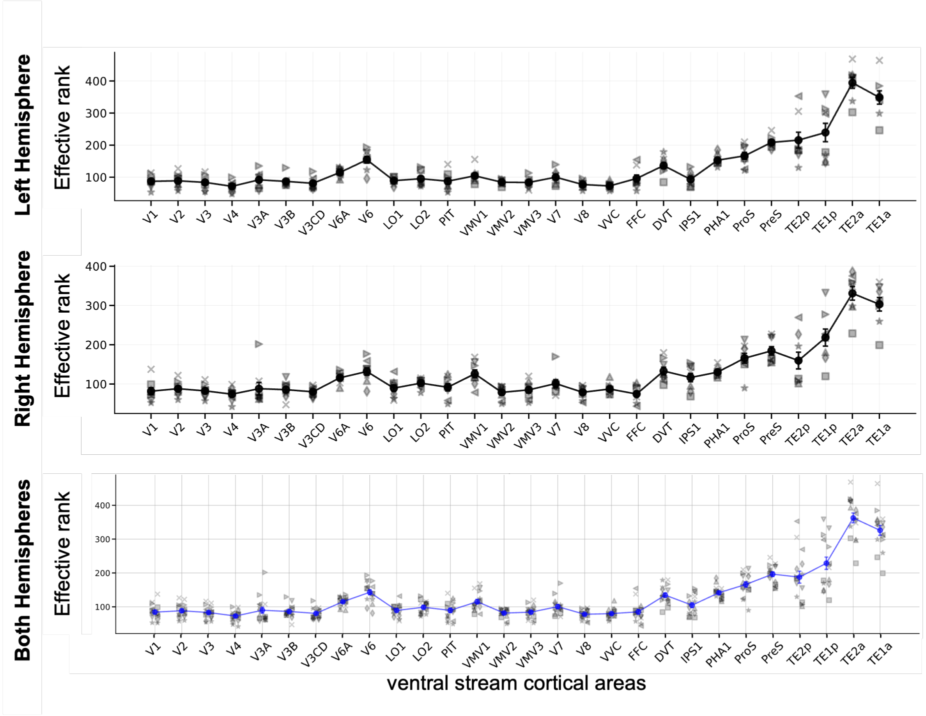

Figure 4: Effective rank for different areas for each hemisphere, and combined for both hemispheres.