The Geometry of Cortical Computation: Manifold Disentanglement and Predictive Dynamics in VCNet

Brennen Hill*
Department of Computer Science
University of Wisconsin-Madison
Madison, WI 53706
bahill4@wisc.edu

Zhang Xinyu
Department of Computer Science
National University of Singapore
Singapore 119077
zhang_xinyu_@u.nus.edu

Timothy Putra Prasetio
Department of Computer Science
National University of Singapore
Singapore 119077
timothy_prasetio@u.nus.edu

Abstract

Despite their success, modern convolutional neural networks (CNNs) exhibit fundamental limitations, including data inefficiency, poor out-of-distribution generalization, and vulnerability to adversarial perturbations. These shortcomings can be traced to a lack of inductive biases that reflect the inherent geometric structure of the visual world. The primate visual system, in contrast, demonstrates superior efficiency and robustness, suggesting that its architectural and computational principles,which evolved to internalize these structures,may offer a blueprint for more capable artificial vision. This paper introduces Visual Cortex Network (VCNet), a novel neural network architecture whose design is informed by the macro-scale organization of the primate visual cortex. VCNet is framed as a geometric framework that emulates key biological mechanisms, including hierarchical processing across distinct cortical areas, dual-stream information segregation for learning disentangled representations, and top-down predictive feedback for representation refinement. We interpret these mechanisms through the lens of geometry and dynamical systems, positing that they guide the learning of structured, low-dimensional neural manifolds. We evaluate VCNet on two specialized benchmarks: the Spots-10 animal pattern dataset, which probes sensitivity to natural textures, and a light field image classification task, which requires processing higher-dimensional visual data. Our results show that VCNet achieves state-of-the-art accuracy of 92.1% on Spots-10 and 74.4% on the light field dataset, surpassing contemporary models of comparable size. This work demonstrates that integrating high-level neuroscientific principles, viewed through a geometric lens, can lead to more efficient and robust models, providing a promising direction for addressing long-standing challenges in machine learning.

1 Introduction

Contemporary deep learning models for image recognition, particularly Convolutional Neural Networks (CNNs), have achieved remarkable success Krizhevsky et al. [2012]. However, their prowess

*Corresponding author.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Symmetry and Geometry in Neural Representations.

is shadowed by critical challenges that impede their widespread and reliable deployment. These models often require vast, meticulously labeled training datasets, exhibit poor generalization to out-of-distribution (OOD) examples Sagawa et al. [2020], and are notoriously vulnerable to adversarial attacks and partial occlusion Liu et al. [2022]. Minor, human-imperceptible perturbations or hidden object parts can cause catastrophic failures, raising concerns about their reliability in safety-critical applications. These persistent issues are not merely implementation details but may point to a fundamental inadequacy in their core architectural assumptions. While CNNs incorporate a crucial geometric prior-translation equivariance,they largely fail to account for other fundamental symmetries and structures of the visual world, such as rotation, scale, and the compositional nature of objects.

In stark contrast, the primate visual system is a paragon of efficiency and robustness. Humans can learn to recognize objects from few examples Lake et al. [2015], generalize effortlessly across novel contexts Geirhos et al. [2018], robustly identify occluded objects Hegdé et al. [2008], and operate with unparalleled energy efficiency Lennie [2003]. Mounting evidence from neuroscience suggests these capabilities are rooted in the specific architectural and computational principles of the visual cortex. Neural representations in the brain appear to be organized on low-dimensional, structured geometric spaces, often referred to as neural manifolds. The brain seems to learn not just features, but the underlying geometric and topological structure of the data-generating process. This is achieved through its unique hierarchical organization Felleman and van Essen [1991], Grill-Spector and Malach [2004] and its use of predictive processing Rao and Ballard [1999], de Lange et al. [2018].

This paper seeks to bridge the gap between the brute-force pattern matching of modern CNNs and the geometrically-aware, structured inference of the brain. We propose VCNet, a novel neural network whose macro-architecture is derived from the primate visual cortex. We go beyond mere biomimicry and interpret the cortex's organization as a computational framework for learning geometrically sound representations. Our work is guided by the thesis that principles of symmetry and geometry can illuminate the foundations of intelligence. Our contributions are threefold:

- We introduce **VCNet**, a deep neural network architecture that models the high-level information flow between major areas of the visual cortex. We provide a **geometric interpretation** of its core components, including dual-stream processing for manifold disentanglement, recurrent connections for representation dynamics, and top-down predictive feedback as a mechanism for geometric refinement on these manifolds.
- We demonstrate the efficacy of VCNet on the **Spots-10 animal pattern benchmark**. This task is selected to test our bio-inspired architecture on a problem that mirrors a key evolutionary pressure for vision, and we show that VCNet outperforms other models of comparable size.
- We further evaluate VCNet on a **light field image classification task**, providing evidence that its geometrically-motivated design is particularly well-suited for processing richer, multi-view data that more closely approximates the input to the human visual system.

2 Related Work

Our research is situated at the confluence of geometric deep learning, computational neuroscience, and neuro-inspired AI.

Geometric Deep Learning and Equivariance Geometric deep learning seeks to incorporate geometric priors and symmetries into neural network architectures Bronstein et al. [2021]. A significant focus has been on achieving equivariance to groups of transformations beyond translation. Steerable CNNs Cohen and Welling [2017] and E(2)-equivariant CNNs Maurice Weiler [2019] generalize convolutions to handle rotation and reflection, leading to improved data efficiency and generalization on tasks with these underlying symmetries. Graph Neural Networks (GNNs) extend these ideas to data defined on non-Euclidean domains like graphs and manifolds. While these approaches enforce geometric constraints at the micro-level of individual filters or operations, VCNet takes a complementary, macro-level approach. We do not engineer specific filter symmetries; instead, we hypothesize that the high-level architectural organization of the visual cortex itself, its network

topology of processing streams and feedback loops, creates an inductive bias that guides the learning process toward geometrically structured representations.

Neuro-Inspired Architectures The brain has long been a source of inspiration for artificial intelligence. Early models like the Neocognitron Fukushima [1980] laid the groundwork for modern CNNs by mimicking the simple and complex cells of V1. More recently, models like CorNet Kubilius et al. [2019] have sought to create architectures that not only perform well but also whose internal activations correlate with neural recordings from the primate brain. These models often focus on replicating the feedforward ventral stream. VCNet differentiates itself by modeling a more comprehensive set of cortical principles: (1) the explicit separation and interaction of the ventral and dorsal streams, (2) the inclusion of recurrent dynamics to model iterative processing, and (3) the implementation of a top-down predictive coding loop, which we argue is critical for robust, generative understanding.

Predictive Coding and Generative Models Predictive coding posits that the brain is fundamentally a generative model of its environment Rao and Ballard [1999]. Higher cortical areas generate predictions about lower-level sensory input, and only the residual error between the prediction and the actual input is propagated forward. This principle is computationally efficient and has deep connections to Bayesian inference and the free-energy principle Friston [2010]. In machine learning, this resonates with the objectives of generative models like Variational Autoencoders (VAEs) and Helmholtz machines, which learn a latent generative model of the data. Our implementation of predictive coding in VCNet serves a similar purpose, encouraging the network to learn an internal model of the visual world. By framing this process geometrically, we view the prediction error as a vector in the tangent space of a learned representation manifold, driving the model's state along a geodesic toward a more accurate representation.

3 The VCNet Architecture: A Geometric Interpretation

While a complete replication of the visual system is infeasible, our research focuses on emulating the macro-scale organization of the visual cortex. We interpret its connectivity and computational patterns not as an arbitrary biological arrangement, but as a sophisticated framework for learning and processing geometric information. VCNet is a deep neural architecture engineered to operationalize these principles.

3.1 Biologically-Inspired Geometric Principles

Our model's design is predicated on three foundational principles of primate vision, which we reformulate in the language of geometry and dynamics.

Hierarchical Processing as Compositional Feature Geometry Visual information propagates from the retina through a hierarchy of cortical areas (V1, V2, V3, V4, V5), each specialized for extracting progressively complex features Huff et al. [2023]. V1 detects simple elements like oriented edges, which can be understood as recognizing local Euclidean symmetries (translations and rotations of local patterns). It projects to V2, which processes intermediate features like contours and textures. V2, in turn, projects to higher-order areas like V4 and V5. We interpret this hierarchy not merely as a cascade of feature extractors, but as a sequence of learned, non-linear transformations $\phi_i : \mathcal{M}_{i-1} \to \mathcal{M}_i$ that map representations from one manifold $\mathcal{M}_{i-1}$ to another $\mathcal{M}_i$. Each stage aims to create a new representation that is more abstract, more disentangled, and more useful for the organism's goals. The final representation should ideally live on a manifold where semantic categories are linearly separable.

Dual-Stream Processing as Manifold Disentanglement The visual cortex is famously organized into two primary processing pathways Sheth and Young [2016]. The **ventral stream** ("what" pathway) is responsible for object recognition, while the **dorsal stream** ("where/how" pathway) handles spatial awareness and motion analysis. We propose a geometric interpretation: these two streams learn to project the high-dimensional sensory input onto two distinct, yet correlated, low-dimensional manifolds.

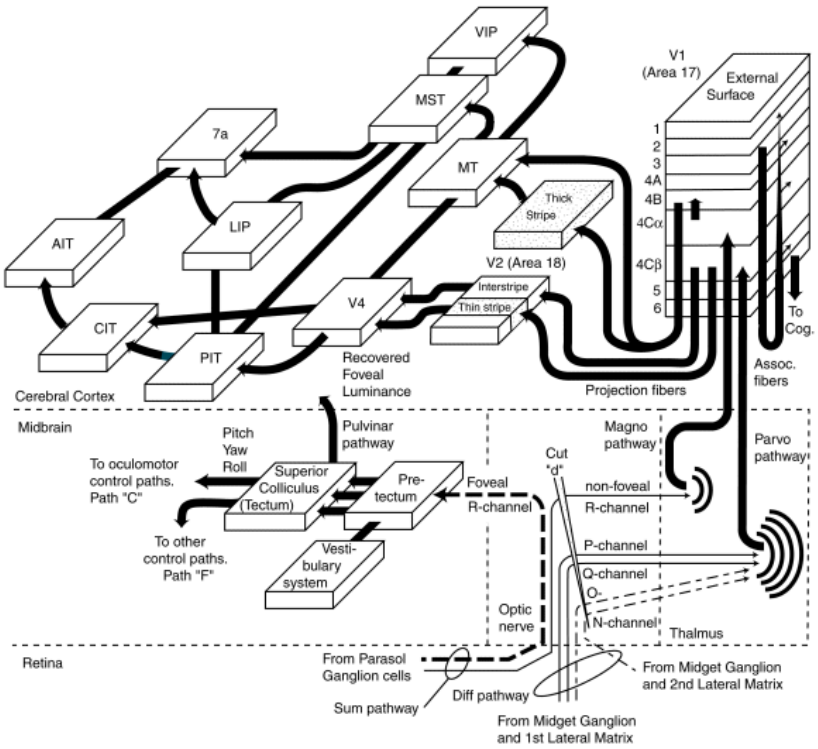

Figure 1: A high-level model of information pathways in the primate visual cortex, illustrating the hierarchical series of feature extraction stages Fulton [2001]. This organization forms the architectural basis of VCNet, which we interpret as a graph of transformations between representation manifolds.

- The **ventral stream** learns an **identity manifold**, $\mathcal{M}_{id}$. The goal is to learn a representation that is *invariant* to changes in pose, illumination, and position. Points on this manifold correspond to object identities.
- The **dorsal stream** learns a **pose/motion manifold**, $\mathcal{M}_{pose}$. The goal here is to learn a representation that is *equivariant* to changes in object position and orientation. Points on this manifold correspond to spatial properties.

By explicitly separating these tasks into different architectural pathways, the model is encouraged to learn disentangled representations, a key goal in representation learning. The interconnections between the streams allow the model to bind "what" information with "where" information.

Predictive Coding as Geodesic Refinement on a Manifold The visual cortex is not a purely feedforward system. It employs predictive coding, where higher-level areas send top-down predictions to lower-level areas Lowet and Uchida [2024]. We formalize this as a process of refinement on a representation manifold. Let z_L be a representation in a high-level area (e.g., AIT). The network learns a top-down mapping $p(z_l|z_L)$ that generates a prediction of the representation z_l in a lower-level area (e.g., V1). The bottom-up processing provides the actual sensory evidence, resulting in an observed representation z_l^{obs}. The discrepancy, or prediction error ϵ, is computed:

$$\epsilon = z_l^{obs} - p(z_l|z_L)$$

This error signal is not just noise; it is a vector in the tangent space of the lower-level manifold $\mathcal{M}_l$. This vector indicates the direction in which the higher-level representation z_L needs to be updated to better explain the sensory data. The learning process, which aims to minimize this error over time and data, can be viewed as an optimization process that encourages the model's internal generative trajectory to follow a geodesic on the manifold of plausible world states. This creates a powerful dynamical system for inference and learning.

3.2 Architectural Framework of VCNet

Departing from conventional, monolithic CNN architectures, VCNet is structured as a directed graph that models the known connectivity between the major visual cortical areas. The channel capacity of each module is scaled to approximate the relative neuronal populations in its biological counterpart.

Ventral Stream This "what" pathway models object recognition, progressing from V1 through modules representing V2 (Interstripe, Thin Stripe), V4, and the inferotemporal (PIT, CIT, AIT) cortices. It is specialized for learning the invariant *identity manifold*.

Dorsal Stream This "where/how" pathway models spatial and motion analysis, flowing from V1 through V2 (Thick Stripe), the middle temporal (MT) and medial superior temporal (MST) areas, and onward to parietal regions. It is specialized for learning the equivariant *pose/motion manifold*.

These streams are interconnected at multiple levels, enabling the integration of object identity with spatial information. The final representation is formed in the AIT module, which receives convergent inputs and feeds into the classification layer. VCNet's functionality is realized through several specialized computational blocks, each with a geometric interpretation.

3.2.1 Multi-Scale Feature Extraction (V1)

Implementation: To emulate the diverse receptive field sizes in V1, the module processes input through three parallel depthwise separable convolution streams with different kernel sizes (3x3, 5x5, 7x7). The resulting feature maps are concatenated. **Geometric Interpretation:** This block acts as a multi-scale probe of the local geometry of the input signal. Different kernel sizes are sensitive to structures at different frequencies and scales, akin to a wavelet decomposition. This provides a rich, multi-scale initial representation that captures the geometry of the input space more effectively than a single-scale approach.

3.2.2 Recurrent Processing Blocks (MT/MST)

Implementation: To model the iterative refinement of representations, the MT and MST modules incorporate Recurrent Blocks. These blocks apply a convolutional transformation with shared weights for a fixed number of iterations ($t = 3$), with each iteration receiving the output of the previous one plus a residual connection from the initial input. **Geometric Interpretation:** This recurrent application of a transformation defines a discrete-time dynamical system on the feature space. The representation z_t at iteration t evolves according to $z_{t+1} = f(z_t) + z_0$. The recurrence allows the representation to iteratively converge toward a stable fixed point on its manifold, effectively refining the estimate of motion or spatial properties over time.

3.2.3 Attentional Modulation (CBAM)

Implementation: To emulate the brain's ability to focus on salient features, key modules (V1, MT, V4) incorporate a Convolutional Block Attention Module (CBAM). CBAM sequentially infers and applies channel-wise and spatial attention maps. **Geometric Interpretation:** Attention can be viewed as a mechanism for dynamically selecting a relevant subspace of the feature manifold. Channel attention re-weights the contribution of different feature dimensions, effectively stretching or shrinking the manifold along certain axes. Spatial attention re-weights different locations, focusing computational resources on a specific region of the manifold's domain (the image space).

3.2.4 Lateral Interaction Module (V1)

Implementation: The V1 module includes a Lateral Interaction block, implemented as a convolution followed by channel-wise self-attention within a residual connection. **Geometric Interpretation:** This simulates the horizontal connections within cortical layers that mediate contextual effects like lateral inhibition. Geometrically, this enforces local consistency constraints on the feature manifold. It encourages nearby points in the image space to have related representations, promoting smoothness and helping to form coherent structures like contours, which can be seen as enforcing a local group structure.

3.2.5 Predictive Coding Loop (AIT to V1)

Implementation: We implement predictive coding via a top-down connection from the highest level of the ventral stream (AIT) back to V1. The AIT module generates a prediction of V1 feature activations. This prediction is subtracted from the actual bottom-up V1 activity to compute a prediction error, $\epsilon = \text{ReLU}(\text{V1}_{\text{bottom-up}} - \text{AIT}_{\text{top-down}})$. This error signal is then used as an additional learning signal. **Geometric Interpretation:** This is the core of our geometric dynamical system. The top-down signal from AIT is a hypothesis about the world, projected back onto the low-level V1 feature manifold. The bottom-up signal is the evidence. The error ϵ is a vector field on the V1 manifold that drives the learning process, forcing the high-level AIT manifold to generate representations that are consistent with the low-level sensory data. This process refines the entire hierarchy of representational geometries.

3.2.6 Neuromodulatory Gating

Implementation: To model the global gain control exerted by neuromodulators, we introduce a Neuromodulation block in key modules (V1, MT, V4). This block applies a learnable, channel-wise multiplicative scaling factor to feature maps. **Geometric Interpretation:** This mechanism can be interpreted as controlling the local curvature or metric of the representation manifold. A high gain (scaling factor > 1) could increase the sensitivity of the representation to small input changes (increasing curvature), while a low gain (scaling factor < 1) could promote invariance by flattening the manifold locally. This allows the network to adapt its representational geometry based on context.

4 Experiments and Results

We benchmarked VCNet's performance against contemporary neural networks of comparable size to assess its image classification capabilities. We chose datasets and tasks that are particularly relevant to the evolution and function of biological vision and that test the geometric principles embedded in our architecture.

4.1 Experimental Setup

All models were trained using the Adam optimizer Kingma and Ba [2015] with a learning rate of 10^{-3}. We used a batch size of 16 and applied standard data augmentation techniques, including random horizontal flips and random rotations. All experiments were conducted using Google Colab Google.

4.2 Experiment 1: Animal Pattern Classification

Motivation Key evolutionary drivers for primate vision include finding food and avoiding predators, tasks that rely heavily on recognizing natural patterns and textures Kaas [2012]. The primate visual cortex is thus highly optimized for this purpose. We therefore evaluated our biologically-inspired model on a benchmark focused on classifying animal patterns, which tests the model's ability to learn representations of complex, semi-structured textures.

Methodology We utilized the Spots-10 dataset, which contains 50,000 grayscale 32x32 pixel images across 10 classes of animal patterns Atanbori [2024]. We trained VCNet Mini and compared its performance against a suite of established models whose weights were derived via knowledge distillation, making them highly compact and efficient. To ensure a consistent evaluation of convergence, all baseline models were finetuned on Spots-10 for the same number of epochs that VCNet Mini was trained.

Results As shown in Table 1, VCNet Mini attains the highest accuracy on Spots-10 (92.08%), outperforming the strongest baseline (DenseNet121 Distiller, 81.84%) by a significant margin of 10.24 percentage points. To ensure a fair comparison with the lightweight distilled baselines, we reduced VCNet's hidden-layer widths to form the *Mini* variant. Remarkably, VCNet Mini achieves this superior performance while using only 0.04 MB of storage, about 43% smaller than the 0.07 MB baselines. These findings strongly indicate that architectures inspired by the geometric

Table 1: Test accuracy and model size on the Spots-10 benchmark. Best values are in bold. VCNet Mini demonstrates superior accuracy with significantly fewer parameters, highlighting the efficiency of its architectural priors.

Model	Test Accuracy (%)	Model Size (MB)
VCNet Mini (Ours)	**92.08**	**0.04**
DenseNet121 Distiller	81.84	0.07
ResNet101V2 Distiller	80.29	0.07
ResNet50V2 Distiller	79.03	0.07
MobileNet Distiller	78.26	0.07
MobileNetV3-Small Distiller	78.04	0.07

and computational principles of the visual cortex can yield models that are both highly accurate and extremely parameter-efficient.

4.3 Experiment 2: Light Field Classification

Motivation Standard 2D images are flat projections of the 3D world, discarding vast amounts of visual information related to depth, parallax, and view-dependent reflectance. The human visual system (HVS) processes a much richer input, leveraging binocular vision and eye movements to interpret a subset of the 7D plenoptic function Adelson and Bergen [1991]. Light field cameras, which capture both the intensity and the angular direction of light rays, provide data that is a much closer analogue to the input processed by the HVS Lin et al. [2024]. We hypothesize that an architecture designed to emulate the visual cortex's dual-stream, geometrically-aware processing will demonstrate superior performance on this richer data modality.

Methodology We evaluated VCNet on a light field image classification task using a standard dataset Raj et al. [2016]. The light field data was processed into a 4D tensor, which was then fed into the models. We compared its performance against benchmark models: ResNet18, VGG11 with Batch Normalization, and MobileNetV2. These baselines were pre-trained on ImageNet and finetuned for the same number of epochs as VCNet was trained.

Table 2: Performance and Size Comparison on Light Field Image Classification. VCNet achieves the highest accuracy while being the most compact model, demonstrating its suitability for processing higher-dimensional, geometrically rich visual data.

Model	Test Accuracy (%)	Model Size (MB)
VCNet (Ours)	**74.42**	**3.52**
MobileNetV2	72.09	8.66
ResNet18	65.12	42.69
VGG11_BN	51.16	491.39

Results The results, summarized in Table 2, highlight VCNet's superior performance and efficiency. VCNet achieved the highest test accuracy (74.42%) while maintaining a minimal model size of 3.52 MB. This is over twice as small as MobileNetV2, over ten times smaller than ResNet18, and over 100 times smaller than VGG11. This result validates our hypothesis that an architecture incorporating principles like dual-stream processing and predictive feedback is particularly effective for processing high-dimensional visual data that contains both object identity and spatial/viewpoint information.

5 Conclusion and Future Work

In this work, we introduced VCNet, an architecture guided by the computational principles and anatomical organization of the primate visual cortex. By interpreting these principles through the lens of geometry and dynamical systems, framing dual-stream processing as manifold disentanglement and predictive coding as geodesic refinement, we developed a model that demonstrates superior performance and parameter efficiency on challenging image classification tasks. Our findings

underscore the significant potential of this approach: by embedding high-level geometric priors into the macro-architecture of a network, we can guide it to learn more robust and efficient representations. This convergence of disciplines not only offers a path toward more capable artificial systems but also provides computational frameworks for testing hypotheses about brain function.

Our model opens several avenues for future research.

- **Integrating Explicit Equivariance:** A powerful next step would be to combine our macro-scale geometric approach with the micro-scale constraints of geometric deep learning. Incorporating steerable filters into the V1 module could enforce explicit rotation and scale equivariance at the lowest level, which could then be integrated into the global representations of the full VCNet architecture.
- **Topological Data Analysis (TDA):** The geometric structure of the learned manifolds could be more formally analyzed using tools from TDA. We could use persistent homology to quantify the topological structure of the representations, testing the hypothesis that better representations have simpler topology (e.g., one connected component per class with no spurious holes).
- **Extension to Spatio-Temporal Dynamics:** The current model is designed for static images. Extending it to video processing is a natural progression. The dorsal stream and recurrent dynamics would become even more critical for modeling the flow of information on spatio-temporal manifolds, potentially leading to more robust action recognition and video prediction models.
- **Geometric Principles in Language:** Finally, it is worth speculating whether the principles of hierarchical, predictive, and geometrically structured representations could be applied to other modalities. Understanding the geometry of representations in large language models is a burgeoning field, and concepts like dual-stream processing (e.g., for syntax vs. semantics) and predictive refinement could offer valuable new architectural ideas.

Author Contributions

Brennen Hill: Project lead, conceptualization, software, engineering, investigation, research, writing
Zhang Xinyu: Software, engineering, investigation, research,writing.
Timothy Putra Prasetio: Software, engineering, investigation, research, writing.

References

Edward H. Adelson and James R. Bergen. The Plenoptic Function and the Elements of Early Vision. In Michael S. Landy and J. Anthony Movshon, editors, *Computational Models of Visual Processing*, pages 3–20. MIT Press, Cambridge, MA, 1991.

John Atanbori. SPOTS-10: Animal Pattern Benchmark Dataset for Machine Learning Algorithms. https://paperswithcode.com/paper/spots-10-animal-pattern-benchmark-dataset-for, 2024. Accessed on November 14, 2024.

Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021. URL https://arxiv.org/abs/2104.13478.

Taco S. Cohen and Max Welling. Steerable cnns. In *International Conference on Learning Representations (ICLR)*, 2017. URL https://arxiv.org/abs/1612.08498.

Floris P. de Lange, Micha Heilbron, and Peter Kok. How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, 22(9):764–779, 2018. doi: 10.1016/j.tics.2018.06.002.

Daniel J. Felleman and David C. van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991. doi: 10.1093/cercor/1.1.1.

Karl Friston. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, 11(2): 127–138, 2010. doi: 10.1038/nrn2787.

Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980. doi: 10.1007/BF00344251.

James T. Fulton. *Processes in Biological Vision*. Self-published, 2001. Available from: https://www.researchgate.net/publication/225026362_Processes_in_Biological_Vision.

Robert Geirhos, Carlos R. Medina Temme, Jonas Rauber, Heiko H. Schütt, Matthias Bethge, and Felix A. Wichmann. Generalisation in Humans and Deep Neural Networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 7547–7558. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper/2018/file/0937fb5864ed06ffb59ae5f9b5ed67a9-Paper.pdf.

Google. Google colaboratory. https://colab.research.google.com/. Accessed October 13, 2024.

Kalanit Grill-Spector and Rafael Malach. THE HUMAN VISUAL CORTEX. *Annual Review of Neuroscience*, 27:649–677, 2004. doi: 10.1146/annurev.neuro.27.070203.144220.

Jay Hegdé, Fang Fang, Scott O. Murray, and Daniel Kersten. Preferential responses to occluded objects in the human visual cortex. *Journal of Vision*, 8(4):16, 2008. doi: 10.1167/8.4.16.

Trevor Huff, Navid Mahabadi, and Prasanna Tadi. Neuroanatomy, Visual Cortex. In *StatPearls*. StatPearls Publishing, Treasure Island (FL), 2023. Updated 2023 Aug 14. Available from: https://www.ncbi.nlm.nih.gov/books/NBK482504/.

Jon H. Kaas. The Evolution of the Visual System in Primates. In Todd M. Preuss and Jon H. Kaas, editors, *Evolution of the Primate Brain*, pages 441–460. Academic Press, 2012. doi: 10.1016/B978-0-12-398315-7.00021-0.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 2012. Available from: https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

Jonas Kubilius, Martin Schrimpf, Kohitij Kar, Rishi Rajalingham, Ha Hong, N. Apurva Majaj, Elias Issa, Pouya Bashivan, Jonathan Prescott-Roy, Kailyn Schmidt, Daniel L. K. Yamins, and James J. DiCarlo. Brain-like object recognition with high-performing shallow recurrent anns. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pages 12785–12796. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/2019/file/941366319835c0652341977412193b9d-Paper.pdf.

Brenden M. Lake, Ruslan Salakhutdinov, and Joshua B. Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015. doi: 10.1126/science.aab3050.

Peter Lennie. The cost of cortical computation. *Current Biology*, 13(6):493–497, 2003.

Bingzhi Lin, Yuan Tian, Yue Zhang, Zhijing Zhu, and Depeng Wang. Deep learning methods for high-resolution microscale light field image reconstruction: a survey. *Frontiers in Bioengineering and Biotechnology*, 12:1500270, 2024. doi: 10.3389/fbioe.2024.1500270.

Xiaofeng Liu, Chaehwa Yoo, Fangxu Xing, Hyejin Oh, Georges El Fakhri, Je-Won Kang, and Jonghye Woo. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Transactions on Signal and Information Processing*, 1:1–48, 2022. Available from: https://arxiv.org/pdf/2208.07422.

Adam S. Lowet and Naoshige Uchida. Predictive coding: A distinction — without a difference. *Current Biology*, 34(20):R926–R929, 2024. doi: 10.1016/j.cub.2024.09.026.

Gabriele Cesa Maurice Weiler. General E(2)-equivariant steerable cnns. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pages 14066–14077. Curran Associates, Inc., 2019. URL https://arxiv.org/abs/1911.08251.

Abhilash Sunder Raj, Michael Lowney, Raj Shah, and Gordon Wetzstein. Stanford Lytro Light Field Archive. http://lightfields.stanford.edu/LF2016.html, 2016. Accessed on November 26, 2024.

Rajesh P. N. Rao and Dana H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999. doi: 10.1038/4580.

Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1xSgCEFvS.

Bhavin R. Sheth and Ryan Young. Two Visual Pathways in Primates Based on Sampling of Space: Exploitation and Exploration of Visual Information. *Frontiers in Integrative Neuroscience*, 10:37, 2016. doi: 10.3389/fnint.2016.00037.