Unsupervised learning on spontaneous retinal activity leads to efficient neural representation geometry

Andrew Ligeralde[1,2*] **Yilun Kuang**[2,3*] **Thomas Edward Yerxa**[2,4]
Miah N. Pitcher[5] **Marla Feller**[5,6] **SueYeon Chung**[2,4]
[1]Biophysics Graduate Group, University of California, Berkeley,
[2]Center for Computational Neuroscience, Flatiron Institute,
[3]Courant Inst. of Mathematical Sciences, New York University,
[4]Center for Neural Science, New York University,
[5]Helen Wills Neuroscience Institute, University of California, Berkeley,
[6]Department of Molecular and Cell Biology, University of California, Berkeley
*Equal contribution
ligeralde@berkeley.edu, yilun.kuang@nyu.edu

Editors: Marco Fumero, Emanuele Rodolà, Clementine Domine, Francesco Locatello, Gintare Karolina Dziugaite, Mathilde Caron

Abstract

Prior to the onset of vision, neurons in the developing mammalian retina spontaneously fire in correlated activity patterns known as retinal waves. Experimental evidence suggests that retinal waves strongly influence the emergence of sensory representations before visual experience. We aim to model this early stage of functional development by using movies of neurally active developing retinas as pre-training data for neural networks. Specifically, we pre-train a ResNet-18 with an unsupervised contrastive learning objective (SimCLR) on both simulated and experimentally-obtained movies of retinal waves, then evaluate its performance on image classification tasks. We find that pre-training on retinal waves significantly improves performance on tasks that test object invariance to spatial translation, while slightly improving performance on more complex tasks like image classification. Notably, these performance boosts are realized on held-out natural images even though the pre-training procedure does not include any natural image data. We then propose a geometrical explanation for the increase in network performance, namely that the spatiotemporal characteristics of retinal waves facilitate the formation of separable feature representations. In particular, we demonstrate that networks pre-trained on retinal waves are more effective at separating image manifolds than randomly initialized networks, especially for manifolds defined by sets of spatial translations. These findings indicate that the broad spatiotemporal properties of retinal waves prepare networks for higher order feature extraction.

1 Introduction

The visual system has an extraordinary capacity for rapidly and accurately recognizing distinct objects in the face of identity-preserving transformations [1, 2, 3]. Neural recordings suggest this is in part possible due to the high degree of linear separability between neural responses to different stimuli [4, 2]. Interestingly, deep convolutional neural networks (DCNNs) trained to classify images can perform invariant object categorization at near human-level accuracy [5] and have been shown to

Proceedings of the I edition of the Workshop on Unifying Representations in Neural Models (UniReps 2023).

exhibit representations similar to neural activities in mammalian systems [6, 7]. Furthermore, DCNN layers have analagous properties to the visual hierarchy, whereby feature transformations at each layer induce linear separability in the object manifolds [8]. DCNNs therefore offer a useful testbed for modeling the visual system [9, 10, 11, 12]. However, the supervised learning methods used to train these models are unlikely to explain how the brain learns object recognition, given that large amounts of labeled examples are not necessary for visual development [13, 14, 15, 16]. In this work, we explore the potential of innate neural activity as pre-training data for DCNNs and ask whether the internal representations that enable object recognition can be learned without access to any external visual information.

The motivation for this work is grounded in developmental neurobiology. Many key aspects of visual system organization are well-established before visual experience, such as topographic maps, orientation selectivity, and ocular dominance [17]. Notably, axon targeting can largely be learned by innately generated signals such as spontaneous neural activity and molecular guidance cues [18]. These findings suggest external stimuli are unnecessary for the initial development of the early visual system.

Here, we investigate whether a particular form of spontaneous activity known as retinal waves can instruct formation of the feed-forward connections that support object recognition. Retinal waves are a developmental phenomenon characterized by correlated patterns of propagating, network-level activity among groups of retinal ganglion cells (RGCs) prior to eye-opening [19]. Experimental and computational evidence suggests that retinal waves instruct the formation of retinotopic maps, enabling RGC axons to reach their targets in the superior colliculus and lateral geniculate nucleus before the onset of visual experience [20, 21, 22, 23, 24, 25].

Our core finding is that DCNNs pre-trained on movies of retinal waves produce more linearly separable representations of natural images compared to randomly initialized networks. To demonstrate this, we pre-train the hidden layers in a ResNet-18 classifier on calcium imaging movies of whole developing mouse retinas using the SimCLR [26] objective. This task-independent phase is meant to simulate the experience-independent period of visual development prior to eye-opening. We then evaluate network performance on a set of image classification tasks and find that networks pre-trained on retinal wave timecourses consistently outperform random controls. We explain this performance increase using the framework of manifold geometry [27]. Specifically, we characterize the geometry of the networks' internal feature representations [28] and find that networks pre-trained on retinal waves more effectively separate object manifolds. We also find the extent of separability is task-dependent and most pronounced for tasks that test spatial invariances. Our results suggest that the spatiotemporal information in retinal waves is relevant for object recognition in natural scenes and point towards an instructive role for retinal waves during early synapse formation in visual circuits.

2 Methods

2.1 Pre-training

To test whether spatiotemporal features of retinal waves learned during pre-training will improve performance on visual tasks, we follow the pipeline described in Fig. 1. Given a movie of a neurally active developing retina (Fig. 1A), we first train a ResNet-18 to compress temporally consecutive frames of the movie in output space, while pushing apart temporally distant frames (Fig. 1B) using the SimCLR training objective [26]. This is in accordance with the finding that temporally close activity bursts convey the most spatial information about relative RGC position [29, 30]. This phase is meant to simulate the period of cortical development *prior* to visual experience. We pre-train two kinds of networks: the first using macroscope movies of retinal waves obtained via calcium imaging of whole retinas dissected from postnatal mice (for experimental methods, see Supplementary Material S0), and the second using simulated movies of retinal waves from a parametrized, reaction-diffusion based model [31]. To isolate the effects of the spatial and temporal characteristics of retinal waves, we pre-train networks on three additional types of datasets created by modifying the original movies: **spatially shuffled**, in which the pixels of each frame are randomly permuted; **temporally shuffled**, in which the frame order is randomly permuted; and **spatiotemporally shuffled**, in which both the pixels of each frame and the frame order are randomly permuted (Fig. 1A). Spatially shuffled waves contain information about how the overall distribution of RGC activities changes over time, but lack the continuously varying spatial structure present in the original movies. As such,

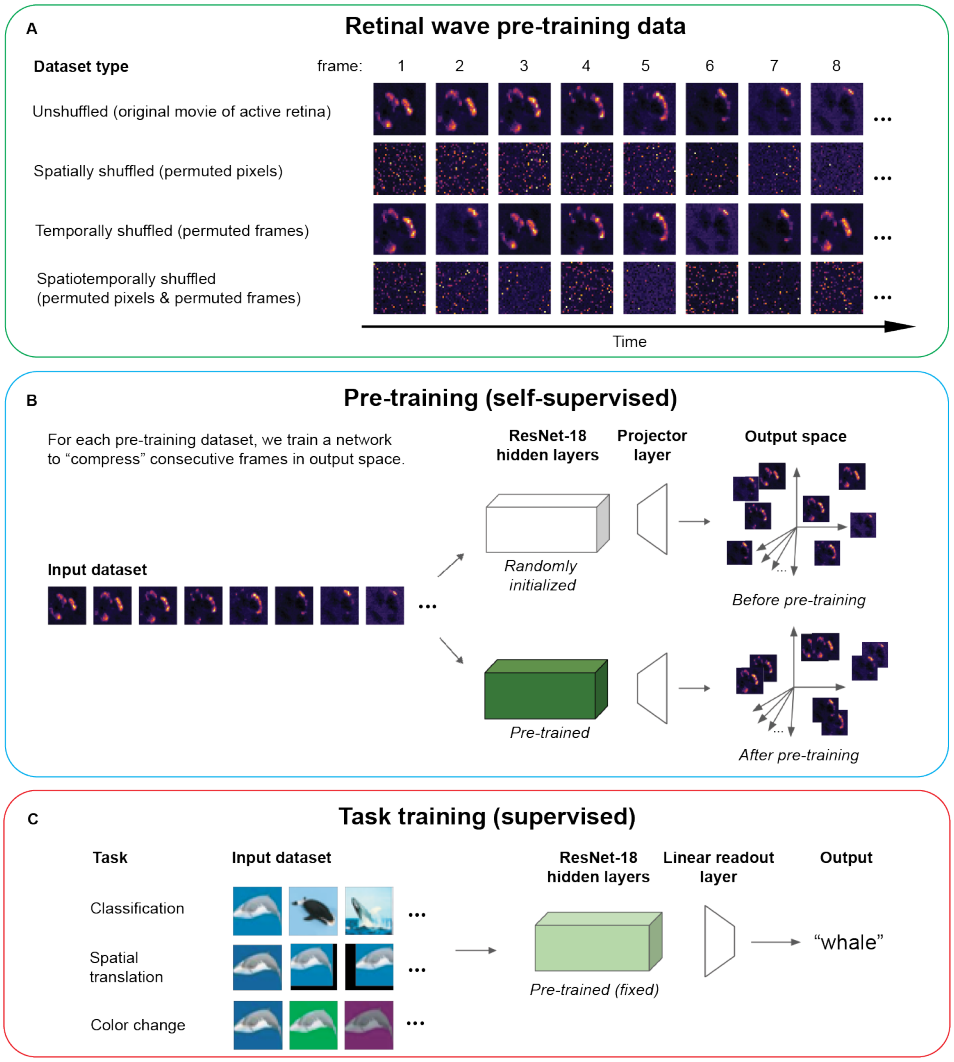

Figure 1: **Network training pipeline.** (A) Retinal wave movies and three permutations of the original movies are used as pre-training datasets (B) contrastive learning to learn temporally close spatial correlations (C) evaluation of network performance on three labeling tasks.

spatially shuffled pre-training controls for how much task information can be inferred only through temporally local changes in the population statistics of RGC activity. Comparing temporally shuffled and spatiotemporally shuffled waves controls for the amount of task-relevant, temporally non-local information in retinal waves. If correlations between temporally distant frames are relevant for a given task, networks trained on temporally shuffled waves should perform better than those trained on spatiotemporally shuffled waves. We compare all pre-training conditions to a He random initialized control network that has not been pre-trained, for a total of nine conditions.

Preprocessing: To filter out calcium transients, periods of inactivity, and random noise in the calcium imaging data, watershed image segmentation is used to identify periods of continuous retinal wave activity spanning a given number of frames, with each period denoted as a "wave event". We use publicly available code for watershed segmentation from `https://github.com/Llamero/Feller_Retinal_Wave_Analysis`. We aggregate movies from four retinas, resulting in $\sim$60,000 total frames of real retinal wave pre-training data. Simulated retinal wave data is generated using the model in [31] "out-of-the-box". The area parameter of the simulation is changed to match the area of the isolated real retinas, and the "strength" parameter α is modified to 0.5 to increase the wave frequency and eliminate long periods of inactivity. The model frame rate is matched to that of the macroscope data. The model is run to obtain a total of $\sim$237,000 frames of simulation data. Because

the simulated data is far less noisy than the real data, wave events are simply taken as the sets of frames in between periods of cell inactivity, without the need for image segmentation.

Hyperparameters: Networks are pre-trained with a projector layer [32] of dimensions $8192 \times 8192 \times 8192$ for 100 epochs with a learning rate of 0.0001 and Adam optimization based on a grid hyperparameter search. Because wave events occur for varying lengths of time, batches are formed by randomly sampling whole wave events from the movie until the total number of sampled frames exceeds a threshold value of 3000. Positive examples are defined as consecutive frames within the same wave event, and negative examples are defined as all frames outside of that wave event.

2.2 Task training

To test the effects of pre-training on task performance, we add a linear readout layer to the pre-trained weights and train linear readout layer weights on labeled images while leaving the pre-trained hidden layer weights fixed (Fig. 1C). This phase is meant to simulate a test of the functionality gained from retinal wave activity at the onset of visual experience. We use this procedure to evaluate network performance on three labeling tasks. We report the mean and standard deviations for test accuracy across the three seeded random network initializations.

Classification task: The first task is standard image classification on CIFAR-10.

Spatial translation task: For the second task, we train networks to classify spatially translated images drawn from CIFAR-100. To generate the task data, we first choose 10 of 100 classes at random and draw a random image from each class, which we denote as a "base" image. An image in the task dataset is then generated as a random affine transformation (up to 16 pixels in the x and y directions) of one of the 10 base images. Using this procedure, each base image is used to generate 5000 training images and 1000 test images, for a total of 50,000 training images and 10,000 test images. The networks are trained to classify a given training image with the label of its original base image.

Color change task: For the third task, we train networks to classify recolorations of the same 10 base images used in the spatial translation task. The task data is generated by the same procedure, only instead of random affine transformations, we apply random color transformations to the base image that range from 50 to 100% changes in saturation, brightness, contrast, and hue. The networks are trained to classify a given training image with the label of its original base image.

Hyperparameters: In task training, the projector dimension used in pre-training is removed and replaced with a 512×10 linear readout layer [32]. The readout layer is trained for 100 epochs, batches of size 100, and learning rate of 0.0001 on 50,000 labeled training images. The performance is evaluated on 10,000 labeled test images.

2.3 Manifold analysis

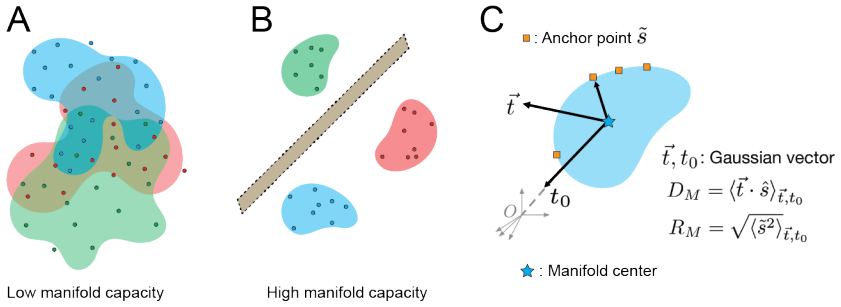

Figure 2: **Illustration of point cloud manifolds**. (A) Tangled manifolds exhibit low capacity (B) untangled manifolds exhibit high capacity and are separable by a hyperplane (C) manifold dimension measures spread of anchor points across the manifold axes by projection of a Gaussian vector onto an anchor point. Manifold radius measures the norm of an anchor point in the manifold subspace.

The set of neural responses to different presentations of a given stimulus define a neural object manifold. The linear separability of these manifolds as a function of their geometry enables discrimination

between stimuli [27]. We use this framework to characterize how pre-training on retinal waves changes the geometry of representation. We examine three quantities of manifolds that determine their separability, namely the capacity α_c, the dimension D_M, and the radius R_M.

Capacity α_c: We consider a set of P object manifolds linearly separable if they can be classified into binary classes by a hyperplane in N-dimensional feature space. The theory of manifold geometry shows that the value of the manifold capacity α_c determines the extent of separability in the limit of large P and N: if $P/N < \alpha_c$, the manifolds are separable with high probability; if $P/N > \alpha_c$, the manifolds are inseparable with high probability. Therefore, the higher the value of α_c, the higher the probability of separability for a given set of manifolds (Figs. 2A,B). For point-cloud manifolds, in which each manifold consists of M data points each corresponding to an example of the given object, the capacity can be shown to be bounded as $\frac{2}{M} \leq \alpha_c \leq 2$ [28]. The theory of manifold geometry also shows that capacity is determined by two quantities which describe the geometry of the object manifolds in N-space: the dimension D_M and the radius R_M. These are statistical quantities defined for each manifold by considering the spread of points in the manifold's convex hull, called anchor points, over variations in the manifold's labeling and location in N-space (Fig. 2C). For large N, α_c is inversely proportional to $\sqrt{D_M}$ and R_M [33]. All three quantities — α_c, D_M, and R_M — are estimated using algorithms based on statistical mechanical mean-field techniques described in [34].

Dimension D_M: Dimension is the spread of anchor points across the manifold axes and estimates the average embedding dimension of the manifold (Fig. 2C).

Radius R_M: Radius is the average distance between the manifold center and anchor points and reflects the scale of the manifold compared to the overall data distribution. (Fig. 2C).

Simulation capacity α_{sim}: We note that α_c is a theoretical estimate of linear separability that may deviate from the true capacity in the regime of finite manifolds P and feature dimensions N [28]. Simulation capacity provides a numerical approximation of the ground-truth manifold capacity. We calculate simulation capacity by first running linear classifications with fixed P and varying N until the probability of manifold separation converges to 0.5. The final value of $N = N_c$ is used to calculate the simulation capacity $\alpha_{sim} = P/N_c$. We report the correspondence between α_c and simulation capacity in Fig. S2.

Task data manifolds: To examine how pre-training with retinal waves affects the geometry, and in turn the separability, of neural object manifolds for each task, we extract the network activations at each ReLU layer for $P = 50$ manifolds consisting of $M = 20$ examples. For **standard classification**, each manifold corresponds to an image class in CIFAR-100. Examples for each manifold are drawn from the given class based on the ranked 20-highest softmax probability scores output by a well-trained classifier. For both **spatial translation** and **color change**, each manifold corresponds to one random base image drawn from CIFAR-100. Examples for each spatial translation manifold are generated by applying random affine shifts up to 3 pixels in both directions to the base image. Examples for each color change manifold are generated by applying random $50 - 150\%$ changes in saturation, brightness, hue, and contrast to the base image. We also measure the capacity and geometry of the manifolds defined by retinal waves, where each manifold consists of frames belonging to a given wave event (Fig. S1).

We report the mean and standard deviations for all manifold quantities across three seeded random network initializations. For all manifold analysis, we use publicly available code from `https://github.com/schung039/neural_manifolds_replicaMFT`.

Pre-training, task training, and manifold analysis was done on an internal cluster using NVIDIA 16 GB V100 (Volta) GPUs. All code for pre-processing, pre-training, task training, and analysis is available at `https://github.com/chung-neuroai-lab/retinal_waves_learning`.

3 Results

3.1 Pre-training on retinal waves improves task performance

Our main result is that self-supervised pre-training of networks on movies of retinal waves improves object separability for labeled natural images. We find that pre-training on the original, unshuffled wave movies yields the highest performance increase in the spatial translation task (Fig. 3). This suggests that retinal waves contain information that supports learning object invariance to spatial

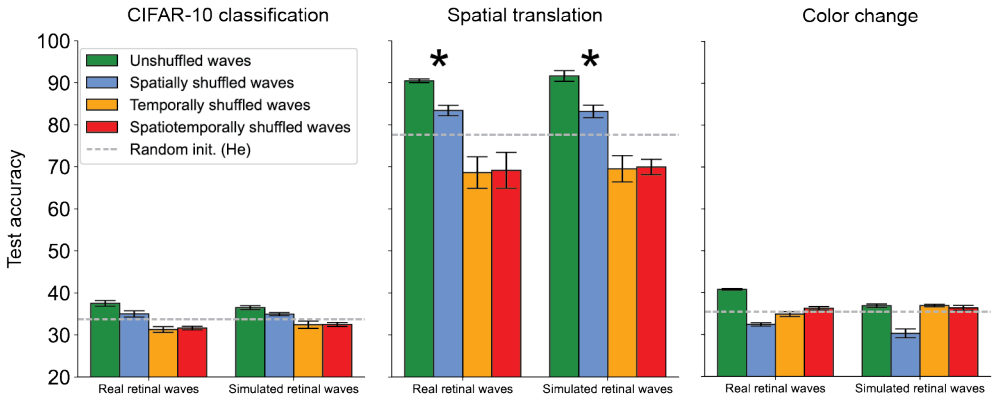

Figure 3: **Test accuracy for pre-trained networks in three labeling tasks.** Asterisks indicate that the performance increase from pre-training on retinal waves is highest for the spatial translation task.

translation. Pre-training on spatially shuffled waves yields a moderate improvement above random initialization in this task, suggesting that learning temporally local changes in the overall distribution of activities is also relevant for this function. Destroying the temporal structure of the waves, however, yields performance below random initialization, as shown in the temporally and spatiotemporally shuffled pre-training conditions. This suggests that temporally local, rather than global correlations in retinal waves are most relevant for learning spatial invariance. This is consistent with the previous finding that little information is gained by considering RGC activity bursts more than 3 sec (around 35 frames) apart [29, 30]. These networks perhaps even learn non-local features that actually hinder task learning, as suggested by their below-random-network performance. We further explore this idea in Sec. 3.3.

Classification is a far more complex task than spatial translation as it requires mapping visual information onto higher level semantic structures, information not present retinal waves. Accordingly, performance for this task is significantly lower for pre-trained networks overall than for spatial translation. However, networks trained on unshuffled waves still perform slightly better than the others (Fig. 3). A similar trend emerges for the color change task, for which we also did not expect pre-training to yield any advantage. A potential reason for the performance increases in both cases is the persistence of similar features across examples in the same class. Visual patterns like edges and curves are features that retinal waves may train the visual system to recognize [35]. We further explore reasons for these small performance boosts in 3.3.

While accuracy provides a proxy for the task-specific relevance of retinal waves, it does not give insight into how retinal waves influence learned feature representations. In the next section, we address this question by examining the geometry of task object manifolds across pre-training conditions.

3.2 Pre-training on retinal waves increases separability for manifolds defined by invariance to spatial translation

Previous work shows that DCNNs trained to classify images increase the object manifold capacity from the input to output layers [8]. We only observe this behavior for the spatial translation manifold. Consistent with the accuracy results, networks trained on unshuffled waves and spatially shuffled waves yield increases in capacity relative to randomly initialized networks, while networks trained on temporally and spatiotemporally shuffled waves do not substantially change the capacity between the input and output layers (Fig. 4). As expected, the spatial translation manifolds in networks pre-trained on unshuffled waves also have lower dimension and radius compared to those in the other networks, while networks pre-trained on spatially shuffled waves only appear to decrease the radius (Fig. 5). These results suggest that pre-training on retinal waves has a direct influence on the geometry and separability of neural object manifolds for tasks that involve learning spatial invariance.

In all networks, the capacity of the CIFAR class manifold (see inset, Fig. 4) remains nearly constant around the theoretical lower bound of 0.1 (Sec. 2.3). All networks also yield a decrease in capacity

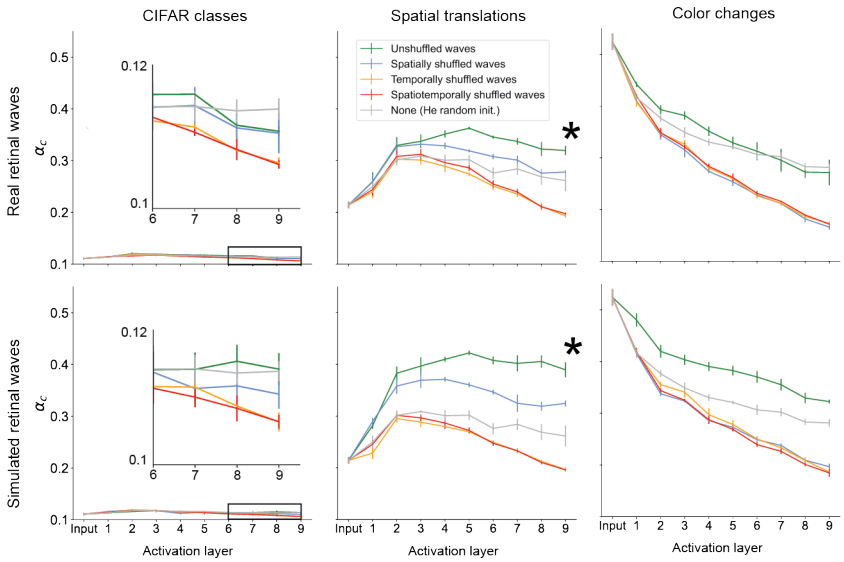

Figure 4: **Changes in classification capacity over network layers.** Asterisks indicate that the capacity of spatial translation manifolds increases the most along the hierarchy of the network pre-trained on unshuffled retinal waves.

for the color change manifold at each successive layer (Fig. 4). (Although the network trained on simulated unshuffled waves appears to have a relatively high capacity for the color change manifold, this particular value actually overestimates the ground truth simulation capacity, which we show in Fig. S2). The dimensions and radii of the CIFAR class and color change manifolds also do not show any consistent ordering that points to a clear advantage of pre-training on retinal waves relative to the random baseline (Fig. 5). These results are consistent with the poor accuracy in these tasks across all networks. However, if pre-training does not substantially affect these object manifolds, what accounts for the slight boost in performance on these tasks for the networks pre-trained on unshuffled waves? To address this question, we explore two factors external to the geometry of individual manifolds, namely the inter-manifold correlation and the effective dimensionality of the feature space.

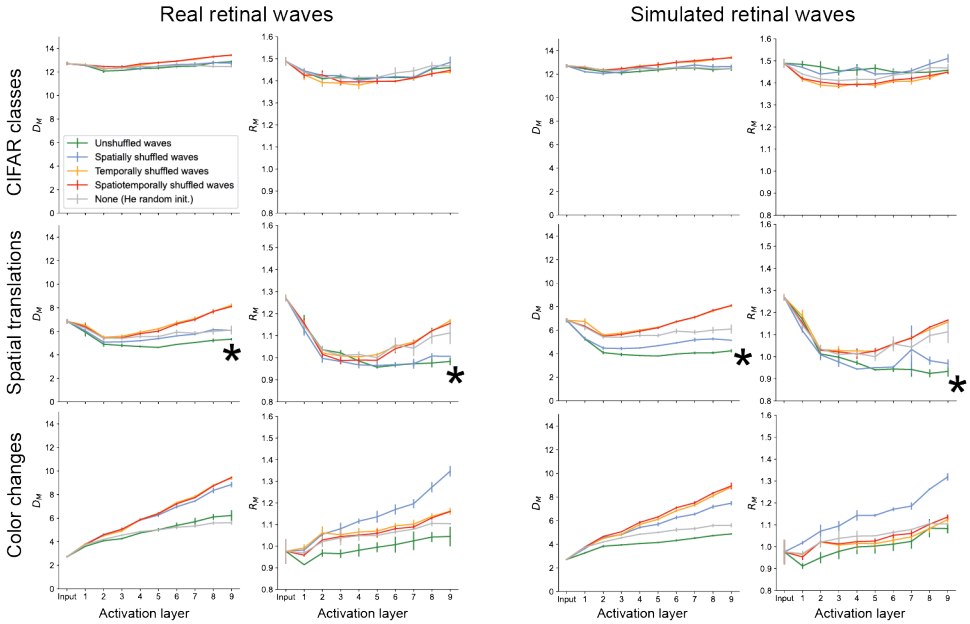

Figure 5: **Changes in manifold geometry over network layers.** Asterisks indicate that networks pre-trained on unshuffled retinal waves most effectively compress spatial translation manifolds.

3.3 Pre-training on retinal waves decreases inter-manifold correlations and maintain effective dimensionality

A high degree of correlation between manifold centers may lead to clustering of object manifolds in feature space, making them more difficult to separate and decreasing the effective capacity. Previous work demonstrates that training DCNNs leads to decorrelation of the manifold centers [8]. Here, we measure the pairwise correlation coefficient between manifold centers at each network layer and find that networks pre-trained on unshuffled retinal waves decrease center correlations relative to randomly initialized networks and networks pre-trained on spatially shuffled waves for all three tasks (Fig. 6). Unshuffled pre-training also leads to a generally consistent decrease in correlation along at each successive network layer. Interestingly, temporally and spatiotemporally shuffled pre-training also produce networks that exhibit this behavior, in addition to having lower correlations than in the unshuffled case. However, based on their poor task performance and low capacities of their feature representations, it is likely this is simply due to the explosion in dimensionality of their respective feature spaces, which we discuss next.

Ideally, a well-trained classifier will extract the features that correspond to the highest sources of variance in the data, while separating out low-variance features that do not correspond to meaningful distinctions between samples. Participation ratio (PR) varies from 1 to N and measures how data variance is spread out across the feature dimensions: if $PR = 1$, the variance is concentrated entirely in one feature; if $PR = N$, the variance is spread out evenly across all features [36]. In general, a good classifier will maintain a $PR > 1$ in the feature dimensions so as not to destroy the structure in the data, while also keeping $PR < N$ so as to preserve only the meaningful features that capture the latent dimensionality of the data. The layer-wise participation ratio suggests that networks pre-trained on unshuffled waves maintain this happy medium in all three tasks (Fig. 6). Networks pre-trained on spatially shuffled waves decrease participation ratio to near the lower bound, consistent with the idea that they broadly capture population-level statistics, but fail to learn many spatially local features that likely lie along other dimensions. The large increase in PR observed in networks trained on temporally and spatiotemporally shuffled waves suggests that they do in fact learn features that are not relevant for the task dataset, as proposed in Section 3.1. These extraneous features would account for the increase in PR above the values observed in other networks. Notably, correlation and PR are inversely related, suggesting that high effective dimensionality is a factor in separation of manifold centers. The trends observed in PR are consistent with the trends in layer-wise explained variance, which measures how many feature dimensions account for a given percentage of variance in the data (Fig. S3).

4 Discussion

To our knowledge, this is the first computational work that directly explores how real retinal waves can influence neural object representations, demonstrating a bioplausible means of learning spatial invariance without training on large datasets of labeled images. While DCNNs trained on labeled images achieve state-of-the-art performance and even predict neural responses [6, 37], these models are unlikely to explain how biological vision develops. Unsupervised and self-supervised learning mechanisms have therefore been proposed as biologically plausible means of learning object recognition [16]. However, standard implementations of these algorithms still require natural images or videos as training inputs, which effectively simulate a visual experience. Though visual experience certainly shapes cortical functional development [38, 39, 40, 41], models that wholly rely on image data do not account for the functionality, connectivity, and feature selectivity already observed in animals prior to the onset of vision [42, 43, 44, 45, 46]. Consistent with our results, previous work has demonstrated that self-supervised learning on structured noise can improve classification accuracy on unseen images [47, 48, 49]. Additionally, simulated retinal waves have been shown to yield V1-like receptive fields when used as inputs for sparse coding algorithms [35, 24, 50] and slow feature analysis [51].

We demonstrate that pre-training on retinal waves has two primary effects on learned representations that can account for increases in task performance. The first is an increase in the separability of individual object manifolds. This effect is pronounced in the spatial translation task, suggesting that the spatiotemporal characteristics of retinal waves train networks to learn spatial translation invariance. To show this, we analyze the geometry of the neural object manifolds defined by affine transformations of a single object (image) and find they are more linearly separable when represented

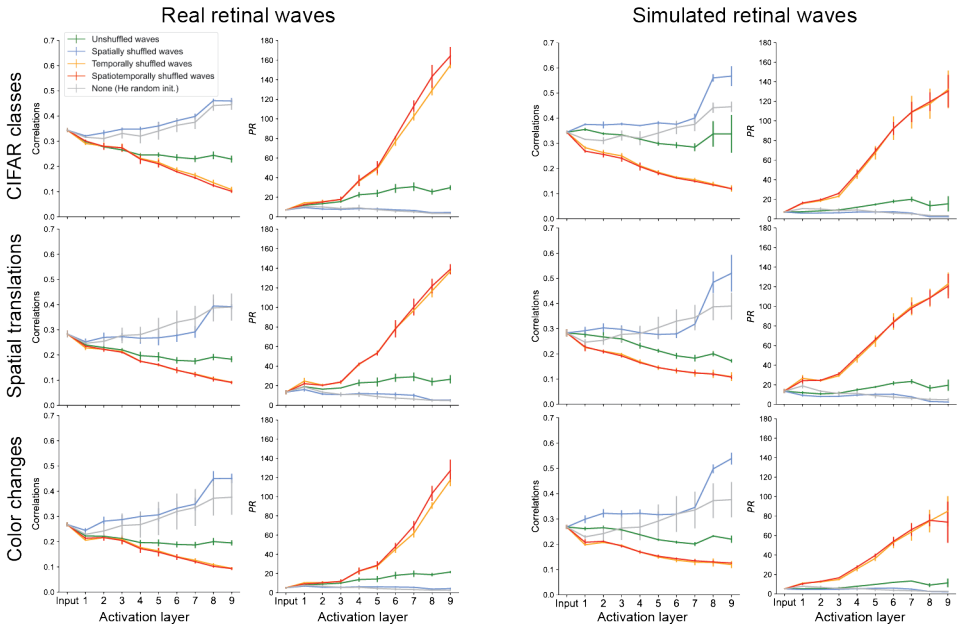

Figure 6: **Changes in inter-manifold correlation and participation ratio along network layers**. Only the network pre-trained on unshuffled waves consistently reduces correlation and avoids vanishing/exploding dimensionality.

in networks pre-trained on unshuffled retinal waves (Figs. 4, 5). Both the spatial and temporal characteristics of retinal waves are necessary for learning this task, as pre-training on spatially and/or temporally shuffled retinal waves leads to poor separability of spatial translation manifolds. Pre-training does not have a significant effect on the separability of the manifolds defined by CIFAR image classes or color changes of a single object (Figs. 4, 5), suggesting a qualitative bound on the scope of tasks for which retinal waves are useful training signals.

We also observe that pre-training on retinal waves reduces center correlations between neural object manifolds and increases the effective dimensionality of the feature space (Figs. 6). Both effects are directly correlated with linear separability and appear to be independent of the effect on individual manifold separability, as they are observed in all three tasks.

Together, these two effects of pre-training on retinal waves correspond to distinct *local* and *global* mechanisms of transforming object representations, both of which are important for separability. At the local level, pre-training increases the compressibility of individual neural object manifolds, as shown in the increase in capacity and the concurrent decreases in dimension and radius. At the global level, pre-training places neural object manifolds in higher dimensional feature space, as shown by the increase in participation ratio and concurrent decrease in center correlation. These two regimes point to distinct ways in which retinal waves may influence emerging sensory representations.

We do not observe a significant difference between pre-training on real versus simulated retinal waves from the model. The advantage of the model is that we can generate an arbitrarily large set of pre-training data, at the risk of introducing free parameters that may lead to deviations from real data. Though we do not perform a direct comparison between the simulated and real data in this work, no clear difference emerges between these two datasets in terms of model performance or the geometry of the object representations. This suggests that for the tasks considered, the common features of these datasets — such as spatiotemporal continuity between frames — are the primary drivers of the observed effects. In future work, the model may be a useful tool for examining the effect of changing the waves' spatiotemporal characteristics on representation learning.

We note that our findings are subject to our choice of network architecture (ResNet-18), learning algorithm (SimCLR), and dataset (postnatal mouse retinal waves). Retinal waves occur during multiple stages of development [52] and drive formation of visual circuitry in numerous ways [19]. Retinal waves are also not the only form of spontaneous activity during development [53]. Along this line of work, future studies may consider the role of cortical feedback [54], introduce bioplausible,

synaptically local learning rules [55], or investigate the role of spontaneous activity in other modalities like temporal prediction [56]. Additionally, laboratory experiments that test object recognition in mice [57] performed at the onset of vision could verify our model predictions and provide richer insight into the capacity of neural object manifolds during this early developmental period.

Acknowledgments and Disclosure of Funding

This work was funded by the Center for Computational Neuroscience at the Flatiron Institute of the Simons Foundation. AL was supported by the Flatiron Machine Learning X Science Summer School and the National Science Foundation under grant DGE 1752814. YK was supported by the NYU Center for Data Science Fellowship. MNP and MF were supported by the National Institutes of Health under grands NIH RO1EY013528, RO1EY019498, and P30EY003176. TY has no competing interests or relevant funding to declare. SC was supported by the Klingenstein-Simons Award. All experiments were performed on the Flatiron Institute high-performance computing cluster.

References

[1] Simon Thorpe, Denis Fize, and Catherine Marlot. Speed of processing in the human visual system. *Nature*, 381(6582):520–522, June 1996. doi: 10.1038/381520a0. URL https://doi.org/10.1038/381520a0.

[2] James J. DiCarlo, Davide Zoccolan, and Nicole C. Rust. How Does the Brain Solve Visual Object Recognition? *Neuron*, 73(3):415–434, February 2012. ISSN 0896-6273. doi: 10.1016/j.neuron.2012.01.010. URL https://www.sciencedirect.com/science/article/pii/S089662731200092X.

[3] Rishi Rajalingham, Kailyn Schmidt, and James J. DiCarlo. Comparison of object recognition behavior in human and monkey. *The Journal of Neuroscience*, 35(35):12127–12136, September 2015. doi: 10.1523/jneurosci.0573-15.2015. URL https://doi.org/10.1523/jneurosci.0573-15.2015.

[4] Chou P. Hung, Gabriel Kreiman, Tomaso Poggio, and James J. DiCarlo. Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866, November 2005. doi: 10.1126/science.1117593. URL https://doi.org/10.1126/science.1117593.

[5] Charles F. Cadieu, Ha Hong, Daniel L. K. Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj, and James J. DiCarlo. Deep neural networks rival the representation of primate IT cortex for core visual object recognition. *PLoS Computational Biology*, 10(12): e1003963, December 2014. doi: 10.1371/journal.pcbi.1003963. URL https://doi.org/10.1371/journal.pcbi.1003963.

[6] Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, May 2014. doi: 10.1073/pnas.1403112111. URL https://doi.org/10.1073/pnas.1403112111.

[7] H Wen, J Shi, W Chen, and Z Liu. Deep residual network predicts cortical representation and organization of visual features for rapid categorization. sci. rep. 8 (1), 3752 (2018), 2018.

[8] Separability and geometry of object manifolds in deep neural networks. 11. ISSN 2041-1723.

[9] Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, 10(11):e1003915, November 2014. doi: 10.1371/journal.pcbi.1003915. URL https://doi.org/10.1371/journal.pcbi.1003915.

[10] Saeed Reza Kheradpisheh, Masoud Ghodrati, Mohammad Ganjtabesh, and Timothée Masquelier. Deep networks can resemble human feed-forward vision in invariant object recognition. *Scientific Reports*, 6(1), September 2016. doi: 10.1038/srep32672. URL https://doi.org/10.1038/srep32672.

[11] Daniel L K Yamins and James J DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, February 2016. doi: 10.1038/nn.4244. URL https://doi.org/10.1038/nn.4244.

[12] Haiguang Wen, Junxing Shi, Wei Chen, and Zhongming Liu. Deep residual network predicts cortical representation and organization of visual features for rapid categorization. *Scientific Reports*, 8(1), February 2018. doi: 10.1038/s41598-018-22160-9. URL https://doi.org/10.1038/s41598-018-22160-9.

[13] Elika Bergelson and Daniel Swingley. At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, 109(9):3253–3258, February 2012. doi: 10.1073/pnas.1113380109. URL https://doi.org/10.1073/pnas.1113380109.

[14] Elika Bergelson and Richard N. Aslin. Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, 114(49):12916–12921, November 2017. doi: 10.1073/pnas.1712966114. URL https://doi.org/10.1073/pnas.1712966114.

[15] Michael C Frank, Mika Braginsky, Daniel Yurovsky, and Virginia A Marchman. *Variability and consistency in early language learning: The Wordbank project*. MIT Press, 2021.

[16] Chengxu Zhuang, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel L. K. Yamins. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3):e2014196118, 2021. doi: 10.1073/pnas.2014196118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2014196118.

[17] J. Sebastian Espinosa and Michael P. Stryker. Development and Plasticity of the Primary Visual Cortex. *Neuron*, 75(2):230–249, July 2012. ISSN 08966273. doi: 10.1016/j.neuron.2012.06.009. URL https://linkinghub.elsevier.com/retrieve/pii/S0896627312005697.

[18] M. B. Feller and D. Kerschensteiner. Chapter 16 - Retinal waves and their role in visual system development. In John Rubenstein, Pasko Rakic, Bin Chen, Kenneth Y. Kwan, Hollis T. Cline, and Jessica Cardin, editors, *Synapse Development and Maturation*, pages 367–382. Academic Press, January 2020. ISBN 978-0-12-823672-7. doi: 10.1016/B978-0-12-823672-7.00016-8. URL https://www.sciencedirect.com/science/article/pii/B9780128236727000168.

[19] David A. Arroyo and Marla B. Feller. Spatiotemporal Features of Retinal Waves Instruct the Wiring of the Visual Circuitry. *Frontiers in Neural Circuits*, 10:54, July 2016. ISSN 1662-5110. doi: 10.3389/fncir.2016.00054. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4960261/.

[20] Jianhua Cang, René C. Rentería, Megumi Kaneko, Xiaorong Liu, David R. Copenhagen, and Michael P. Stryker. Development of precise maps in visual cortex requires patterned spontaneous activity in the retina. *Neuron*, 48(5):797–809, December 2005. ISSN 0896-6273. doi: 10.1016/j.neuron.2005.09.015.

[21] Anand R. Chandrasekaran, Daniel T. Plas, Ernesto Gonzalez, and Michael C. Crair. Evidence for an instructive role of retinal activity in retinotopic map refinement in the superior colliculus of the mouse. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(29):6929–6938, July 2005. ISSN 1529-2401. doi: 10.1523/JNEUROSCI.1470-05.2005.

[22] Andrew D. Huberman, Colenso M. Speer, and Barbara Chapman. Spontaneous Retinal Activity Mediates Development of Ocular Dominance Columns and Binocular Receptive Fields in V1. *Neuron*, 52(2):247–254, October 2006. ISSN 0896-6273. doi: 10.1016/j.neuron.2006.07.028. URL https://www.cell.com/neuron/abstract/S0896-6273(06)00625-8. Publisher: Elsevier.

[23] Jeffrey Markowitz, Yongqiang Cao, and Stephen Grossberg. From Retinal Waves to Activity-Dependent Retinogeniculate Map Development. *PLOS ONE*, 7(2):e31553, February 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0031553. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0031553. Publisher: Public Library of Science.

[24] Jonathan J. Hunt, Michael Ibbotson, and Geoffrey J. Goodhill. Sparse Coding on the Spot: Spontaneous Retinal Waves Suffice for Orientation Selectivity. *Neural Computation*, 24 (9):2422–2433, September 2012. ISSN 0899-7667. doi: 10.1162/NECO_a_00333. URL https://doi.org/10.1162/NECO_a_00333.

[25] Ben Jiwon Choi, Yu-Chieh David Chen, and Claude Desplan. Building a circuit through correlated spontaneous neuronal activity in the developing vertebrate and invertebrate visual systems. *Genes & Development*, 35(9-10):677–691, May 2021. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.348241.121. URL http://genesdev.cshlp.org/lookup/doi/10.1101/gad.348241.121.

[26] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 13–18 Jul 2020.

[27] 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 09594388. doi: 10.1016/j.conb.2021.10.010. URL http://arxiv.org/abs/2104.07059. arXiv: 2104.07059.

[28] SueYeon Chung, Daniel D. Lee, and Haim Sompolinsky. Classification and Geometry of General Perceptual Manifolds. *Physical Review X*, 8(3):031003, July 2018. ISSN 2160-3308. doi: 10.1103/PhysRevX.8.031003. URL https://link.aps.org/doi/10.1103/PhysRevX.8.031003.

[29] Daniel A. Butts and Daniel S. Rokhsar. The information content of spontaneous retinal waves. *The Journal of Neuroscience*, 21(3):961–973, February 2001. doi: 10.1523/jneurosci.21-03-00961.2001. URL https://doi.org/10.1523/jneurosci.21-03-00961.2001.

[30] Daniel A. Butts. Retinal waves: Implications for synaptic learning rules during development. *The Neuroscientist*, 8(3):243–253, June 2002. doi: 10.1177/1073858402008003010. URL https://doi.org/10.1177/1073858402008003010.

[31] Benjamin Lansdell, Kevin Ford, and J. Nathan Kutz. A reaction-diffusion model of cholinergic retinal waves. *PLoS Computational Biology*, 10(12):e1003953, December 2014. doi: 10.1371/journal.pcbi.1003953. URL https://doi.org/10.1371/journal.pcbi.1003953.

[32] 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.

[33] SueYeon Chung. Statistical Mechanics of Neural Processing of Object Manifolds. Technical Report arXiv:2106.00790, arXiv, June 2021. URL http://arxiv.org/abs/2106.00790. arXiv:2106.00790 [cond-mat, q-bio] type: article.

[34] Cory Stephenson, Jenelle Feather, Suchismita Padhy, Oguz H. Elibol, Hanlin Tang, Josh H. McDermott, and SueYeon Chung. Untangling in invariant speech recognition. In *Neural Information Processing Systems*, 2020.

[35] Mark V Albert, Adam Schnabel, and David J Field. Innate Visual Learning through Spontaneous Activity Patterns. *PLoS Computational Biology*, 4(8):8, 2008.

[36] Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. A theory of multineuronal dimensionality, dynamics and measurement. November 2017. doi: 10.1101/214262. URL https://doi.org/10.1101/214262.

[37] 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? September 2018. doi: 10.1101/407007. URL https://doi.org/10.1101/407007.

[38] Michael Pecka, Yunyun Han, Elie Sader, and Thomas D. Mrsic-Flogel. Experience-dependent specialization of receptive field surround for selective coding of natural scenes. *Neuron*, 84 (2):457–469, October 2014. doi: 10.1016/j.neuron.2014.09.010. URL https://doi.org/10.1016/j.neuron.2014.09.010.

[39] Giulio Matteucci and Davide Zoccolan. Unsupervised experience with temporal continuity of the visual environment is causally involved in the development of v1 complex cells. *Science Advances*, 6(22), May 2020. doi: 10.1126/sciadv.aba3742. URL https://doi.org/10.1126/sciadv.aba3742.

[40] Nina N. Kowalewski, Janne Kauttonen, Patricia L. Stan, Brian B. Jeon, Thomas Fuchs, Steven M. Chase, Tai Sing Lee, and Sandra J. Kuhlman. Development of natural scene representation in primary visual cortex requires early postnatal experience. *Current Biology*, 31(2):369–380.e5, January 2021. doi: 10.1016/j.cub.2020.10.046. URL https://doi.org/10.1016/j.cub.2020.10.046.

[41] Nana Nishio, Kenji Hayashi, Ayako Wendy Ishikawa, and Yumiko Yoshimura. The role of early visual experience in the development of spatial-frequency preference in the primary visual cortex. *The Journal of Physiology*, 599(17):4131–4152, August 2021. doi: 10.1113/jp281463. URL https://doi.org/10.1113/jp281463.

[42] H. Sherk and M. P. Stryker. Quantitative study of cortical orientation selectivity in visually inexperienced kitten. *Journal of Neurophysiology*, 39(1):63–70, January 1976. doi: 10.1152/jn.1976.39.1.63. URL https://doi.org/10.1152/jn.1976.39.1.63.

[43] H. Ko, T. D. Mrsic-Flogel, and S. B. Hofer. Emergence of feature-specific connectivity in cortical microcircuits in the absence of visual experience. *Journal of Neuroscience*, 34(29): 9812–9816, July 2014. doi: 10.1523/jneurosci.0875-14.2014. URL https://doi.org/10.1523/jneurosci.0875-14.2014.

[44] James B Ackman and Michael C Crair. Role of emergent neural activity in visual map development. *Current Opinion in Neurobiology*, 24:166–175, February 2014. ISSN 09594388. doi: 10.1016/j.conb.2013.11.011. URL https://linkinghub.elsevier.com/retrieve/pii/S0959438813002225.

[45] Hong-Ping Xu, Timothy J. Burbridge, Meijun Ye, Minggang Chen, Xinxin Ge, Z. Jimmy Zhou, and Michael C. Crair. Retinal wave patterns are governed by mutual excitation among starburst amacrine cells and drive the refinement and maintenance of visual circuits. *The Journal of Neuroscience*, 36(13):3871–3886, March 2016. doi: 10.1523/jneurosci.3549-15.2016. URL https://doi.org/10.1523/jneurosci.3549-15.2016.

[46] Alexandre Tiriac, Karina Bistrong, and Marla B. Feller. Retinal waves but not visual experience are required for development of retinal direction selectivity maps. Technical report, bioRxiv, March 2021. URL https://www.biorxiv.org/content/10.1101/2021.03.25.437067v1. Section: New Results Type: article.

[47] Guruprasad Raghavan and Matt Thomson. Neural networks grown and self-organized by noise. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/1e6e0a04d20f50967c64dac2d639a577-Paper.pdf.

[48] Guruprasad Raghavan, Cong Lin, and Matt Thomson. Self-organization of multi-layer spiking neural networks, 2020. URL https://arxiv.org/abs/2006.06902.

[49] Manel Baradad, Jonas Wulff, Tongzhou Wang, Phillip Isola, and Antonio Torralba. Learning to see by looking at noise. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=RQUl8gZnN7O.

[50] Sahar Behpour, David J. Field, and Mark V. Albert. On the Role of LGN/V1 Spontaneous Activity as an Innate Learning Pattern for Visual Development. *Frontiers in Physiology*, 12, 2021. ISSN 1664-042X. URL https://www.frontiersin.org/article/10.3389/fphys.2021.695431.

[51] Sven Dähne, Niko Wilbert, and Laurenz Wiskott. Slow Feature Analysis on Retinal Waves Leads to V1 Complex Cells. *PLoS Computational Biology*, 10(5):e1003564, May 2014. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1003564. URL https://dx.plos.org/10.1371/journal.pcbi.1003564.

[52] Christiane Voufo, Andy Quaen Chen, Benjamin E Smith, Rongshan Yan, Marla B Feller, and Alexandre Tiriac. Circuit mechanisms underlying embryonic retinal waves. *eLife*, 12, February 2023. doi: 10.7554/elife.81983. URL https://doi.org/10.7554/elife.81983.

[53] I. L. Hanganu, Y. Ben-Ari, and R. Khazipov. Retinal waves trigger spindle bursts in the neonatal rat visual cortex. *Journal of Neuroscience*, 26(25):6728–6736, June 2006. doi: 10.1523/jneurosci.0752-06.2006. URL https://doi.org/10.1523/jneurosci.0752-06.2006.

[54] Yasunobu Murata and Matthew T Colonnese. An excitatory cortical feedback loop gates retinal wave transmission in rodent thalamus. *eLife*, 5, October 2016. doi: 10.7554/elife.18816. URL https://doi.org/10.7554/elife.18816.

[55] Bernd Illing, Jean Robin Ventura, Guillaume Bellec, and Wulfram Gerstner. Local plasticity rules can learn deep representations using self-supervised contrastive predictions. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=Yu8Q6341U7W.

[56] Artur Luczak, Bruce L. McNaughton, and Yoshimasa Kubo. Neurons learn by predicting future activity. *Nature Machine Intelligence*, 4(1):62–72, January 2022. doi: 10.1038/s42256-021-00430-y. URL https://doi.org/10.1038/s42256-021-00430-y.

[57] Davide Zoccolan, Nadja Oertelt, James J. DiCarlo, and David D. Cox. A rodent model for the study of invariant visual object recognition. *Proceedings of the National Academy of Sciences*, 106(21):8748–8753, May 2009. doi: 10.1073/pnas.0811583106. URL https://doi.org/10.1073/pnas.0811583106.