Computational constraints underlying the emergence of functional domains in the topological map of Macaque V4

Dunhan Jiang[1], Tianye Wang[2,3,4,5], Shiming Tang[2,3,4,5], and Tai-Sing Lee[6]

[1]School of Computer Science and Mellon College of Science, Carnegie Mellon University, Pittsburgh, PA 15213
[2]School of Life Sciences, Peking University, Beijing, China, 100871
[3]Peking-Tsinghua Center for Life Sciences, Beijing, China, 100871
[4]IDG/McGovern Institute for Brain Research, Peking University, Beijing, China, 100871
[5]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing, China, 100871
[6]Computer Science Department and Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213

V4, an intermediate visual area in the ventral visual stream of primates, is known to contain neurons tuned to color, complex local patterns, shape, and texture. Neurons with similar visual attribute preferences are closely positioned on the cortical surface, forming a topological map. Recent studies based on multielectrode arrays and calcium imaging revealed the macaque V4 has neuronal columns tuned to specific natural image features, and these columns are clustered into various functional domains. There are domains tuned to attributes generally associated with object surface properties such as texture or color, as well as domains associated with the shape and form of object boundaries reminiscent of the blobs and inter-blobs in the primary visual cortex. Here, we explored the computational constraints underlying the development of the V4 topological map. We found that the map learned based on self-organizing principles constrained by neuronal column's tuning and retinotopy position can account for many characteristics of the observed V4 map, including the interwoven organization of texture and shape processing clusters. These anatomical clustering, with the implied local recurrent connectivity, might facilitate a modular parallel processing of surfaces and boundaries of objects along the ventral visual system.

macaque V4 | organizational principle | texture and shape | retinotopy | self-organizing map

Correspondence: *taislee@andrew.cmu.edu*

Introduction

The ventral visual stream extracts a variety of visual attributes to support object recognition (1–3). V4, an intermediate visual area along the ventral stream, is known to contain neurons with selectivity to color (4), orientation (5), spatial frequency (6, 7), curvature (8, 9), figure-ground segregation (10), texture (11–14), and shape information (15–17). It is found that there are distinct groups of neurons processing texture and shape information rather independently (18, 19). Recent studies further revealed that V4 neurons encode a wide variety of natural image features, and they are organized in columns of distinct functional groups with their tunings ranging from complex patterns such as curvatures to texture and boundary, color, and even facial parts such as eye features (20). Wide-field imaging (21) revealed that the top layer "pixels" of V4 are organized into functional domains.

These pixels are about 90μm x 90μm in resolution, roughly in the scale of a column, and thus each reflecting the feature tuning of neurons in a column, analogous to a V1 orientation column. Some of these feature columns are characterized with distinct surface information, such as texture and color. Other feature columns are embedded with shape information, characterized by object boundaries. These neuronal columns cluster together to form functional domains. The functional domains with surface properties and those with shape properties are interleaved in the map, akin to the blob and inter-blob in the primary visual cortex (21), suggesting there might be distinct modules for parallel surface and shape processing.

What are the logic and constraints underlying the development of the topological map in V4, particularly in the organization of shape and surface processing modules? Along the hierarchical ventral visual pathway, the topological map organization in each area has to resolve the tension that arises from trying to fit multiple visual feature dimensions within a single retinotopic map (22–24). In V1, orientation, spatial frequency, and color feature dimensions are fitted in the spin-wheel hyper-columns of the global retinotopic map. As the neuron receptive field becomes larger going up the visual hierarchical system, the retinotopic constraint becomes looser, and the topological map becomes more dominated by visual features. In the object recognition areas such as the inferotemporal cortex (ITC), the cortical columns become organized in various object-centric functional domains (25) such as animacy (26), object sizes (27), scenes (28), and faces (29–31). While ITC neurons are believed to be organized on the cortical map with ordered features (32–34), retinotopy is still detectable (35–39). V4, an intermediate visual area, has to accommodate the encoding of a much larger set of feature patterns under the confinement of retinotopy (40). The observed topological organization of the functional domains in V4 likely results from the interplay of these constraints. Here, we used a self-organizing algorithm simulated cortical map to better understand these constraints.

The self-organizing map is a computational model based on Hebbian learning, local recurrent excitation, and global inhibition to explain the development of a topological map in the brain (41, 42). It is a process that brings neurons with simi-
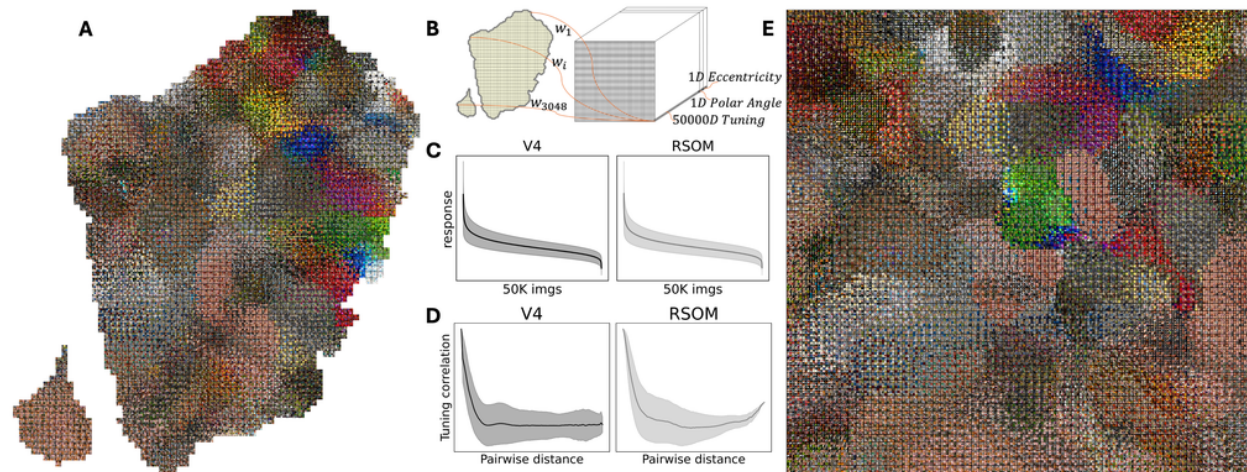
Fig. 1. A. Monkey V4 digital twin image preference map, each column shows its most preferred nine images out of 50000 color natural images. **B.** RSOM connects to neural column image space. Each column is described by a 50000D tuning curve and its weighted polar angle and eccentricity values. A 60 by 60 gridded simulated RSOM cortical map thus learns a weight vector of 60 by 60 by 50002 dimensions. **C.** Averaged sorted tuning curve shape of the V4 digital twin and our RSOM in darker solid lines, with one standard deviation above and below marked by shadowed areas. **D.** Pairwise V4 columns and RSOM grids tuning correlation as a function of map distance, with function average and standard deviation vectors shown. For every grid, we compute the correlation between its tuning curve and all other grid tuning curves, where correlations are ranked based on the order of pairwise grid Euclidean map distance. **E.** RSOM simulated map image preference. Each grid shows its most preferred nine images from its weight vector first 50000D, representing its tuning curve to the same set of images.

lar tunings together in a cortical spatial neighborhood. Self-organizing algorithms and models sharing similar assumptions have been used to explain the formation of V1 orientation columns and the ITC object-centric functional domains (43–45). One rationale for this design is to minimize connections (46) between neurons encoding semantically related concepts together for encoding the relationship between these concepts. Thus, topological maps provide a valuable window into neural codes and cortical computation. However, the self-organizing map approach has not been applied to V4 topological maps. A recent spatially regularized topological deep learning network (TDANN) purports to explain cortical map formation throughout the ventral stream from V1 to ITC (45). Yet, the V4 map predicted from this model fails to match some of the key characteristics of the topological map of functional domains we observed in V4, suggesting either some key computational constraints are missing or perhaps there are inadequacies in the algorithm.

Given the measured neuronal columns with different tunings to color natural images, what would be the constraints essential for the self-organizing algorithm to map these neuronal columns to reproduce the observed topology of V4 functional domains (21)? We found that balancing the retinotopic constraint and the feature tuning similarity constraint is key to reproducing a topological map of the functional domains similar to the observed map in many measures, particularly in terms of (1) the distributions of size and adjacency relationship of the functional domains in the map, (2) the simultaneous harboring of a continuous feature map and retinotopic maps, and (3) spatial arrangement of neural clusters with tunings associated with surface properties such as texture and color, and clusters with tunings associated with shape properties such as curvature, junctions and lines. In this paper, we will be referring to this retinotopically constrained self-organizing map as RSOM and a self-organizing map constrained only on image tuning as SOM. Our findings provide further evidence in support of the self-organizing principle as the governing rule in the development of the topological organization of neuronal columns and domains. However, what biological processes underlie the formation of topological maps and how such maps enhance neural computation are mysteries that remain to be unveiled.

Results

From the macaque V4 to our RSOM simulated cortical map. To organize all neuronal columns in the V4 digital twin (21) shown in Figure 1A, we leveraged the self-organizing algorithm to train our RSOM mutually constrained by their predicted tuning and estimated retinotopy. With balanced mutual constraints, we allow the Gaussian neighborhood to be big at first and gradually shrink, similar to limiting the wiring cost (46) (see Materials and Methods). This process is biologically plausible as it starts with a long-range horizontal connection that branches to facilitate short-range connections (47, 48). The resulting simulated cortical map in Figure 1E consists of 60 by 60 grids, each learning a tuning curve. We first ask whether the shape of these learned tuning curves is similar to that of the V4. We sort every tuning curve to compute an averaged curve along with a standard deviation from all RSOM grids. This averaged tuning curve responds to 1157 images before dropping to half of the maximum response, while the V4 does so for 842 images out of 50000 images (Figure 1C), corresponding to a sparse code (49). When comparing the averaged tuning of the V4 against our RSOM, it gives a 0.996 Pearson correlation. This drops to 0.992 when comparing the SOM to the V4 benchmark, indicating the retinotopy constraint facilitates the learning of more V4-like tuning curves.
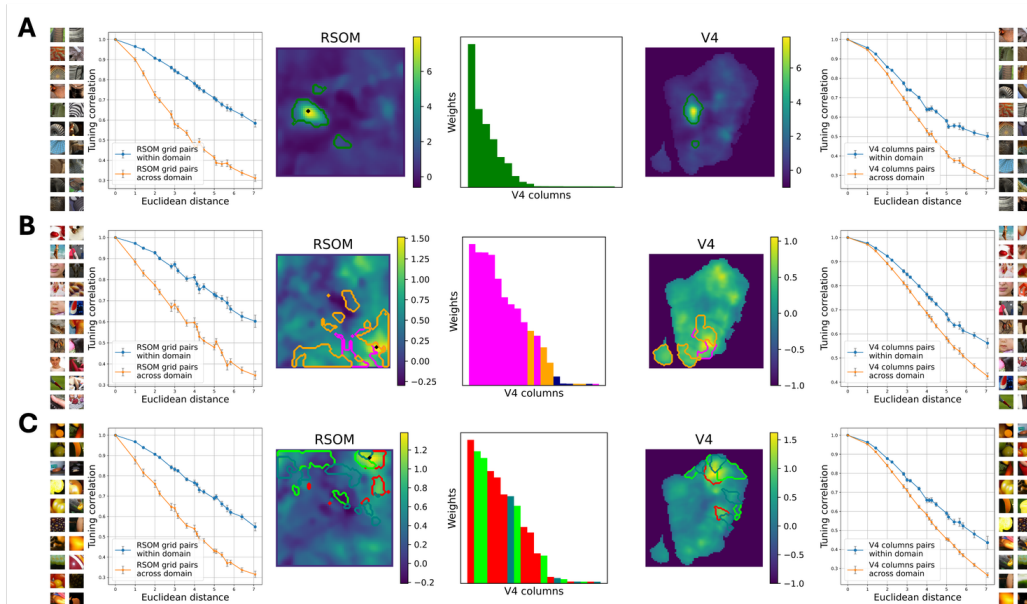
Fig. 2. A. Texture domain. The left-most images are preferred by RSOM grids as enclosed by green line. They elicit RSOM grid responses, as shown in the left RSOM heatmap. The right-most images are preferred by the same V4 domain with their responses shown in the right V4 heatmap. For all boundary grids, we compute each grid's tuning correlation to adjacent grids as a function of map distance, the mean and SEM vectors for grids within and across the home domain are shown for the RSOM (left) and the V4 (right). The RSOM example grid's (a black dot in the heatmap) top one thousand tuning is decomposed as a linear combination of V4 columns through a LASSO linear regression. The middle bar chart shows the top 20 V4 columns with the largest model weights, each as a single bar whose color represents the V4 column domain label. **B.** Face domain. **C.** Curvature domain.

Besides a similar tuning shape, how the tuning curve changes across the cortical surface in these maps is another question. We specifically ask how similar are pairwise grid tuning curves as a function of the physical distance between them on the map. For each RSOM grid, we compute the Pearson correlation between its tuning curve and all other grid tuning curves while recording a pairwise grid distance. This vector of tuning correlations is sorted according to the order of the distance vector. Figure 1D shows the averaged correlation values as a function of distance along with a standard deviation, where spatially adjacent grids have much more similar tuning curves as compared with spatially distant grids. This means spatially closed grids respond particularly to their shared most preferred images, whereas distant grids respond selectively to distinct sets of images. Computing the averaged function vector gives a Pearson correlation of 0.95 between our RSOM and the V4, which drops to 0.89 when comparing the SOM to the V4. This indicates the V4 cortical surface may not be a complete tuning continuum as retinotopy constraint plays a noticeable role in driving a simulated map with a more similar tuning correlation pattern against the benchmark.

The emergence of V4 domains in the RSOM. Since the V4 map is assembled from sixteen different domains (21), we hypothesize the domain boundary area may experience a sharper tuning curve transition as compared to the tuning transition happening within the domain, thus explaining the above benefit of having an extra constraint other than image tuning. We then look for the emergence of V4 domains by assigning the most similar V4 neuronal column onto every RSOM grid such that the grid can borrow the domain label from the V4 column. This is done by computing correlations between an RSOM grid learned tuning curve and all V4 neuronal columns' tuning curves such that the column with the highest similarity gets assigned to that grid. We observe the

emergence of all V4 domains in our RSOM, as Figure 2A heatmaps demonstrates an example curvature preferring domain enclosed by green line. Two columns of images on the left- and right-most show unique images in the set of top nine most preferred images by all RSOM grids or V4 columns within this domain. Both heatmap colorings indicate averaged response magnitude to these sets of unique images. For the RSOM, we access every boundary grid tuning curve correlation against another nearby grid in a square neighborhood, either within or across the home domain, while recording a pairwise distance. The averaged correlation value as a function of distance is lower for grid pairs across the domain than within the same domain, which is the same for the V4 as well as other domains in both maps. This means tuning transitions more smoothly as a continuum within a domain rather than a more sudden change happening at the domain boundary.

To better understand this perspective, we look at the tuning composition of example RSOM grids at both the center and peripheral of some domains. The black dot in the RSOM heatmap of Figure 2A represents an example grid that has the strongest response inside the curvature-preferring domain, which happens to sit at the domain center. We look at this example grid top one thousand tuning curve and ignore the remaining baseline tuning curve, which is not meaningful. We train an L1-regularized LASSO linear regression model to the fullest to represent its tuning as a linear combination of all V4 column tuning curves. We display the top 20 V4 columns with the biggest linear regression model weights as 20 bars in the center bar chart where its color corresponds to the V4 columns domain label. This example grid is thus learning a tuning curve as a linear combination of only a few V4 columns entirely coming from the same domain in V4.

This story becomes more complicated for an example grid on the more peripheral area, instead of the center, of a shape-preferring domain enclosed by the pink line in Figure 2B
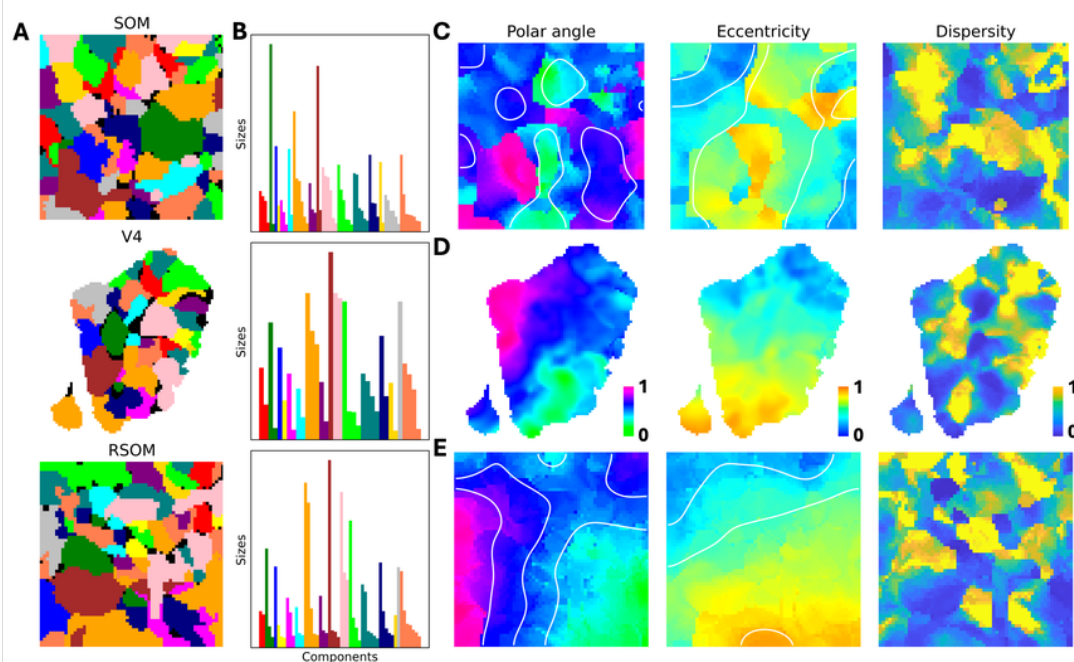
Fig. 3. **A.** Domain label map by assigning V4 column onto SOM/RSOM grids. Sixteen distinct colors each represent one domain. Black region means erased grids/columns, whose original connected-component with the same domain label is not large enough. **B.** Domain connected-component sizes. Each bar represents a single component where the same set of colors are used across subfigures to indicate domain label. **C.** SOM polar angle, eccentricity, and feature dispersity maps from V4 column assignment, with contour lines added for the Gaussian smoothed retinotopy maps, indicating their global pattern. **D.** V4 retinotopy and dispersity maps. **E.** RSOM retinotopy and dispersity maps, with contour lines similarly added.

heatmaps. While this domain is adjacent to another face-preferring domain enclosed by an orange line, the example black grid in the RSOM heatmap is learning a tuning curve from a slightly more sophisticated subset of V4 columns, including both some home domain columns and a few other columns coming from its adjacent domain. In Figure 2C, we show an extreme case where the example black grid in the red line enclosed domain sits directly adjacent to another two domains that prefer red images (enclosed by the dark green line) and curvature images (enclosed by the light green line). It learns a tuning curve from a very diverse set of V4 columns coming from these three domains. We thereby argue the V4 map is not a perfect tuning continuum, with tuning transitions being smoother within a domain and undergoing more sudden changes at domain boundary lines, as captured by our RSOM. This hints that the V4 map domain assembly and adjacency pattern should not be completely driven by the image tuning constraint.

Simulated cortical map domain adjacency, retinotopy, and feature dispersity patterns. We visualize domain label maps given by the V4 column assignment in Figure 3A, where each color indicates a specific domain that can split into a few components across the map. All domain component sizes are visualized as a bar chart in Figure 3B with the same coloring. The average domain component sizes of the RSOM and V4 achieve a Pearson correlation of 0.791, whereas this correlation drops to 0.427 for the SOM. We use a square domain adjacency matrix to characterize cortical map domain assembly pattern, where the ith row and jth column value indicates the number of grids in the ith domain being adjacent to grids in the jth domain, where each row is normalized. The Pearson correlation between the flattened adjacency matrix of the RSOM and the V4 reaches 0.755, whereas this value drops to 0.484 for the SOM. These find-

ings indicate the crucial role of retinotopy in driving the simulated cortical map to be more V4-like, where it guides domains into their relative positioning onto the cortical map. This perspective matches evidence in the ITC pointing to an experience-insensitive and predetermined innate network structure (50, 51), although similar domain adjacency patterns were only studied in the ventral visual pathway at a macro-scale (52, 53).

Cortical Sheets	num	sum	mean	std	min	max
1. V4	7	0.31	0.04	0.04	0.01	0.09
2. SOM	6	0.32	0.05	0.06	0.01	0.17
3. RSOM	9	0.33	0.04	0.03	0.01	0.09

Table 1. High dispersity component size statistics, including the total number of components, total map area ratio, averaged component area ratio, standard deviation of all component area ratios, minimum component area ratio, and maximum component area ratio.

V4 column assignment also brings retinotopy and feature dispersity measurements for every simulated map grid. We visualize these map patterns for the SOM (Figure 3C), V4 benchmark (Figure 3D), and our RSOM (Figure 3E). Due to the retinotopy constraint, contour lines in the polar angle map intersect with those lines in the eccentricity map roughly perpendicularly in the RSOM simulated cortical map, similar to the V4, instead of the chaotic pattern found in the regular SOM. For the dispersity pattern, we did a two-way hierarchical clustering on all maps to cluster grids into either low or high dispersity types. This finds the connected high dispersity components in all maps, and their sizing statistics are shown in Table 1. Although there is no explicit dispersity driving force in our RSOM, the feature dispersity map splits into many high dispersity components in yellow that prefer dispersed texture features and low dispersity areas in blue that tune to more spatially localized shape and object boundaries, resembling the high- and low-dispersity segregation in
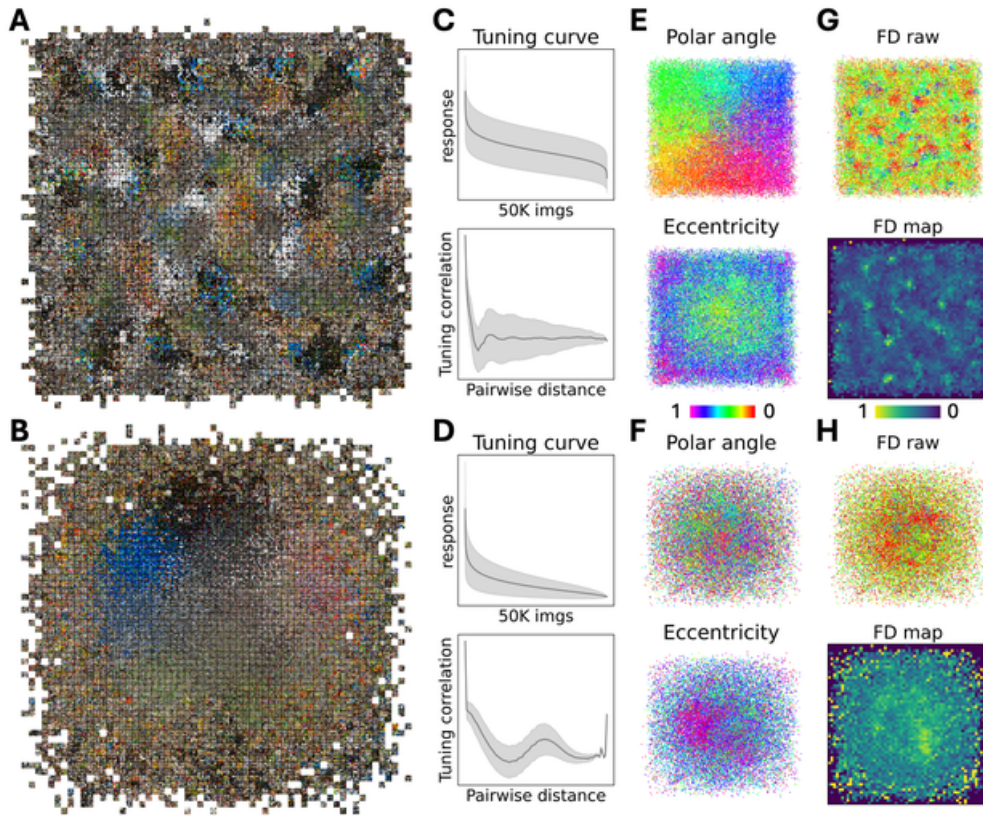
Fig. 4. A. TDANN purported V4 image preference map, with 60 by 60 pixels each showing nine most preferred images. This cortical map was equally segmented into pixels, where all units within a pixel yield a mean tuning. **B.** TDANN purported ITC 60 by 60 image preference map. **C.** Top: averaged sorted tuning curve for V4 layer pixels with a much wider shape than that of the macaque V4 (corr=0.708). Bottom: V4 layer pixels pairwise tuning curve correlation as a function of map Euclidean distance (corr=0.838 against the macaque V4). **D.** Top: ITC layer pixels mean tuning has a similar shape against the macaque V4 (corr=0.947). Bottom: ITC layer pixels pairwise tuning correlation as a function of distance (corr=0.778). **E.** Top: V4 layer polar angle map. Bottom: V4 layer eccentricity map. **F.** Top: ITC layer polar angle map. Bottom: ITC layer eccentricity map. **G.** Top: V4 layer feature dispersity map. Bottom: aggregated V4 layer feature dispersity map with 60 by 60 pixels. **H.** Top: ITC layer feature dispersity map. Bottom: aggregated ITC layer feature dispersity map with 60 by 60 pixels.

V4. Without the retinotopy constraint, the regular SOM features less segregated high dispersity components, where one large high dispersity cluster emerges across the map instead. This exciting finding points to a retinotopy-dependent organizational principle being the driving force behind the observed modular parallel processing of segregated texture and shape clusters in the V4 and along the ventral visual processing stream.

TDANN purported V4 and ITC layer map topologies. As supervised models tend to encode images differently from the brain (54, 55), innovation of spatially regularized goal-driven network (45) was trained from contrastive learning to model the ventral visual stream cortical map. Here, we examine TDANN's simulated V4 and ITC cortical maps. For each map, we extract all artificial unit responses to the same set of 50000 natural images. The simulated cortical map area was segmented into 60 by 60 pixels such that each pixel have an aggregated tuning curve from all artificial units within this pixel area. We show the top nine most preferred images of these pixels for both the V4 (Figure 4A) and the ITC (Figure 4B) layer. Upon an initial examination, the V4 map forms small domains tuned to color and object but no faces, while the ITC map forms larger domains only at the map center tuned to color, face, and object. While this TDANN model was not trained to match the macaque V4 neural code, we examine its tuning curve shapes. We found its ITC average tuning curve shape achieves a 0.947 correlation against our benchmark macaque V4, which drops to only 0.708 for its purported V4 layer, whose tuning curve also becomes much

less sparse. We repeat the pairwise pixel tuning curve correlation as a function of simulated cortical map distance analysis and find 0.838 and 0.787 correlations for the purported V4 and ITC layers against our benchmark V4 (Figure 4C and 4D). Due to distinct image preferences, we select all unique top nine most preferred images from macaque V4 neuronal columns, to build a representation similarity matrix for every cortical map. This gives a correlation of 0.303 for the TDANN V4 map and a 0.242 correlation for the TDANN ITC map against the benchmark V4, indicating a divergence between goal-driven model image representation and the biological visual cortex representation on a population code level (56).

We further tested the retinotopy organization in these simulated cortical maps, as the position of artificial unit was retinotopically initialized based on their feature map location in the convolution structure. Here, we use this convolution layer feature map structure to assign retinotopy positions for all units based on their position in the feature map, assuming the center of this feature map is the fovea of the visual field. We show both polar angle and eccentricity values for all artificial units in the purported V4 (Figure 4E) and ITC (Figure 4F) layers, indicating some retinotopy preserved at the V4 simulated map but not at the ITC map where long-range horizontal connection and neighborhood position optimization likely disrupted the retinotopic structure. But whether the preserved retinotopic organization at the simulated V4 cortical map gives rise to segregated shape and texture parallel and modular processing remains unknown. We applied the

same feature attribution analysis (21) to investigate the feature dispersity properties of artificial units in both the V4 and the ITC layers. As a result of our test images, where the peripheral portion was occluded to mimic what the monkey saw during the experiment, we chose the center feature map artificial unit as a representative of all other units in the same feature channel, disregarding their feature map locations. However, starting from the purported V4 layer to the ITC layer, we found TDANN units' feature dispersity values follow a different distribution from the monkey V4 neuronal columns in the digital twin. As a result, we did not observe the high and low feature dispersity segregation across both simulated cortical maps here in Figure 4G and 4H, indicating different feature processing and learning between the TDANN and the macaque V4.

Conclusions

In this paper, we showed that the recently observed V4 topological map of natural image feature selectivity is compatible with the self-organizing algorithm advanced by Kohnon (41). Self-organizing and similar algorithms have been used to explain V1 and ITC maps (44, 57, 58), which have not been used to explain V4 maps, particularly the topological natural image preference map (21). Using a digital twin built by fitting wide-field calcium imaging neural responses over a large stretch of V4 at 90 µm scale to a large set of natural images, we have come to a better understanding of how natural scenes are embedded in V4 in the form of a topological map. Here, we show that the self-organizing algorithm with retinotopic constraint, as well as feature similarity constraint, is sufficient to reproduce cortical maps that match the observed V4 map of functional domains of natural image feature preferences in a number of metrics, particularly in terms of (1) the distributions of size and adjacency relationship of the functional domains in the map, (2) the simultaneous harboring of a continuous feature map and a continuous retinotopic map, and (3) spatial arrangement of neural clusters with tunings associated with surface properties such as texture and color, and clusters with tunings associated with boundary shape properties such as curvature, junctions, and lines.

While the functional roles of the observed cortical topology and the clustering of domains remain uncertain, minimizing horizontal connection (46) and top-down feedback connections to encode related concepts to facilitate neural computation is likely a key factor. Visual independent factors such as other sensory modalities and memories that affect the functions of the visual cortex might also contribute to the formation of topological maps, as there are long-range connectivity extending outside of the visual cortex (59–61) and functional domains in extrastriate visual cortex can be activated by concepts communicated with sound in the congenitally blind group (62). Retinotopy, as a visual dependent factor, is known to be an important constraint guiding the formation of the visual cortical map. Current evidence suggests the existence of retinotopy organization of the visual space in primate visual systems at birth, even prior to the formation of neuronal selectivity and functional domains (22–24).

Traces of such inborn retinotopy organization are present at matured primate V4 (40). Eccentricity remains an important factor (35–39) in cortical maps of the ITC, where even scene domain (28) and face domains (30, 31) within the ITC are found to exhibit retinotopic organization at the mesoscale.

We argue that the organization of multiple visual feature dimensions in the macaque V4 area should also simultaneously satisfy the retinotopic organization projected from the retina visual input. Without this retinotopic constraint, as shown in our SOM, functional domains tend to appear in larger chunks. The pressure due to the retinotopic constraint, as in our RSOM, could break a functional domain, such as the facial feature domain, into multiple pieces, conceptually similar to orientation columns in the different hypercolumns in V1. The retinotopy constraint also breaks the high-dispersity domain clusters into pieces, surrounded by low-dispersity domains. It is possible that the high dispersity domains might derive their major input from the V1 blobs / V2 thin stripes and the low dispersity domains from the inter-blobs / V2 thick stripes, constituting parallel streams of specialized surface (color, texture) and shape (orientation, curvature, and form) processing in the ventral visual system. However, further anatomical and physiological studies are needed to evaluate this conjecture.

On the other hand, neural networks have been shown to match the ventral visual systems reasonably well by measures of neural response prediction and representational similarity analysis (54, 55). Neural response prediction suggests different network layers have a rough correspondence with the different visual areas of the ventral stream in terms of the embedding features, while representational similarities suggest a matching in the geometry of the neural manifold of these embedded features. Deep convolutional neural networks, however, do not exhibit topological maps, which is a key characteristic of the organization of the visual cortex. A recently proposed spatially regularized topological deep artificial neural network (45) is an attempt to address this gap by imposing a loss that encourages neurons with similar tunings to come together spatially but subject to the normal convolutional neural network model architecture. Unfortunately, this model fails to predict the V4 map of natural image features that we observed, though it generates a V1-like map in the early layer and an ITC-like map in the higher layer. The features in the purported V4 layer, specifically block 3.1 in the ResNet18 model, appear to be missing some of the V4 feature tunings, most notably facial feature tunings, as well as a greater variety of color tunings. It does exhibit functional domains but with a coarser retinotopic map and completely scattered and chaotic high and low dispersity clusters, similar to our SOM simulated map with only the tuning feature constraint. This suggests the segregation of high and low dispersity clusters, presumably corresponding to surface and shape processing, requires interaction with a relatively high retinotopic constraint during the learning of the map.

Both self-organizing maps and topological CNN primarily address the issues of understanding the topological map at an algorithmic and representational level (63). The theoret-

ical level account and the implementation level account remain to be explored and developed. At the theoretical level, the topological map might be an efficient embedding of the efficient codes of natural scenes related to the sparse manifold transform (SMT) (64), where neurons encoding spatially and/or temporally co-occurring visual events become more connected due to Hebbian learning and slow-feature analysis (65). The excitatory mutual connection might bring the neurons with close semantic relationships closer together in cortical space to facilitate building connections. At the implementation level, further empirical experiments are required to understand the exact biological processes and mechanisms that realize map formation under the retinotopic and feature similarity constraints.

Materials and methods

RSOM parameters and data. We applied a PyTorch-based Kohonen self-organizing algorithm to the image space of V4 digital twin columns, each coupled with its retinotopy, including one polar angle and one eccentricity value. RSOM learns a simulated cortical map of 60 x 60 (H x W) grids where each grid learns a tuning curve. The weight of this RSOM is of shape H x W x F (60 x 60 x 50002), of which the third axis first 50000D learns the tuning and the last 2D corresponds to polar angle and eccentricity. This design allows a direct constraint in finding every best matching unit from both tuning and retinotopy. The training data is 3048 monkey V4 digital twin column vectors, each consisting of 50000D tuning plus its polar angle and eccentricity values.

RSOM training. We do a complete random weight initialization while normalizing the weight vector for each grid. This ensures no V4 topology information was pre-planted into the simulated cortical map. The RSOM only sees one V4 column vector at a time when training. The minimum Euclidean distance between the input vector and all RSOM grid weight vectors finds the best matching unit (BMU, see Equation 1), where $w = 1$ $if \in [1:50000]$ and $w = 125$ $if \in [50001:50002]$.

$$\text{BMU} = \arg\min_{w,h} \sqrt{\sum_{f=0}^{f=50002} w * \left(\text{input}_f - \text{weight}_{(w,h),f}\right)^2}$$

(1)

This winning grid serves as the center of a Gaussian neighborhood function in Equation 2 to update all grid weights, with grids closer to the BMU getting more weight changes toward the input V4 column vector while more distant grids get fewer updates. This filter function η_t starts big and decays to an intermediate value. The weight update (Equation 3) uses a constant learning rate α on top of the neighborhood function. This RSOM is trained for 120 epochs, where a single epoch iterates through all 3048 V4 neuronal columns for one time. The SOM is similarly trained as well.

$$\eta_t = e^{\frac{-\text{Dist}\left(\text{BMU}_{(w,h)}, \text{Grid}_{(i,j)}\right)^2}{2\sigma_t^2}}$$

(2)

$$\text{RSOM}_w = \text{RSOM}_w + (\text{input} - \text{RSOM}_w) * (\alpha * \eta_t)$$

(3)

RSOM example grid tuning composition. For an example grid, we select its top one thousand tuning curve portion. We train a LASSO linear regression model to fit this partial tuning curve as a linear combination of 3048 V4 columns' partial tuning to exactly these one thousand images (Equation 4). The loss is a mean squared error plus another L1 punishment term applied to all 3048 weights. This regularization is multiplied by 0.01 to control the penalty. Each model is trained with a constant learning rate α=0.05 for 7500 epochs to the fullest, giving a final loss around 0.01.

$$\text{example-grid}_{\text{top1k}} = \sum_{i=1}^{3048} \text{w}_i \cdot \text{pixel-i}_{\text{top1k}}$$

(4)

Map connected components and adjacency matrix. In both simulated cortical maps and the V4 map, we find domains as connected components. This means a cluster of grids that are connected also share the same domain label. We define an eight-connectivity where both direct adjacent and diagonal adjacent grids of the same label will be considered as connected. To eliminate small and noisy components in the map, we filter out those connected components whose size is smaller than ten by erasing the label of these grids. We compute an adjacency matrix for each map to quantify relative domain positioning patterns. This is a 16 by 16 matrix where each row and column correspond to a specific domain. For the ith domain represented by the ith row, we look at all of its connected components. We survey each component's adjacent grids where whenever it is adjacent to one grid labeled as the jth domain, we add a value of one to the ith row jth column and the jth row ith column. Each row vector is normalized after surveying the entire map.

Code. Scripts for model training and simulated map analysis are available at: https://github.com/777dunhan/Topological-map-of-Macaque-V4

Bibliography

1. Dwight J. Kravitz, Kadharbatcha S. Saleem, Chris I. Baker, Leslie G. Ungerleider, and Mortimer Mishkin. The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1):26–49, 2013.
2. Maximilian Riesenhuber and Tomaso Poggio. Neural mechanisms of object recognition. *Current Opinion in Neurobiology*, 12(2):162–168, 2002.
3. James J. DiCarlo, Davide Zoccolan, and Nicole C. Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.
4. Ye Liu, Ming Li, Xian Zhang, Yiliang Lu, Hongliang Gong, Jiapeng Yin, Zheyuan Chen, Liling Qian, Yupeng Yang, Ian Max Andolina, Stewart Shipp, Niall Mcloughlin, Shiming Tang, and Wei Wang. Hierarchical representation for chromatic processing across macaque v1, v2, and v4. *Neuron*, 108(3):538–550, 2020.
5. Hisashi Tanigawa, Haidong D Lu, and Anna W Roe. Functional organization for color and orientation in macaque v4. *Nature Neuroscience*, 13:1542–1548, 2010.
6. Yiliang Lu, Jiapeng Yin, Zheyuan Chen, Hongliang Gong, Ye Liu, Liling Qian, Xiaohong Li, Rui Liu, Ian Max Andolina, and Wei Wang. Revealing detail along the visual hierarchy: Neural clustering preserves acuity from v1 to v4. *Neuron*, 98(2):417–428, 2018.
7. Ying Zhang, Kenneth E Schriver, Jia Ming Hu, and Anna Wang Roe. Spatial frequency representation in v2 and v4 of macaque monkey. *eLife*, 12(2050-084X):e81794, 2023.
8. Jia Ming Hu, Xue Mei Song, Qiannan Wang, and Anna Wang Roe. Curvature domains in v4 of macaque monkey. *eLife*, 9(2050-084X):e57261, 2020.
9. Rendong Tang, Qianling Song, Ying Li, Rui Zhang, Xingya Cai, and Haidong D Lu. Curvature-processing domains in primate v4. *eLife*, 9(2050-084X):e57502, 2020.

10. Anna W. Roe, Leonardo Chelazzi, Charles E. Connor, Bevil R. Conway, Ichiro Fujita, Jack L. Gallant, Haidong Lu, and Wim Vanduffel. Toward a unified theory of visual area v4. *Neuron*, 74(1):12–29, 2012.

11. Gouki Okazawa, Satohiro Tajima, and Hidehiko Komatsu. Image statistics underlying natural texture selectivity of neurons in macaque v4. *Proceedings of the National Academy of Sciences*, 112(4):E351–E360, 2014.

12. Taekjun Kim, Wyeth Bair, and Anitha Pasupathy. Perceptual texture dimensions modulate neuronal response dynamics in visual cortical area v4. *The Journal of neuroscience*, 42(4):631–642, 2022.

13. Michele A. Cox, Michael C. Schmid, Andrew J. Peters, Richard C. Saunders, David A. Leopold, and Alexander Maier. Receptive field focus of visual area v4 neurons determines responses to illusory surfaces. *Proceedings of the National Academy of Sciences*, 110(42):17095–17100, 2013.

14. Rundong Jiang, Ian Max Andolina, Ming Li, and Shiming Tang. Clustered functional domains for curves and corners in cortical area v4. *eLife*, 10(2050-084X):e63798, 2021.

15. A Pasupathy and C E Connor. Shape representation in area v4: position-specific tuning for boundary conformation. *Journal of Neurophysiology*, 86(5):2505–2519, 2001.

16. Anitha Pasupathy and Charles E. Connor. Population coding of shape in area v4. *Nature Neuroscience*, 5(12):1332–1338, 2002.

17. Anirvan S. Nandy, Tatyana O. Sharpee, John H. Reynolds, and Jude F. Mitchell. The fine structure of shape tuning in area v4. *Neuron*, 78(6):1102–1115, 2013.

18. Taekjun Kim, Wyeth Bair, and Anitha Pasupathy. Neural coding for shape and texture in macaque area v4. *The Journal of Neuroscience*, 39(24):4760–4774, 2019.

19. Anitha Pasupathy, Taekjun Kim, and Dina V Popovkina. Object shape and surface properties are jointly encoded in mid-level ventral visual cortex. *Current Opinion in Neurobiology*, 58:199–208, 2019.

20. Konstantin F. Willeke, Kelli Restivo, Katrin Franke, Arne F. Nix, Santiago A. Cadena, Tori Shinn, Cate Nealley, Gabrielle Rodriguez, Saumil Patel, Alexander S. Ecker, Fabian H. Sinz, and Andreas S. Tolias. Deep learning-driven characterization of single cell tuning in primate visual area v4 unveils topological organization. *bioRxiv*, 2023.

21. Tianye Wang, Tai Sing Lee, Haoxuan Yao, Jiayi Hong, Yang Li, Hongfei Jiang, Ian Max Andolina, and Shiming Tang. Large-scale calcium imaging reveals a systematic v4 map for encoding natural scenes. *Nature Communications*, 15(1):1–15, 2024.

22. Michael J Arcaro and Margaret S Livingstone. A hierarchical, retinotopic proto-organization of the primate visual system at birth. *eLife*, 6(2050-084X):e26196, 2017.

23. Andrew D Huberman, Marla B Feller, and Barbara Chapman. Mechanisms underlying development of visual maps and receptive fields. *Annual Review of Neuroscience*, 31:479–509, 2008.

24. Sohrab Najafian, Erin Koch, Kai Lun Teh, Jianzhong Jin, Hamed Rahimi-Nasrabadi, Qasim Zaidi, Jens Kremkow, and Jose-Manuel Alonso. A theory of cortical map formation in the visual brain. *Nature Communications*, 13(1):2303, 2022.

25. P. E. Downing, A. W.-Y. Chan, M. V. Peelen, C. M. Dodds, and N. Kanwisher. Domain specificity in visual cortex. *Cerebral Cortex*, 16(10):1453–1461, 2006.

26. Talia Konkle and Alfonso Caramazza. Tripartite organization of the ventral stream by animacy and object size. *The Journal of neuroscience*, 33(25):10235–10242, 2013.

27. Talia Konkle and Aude Oliva. A real-world size organization of object responses in occipitotemporal cortex. *Neuron*, 74(6):1114–1124, 2012.

28. Michael J Arcaro and Margaret S Livingstone. Retinotopic organization of scene areas in macaque inferior temporal cortex. *The Journal of neuroscience*, 37(31):7373–7389, 2017.

29. Rankin Williams McGugin, J. Christopher Gatenby, John C. Gore, and Isabel Gauthier. High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences*, 109 (42):17063–17068, 2012.

30. Thomas Janssens, Qi Zhu, Ivo D. Popivanov, and Wim Vanduffel. Probabilistic and single-subject retinotopic maps reveal the topographic organization of face patches in the macaque cortex. *The Journal of neuroscience*, 34(31):10156–10167, 2014.

31. Reza Rajimehr, Natalia Y Bilenko, Wim Vanduffel, and Roger B H Tootell. Retinotopy versus face selectivity in macaque visual cortex. *Journal of Cognitive Neuroscience*, 26(12):2691–2700, 2014.

32. Richard Durbin and Graeme Mitchison. A dimension reduction framework for understanding cortical maps. *Nature*, 343(6259):644–647, 1990.

33. Kalanit Grill-Spector and Kevin S Weiner. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8):536–548, 2014.

34. Pinglei Bao, Liang She, Mason McGill, and Doris Y. Tsao. A map of object space in primate inferotemporal cortex. *Nature*, 583(7814):103–108, 2020.

35. Bevil R. Conway. The organization and operation of inferior temporal cortex. *Annual Review of Vision Science*, 4:381–402, 2018.

36. Uri Hasson, Ifat Levy, Marlene Behrmann, Talma Hendler, and Rafael Malach. Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, 34(3):479–490, 2002.

37. Rafael Malach, Ifat Levy, and Uri Hasson. The topography of high-order human object areas. *Trends in Cognitive Sciences*, 6(4):176–184, 2002.

38. Michael J. Arcaro, Stephanie A. McMains, Benjamin D. Singer, and Sabine Kastner. Retinotopic organization of human ventral visual cortex. *The Journal of Neuroscience*, 29(34):10638–10652, 2009.

39. Ifat Levy, Uri Hasson, Galia Avidan, Talma Hendler, and Rafael Malach. Center–periphery organization of human object areas. *nature neuroscience*, 4(5):533–539, 2001.

40. Denis Fize, Wim Vanduffel, Koen Nelissen, Katrien Denys, Christophe Chef d'Hotel, Olivier Faugeras, and Guy A. Orban1. The retinotopic organization of primate dorsal v4 and surrounding areas: A functional magnetic resonance imaging study in awake monkeys. *The Journal of Neuroscience*, 23(19):7395–7406, 2003.

41. Teuvo Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

42. Benoît Siri, Mathias Quoy, Bruno Delord, Bruno Cessac, and Hugues Berry. Effects of hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. *Journal of Physiology-Paris*, 101(1-3):136–148, 2007.

43. James A. Bednar, Amol Kelkar, and Risto Miikkulainen. Scaling self-organizing maps to model large cortical networks. *Neuroinformatics*, 2(3):275–301, 2004.

44. Fenil R. Doshi and Talia Konkle. Cortical topographic motifs emerge in a self-organized map of object space. *Science Advances*, 9(25):eade8187, 2023.

45. Eshed Margalit, Hyodong Lee, Dawn Finzi, James J. DiCarlo, Kalanit Grill-Spector, and Daniel L. K. Yamins. A unifying framework for functional organization in early and higher ventral visual cortex. *Neuron*, 112(14):2435–2451, 2024.

46. Nicholas M. Blauch, Marlene Behrmann, and David C. Plaut. A connectivity-constrained computational account of topographic organization in primate high-level visual cortex. *Proceedings of the National Academy of Sciences*, 119(3):e2112566119, 2022.

47. E M Callaway and L C Katz. Emergence and refinement of clustered horizontal connections in cat striate cortex. *The Journal of neuroscience*, 10(4):1134–1153, 1990.

48. S Löwel and W Singer. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science*, 255(5041):209–212, 1992.

49. Eric T. Carlson, Russell J. Rasquinha, Kechen Zhang, and Charles E. Connor. A sparse object coding scheme in area v4. *Current Biology*, 21(4):288–293, 2011.

50. Krishna Srihasam, Justin L Vincent, and Margaret S Livingstone. Novel domain formation reveals proto-architecture in inferotemporal cortex. *nature neuroscience*, 17(12):1776–1783, 2014.

51. Bradford Z Mahon and Alfonso Caramazza. What drives the organization of object knowledge in the brain? *Trends in Cognitive Sciences*, 15(3):97–103, 2011.

52. Rosa Lafer-Sousa, Bevil R. Conway, and Nancy G. Kanwisher. Color-biased regions of the ventral visual pathway lie between face- and place-selective regions in humans, as in macaques. *The Journal of neuroscience*, 36(5):1682–1697, 2016.

53. Yanchao Bi, Xiaoying Wang, and Alfonso Caramazza. Object domain and modality in the ventral visual pathway. *Trends in Cognitive Sciences*, 20(4):282–290, 2016.

54. Chengxu Zhuang, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel L. K. Yamins-3. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3):e2014196118, 2021.

55. Talia Konkle and George A. Alvarez. A self-supervised domain-general learning framework for human ventral stream representation. *Nature Communications*, 13(491):1–12, 2022.

56. Luciano Dyballa, Andra M. Rudzite, Mahmood S. Hoseini, Mishek Thapa, Michael P. Stryker, Greg D. Field, and Steven W. Zucker. Population encoding of stimulus features along the visual hierarchy. *Proceedings of the National Academy of Sciences*, 121(4):e2317773121, 2024.

57. Yiyuan Zhang, Ke Zhou, Pinglei Bao, and Jia Liu. A biologically inspired computational model of human ventral temporal cortex. *Neural Networks*, 178(0893-6080):106437, 2024.

58. T. Anderson Keller, Qinghe Gao, and Max Welling. Modeling category-selective cortical regions with topographic variational autoencoders. *arXiv:2110.13911*, 2021.

59. Zeynep M Saygin, David E Osher, Elizabeth S Norton, Deanna A Youssoufian, Sara D Beach, Jenelle Feather, Nadine Gaab, John D E Gabrieli, and Nancy Kanwisher. Connectivity precedes function in the development of the visual word form area. *Nature Neuroscience*, 19(9):1250–1255, 2016.

60. David E. Osher, Rebecca R. Saxe, Kami Koldewyn, John D.E. Gabrieli, Nancy Kanwisher, and Zeynep M. Saygin. Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cerebral Cortex*, 26(4):1668–1683, 2015.

61. Talia Konkle and Alfonso Caramazza. The large-scale organization of object-responsive cortex is reflected in resting-state network architecture. *Cerebral Cortex*, 27(10):4933–4945, 2017.

62. Ella Striem-Amit and Amir Amedi. Visual cortex extrastriate body-selective area activation in congenitally blind people "seeing" by using sounds. *Current Biiology*, 24(6):687–692, 2014.

63. David Marr. *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman and Company, 1982.

64. Yubei Chen, Dylan M Paiton, and Bruno A Olshausen. The sparse manifold transform. *arXiv:1806.08887*, 2021.

65. Laurenz Wiskott and Terrence J Sejnowski. Slow feature analysis: unsupervised learning of invariances. *Neural Comput*, 14(4):715–70, 2002.