Local lateral connectivity is sufficient for replicating cortex-like topographical organization in deep neural networks

Xinyu Qian
Department of Computer Science
McGill University

Amir Ozhan Dehghani
Department of Physiology
McGill University

Asa Borzabadifarahani
Department of Physiology
McGill University

Pouya Bashivan
Department of Physiology, Computer Science
McGill University
Mila, Universite de montreal

Abstract:

Across the primate cortex, neurons that perform similar functions tend to be spatially grouped together. In the high-level visual cortex, this widely observed biological rule manifests itself as a modular organization of neuronal clusters, each tuned to a specific object category. The tendency towards short connections is one of the most widely accepted views of why such an organization exists in many animals' brains. Yet, how such a feat is implemented at the neural level remains unclear. Here, using artificial deep neural networks as test beds, we demonstrate that topographical organization similar to that in the primary, intermediate, and high-level human visual cortex emerges when units in these models are laterally connected and their weight parameters are tuned using top-down credit assignment. Importantly, the emergence of the modular organization without any explicit topography-inducing learning rules and learning objectives questions their necessity and suggests that local lateral connections alone may be sufficient for the formation of the topographic organization across the cortex.
Keywords: Topographical organization; Modular learning; Deep convolutional networks

Methods

Locally-laterally connected neural network. We used the ResNet18 architecture [10] with two key changes made to each convolutional layer to implement the local lateral connections: Firstly, inspired by the organization of lateral connections in the visual cortex [3], we incorporated Kernel Pooling (KP) layers [2] in all layers of the model. This layer simulated "cortical sheets" by reshaping units in the kernel dimension into a 2D structure, followed by pooling on these sheets. This approach mirrors the principles of average pooling but applied on the kernel dimension of the layer activations. Secondly, inspired by the physical proximity of neurons across cortical areas that are hierarchically close (e.g. V1 and V2), we replaced traditional zero padding with Continuous Padding which involved appending the activation from the preceding layer to the current layer's activation before applying the KP layer.

Furthermore, we progressively reduced the KP kernel size during the training, enabling topographical organization to emerge at increasingly finer scales. These networks were fully trained to solely minimize the object classification cross-entropy loss on the ImageNet dataset [6]. Several model variations with different choices of KP operation were trained: 1) Kernel Average Pooling (Mean): computes the average of unit activations within a local neighborhood of each unit. 2,3) Kernel Gaussian Pooling (Gaussian), Kernel Mexican-hat Pooling (Mexicanhat): similar to KAP but with a Gaussian and Mexican-hat weighting function respectively.

Results

V1 topography. We first evaluated the topographical similarity of our model with that in the primate V1 by evaluating unit responses to sine grating images of varying orientation, spatial frequency, and color, similar to reference [1]. We observed 1) smoothly changing selectivity when considering each of the three factors (Fig. 1 A); 2) the similarity decayed exponentially with distances (Fig. 1 B); 3) difference in feature selectivity as a function of distance in an early layer (Fig. 1 C); 4) distribution of orientation difference within the laterally connected area. The proportion of orientation difference < 45deg is ~60% which aligns with the experimental observation from [3] (Fig. 1 D); 5) a tendency towards orthogonal angles between spatial frequency and orientation gradients similar to prior experimental work [9] (Fig. 1 E)

IT topography. We next investigated the similarity of topographical organization in deeper layers of the network and IT cortex by quantifying unit selectivity using t-value measure [1]. Unit responses were assessed concerning six distinct categories of images, namely face, scene, body, characters, objects, no-man's land [7] as well as to animacy and size [8].

We observed that 1) continuous and smooth patches selective each of the six categories emerged in the deeper layers of the model (blocks 3-4) that were extended along the shallow-deep axis of the model, similar to typical elongation of category selective patches along the posterior-anterior axis of ventral visual cortex (Fig. 2 A, D); 2) Pairwise unit correlations decayed exponentially as a function of distance (Fig. 2 B); 3) the patch elongation was decreased as a function of how fast the lateral connection range was decayed (Fig. 2 D); 4) patch sizes were modulated by the range of lateral connectivity (Fig. 2 E). Furthermore, the model displayed two parallel streams that encoded animacy of objects and their size, similar to prior observations from human visual cortex [8] (Fig. 3).

Behavioral performance and connectivity. While the model displayed a significant drop in its object recognition performance compared to its non-topographical counterpart, its accuracy was still substantially higher than the state-of-the-art topographical model [1] (TDANN=43.9%, Gaussian=53%, RN18=69.57%; Fig. 4 B).

We made two additional surprising observations: 1) the neural network with lateral connections displayed strong resilience to small pixel perturbations compared to the non-topographical model that also increased with larger lateral connection range (Fig. 4 A; AutoAttack ϵ_{L2} = 1); 2) optimization of the locally laterally connected model on object recognition led to minimization of wiring cost as a byproduct, exceeding other topographical models such as TDANN [1] (Fig. 4 C).

Conclusion. We presented a new topographical deep neural network model by incorporating local lateral connectivity into a typical DNN architecture (RN18) and showed that parameter tuning in this network using backpropagation leads to a fully topographical model that not only reproduces hallmarks of cortical topography in early and high-level primate visual cortex, but also achieves improved object-recognition performance, higher recognition robustness, and reduces the wiring cost. Our work signifies the critical role of local lateral connections in the cortex in shaping its regular organization and also suggests a possible double role of these connections in both wiring cost minimization and robust representation learning.

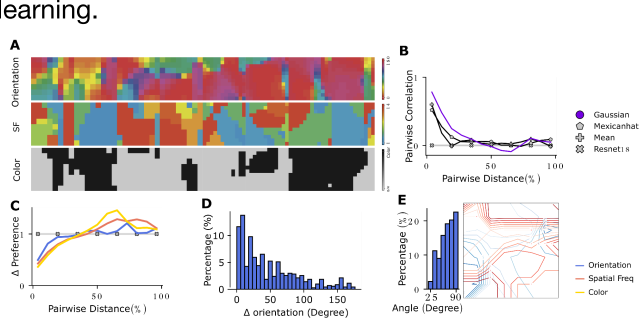

Figure 1: **V1 topography**

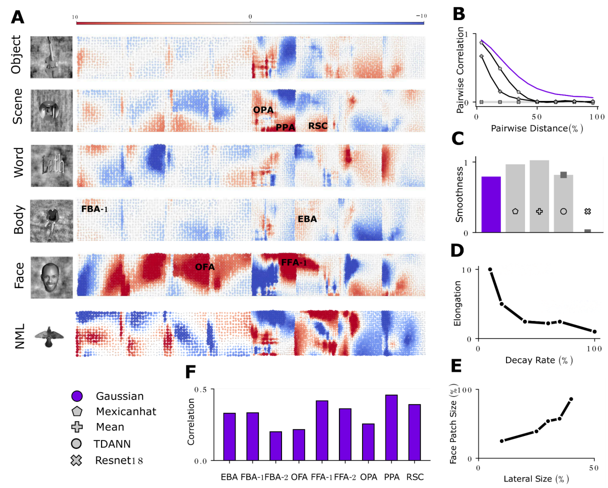

Figure 2: **IT topography**

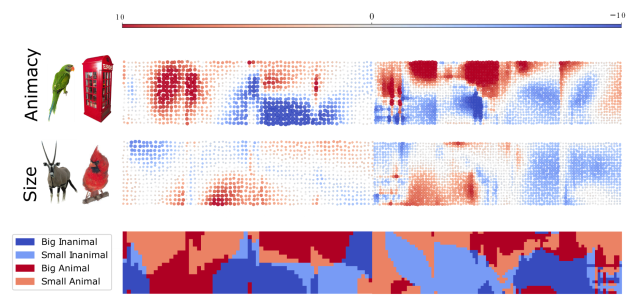

Figure 3: **Animacy and Size**

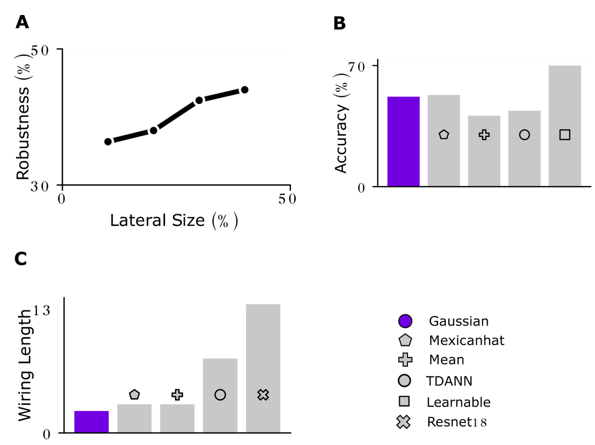

Figure 4: **Behavioral performance and connectivity**

References

[1] Margalit et al. A unifying principle for the functional organization of visual cortex, 2023

[2] Bashivan et al. Learning robust kernel ensembles with kernel average pooling, 2022

[3] Muir et al, Embedding of cortical representations by the superficial patch system, 2011

[4] Blauch et al. A connectivity-constrained computational account of topographical organization in primate high-level visual cortex, 2022

[5] Lu et al. End-to-end topographic networks as models of cortical map formation and human visual behavior: moving beyond convolutions, 2023

[6] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009

[7] Bao P, She L, McGill M, Tsao DY. A map of object space in the primate inferotemporal cortex. Nature. 2020

[8] Konkle T, Caramazza A. Tripartite organization of the ventral stream by animacy and object size. J Neurosci. 2013

[9] Nauhaus I, Nielsen KJ, Disney AA, Callaway EM. Orthogonal micro-organization of orientation and spatial frequency in primate primary visual cortex. Nat Neurosci. 2012

[10] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016