Reset Networks: Emergent Topography in Networks of Convolutional Neural Networks

T. Hannagan[1]

[1] Connecticut Institute for the Brain and Cognitive Sciences, University of Connecticut

Abstract

We introduce Reset networks, which are compositions of several neural networks - typically several levels of CNNs - where possibly non-spatial outputs at one level are reshaped into spatial inputs for the next level. We demonstrate that Reset networks exhibit emergent topographic organization for numbers, as well as for visual categories taken from CIFAR-100. We outline the implications of this model for theories of the cortex and developmental neuroscience.

Introduction

CNN classifiers are scalable and high-performing deep learning models, that have now been shown beyond any reasonable doubt to predict activity in the visual system. However, explaining the existence of categorical areas deep in ventral Occipitotemporal cortex has remained a challenge. This is because categorical areas respond to high level features and yet are spatially extended objects in vOTC, whereas by design, CNN classifiers trade-off spatial dimensions for feature channels as information is fed-forward. In the deepest layers of the network, features have little if any spatial arrangement left.

Another limit of the current CNN-to-visual-cortex mapping endeavor is that most if not all studies attempt to predict cortical responses from a single deep CNN classifier, trained on a single task. Though understandable, these two simplifications nevertheless make the model qualitatively quite different from the visual system, which is shaped by many different tasks other than classification (e.g. visual tracking, naming), and involves different processing streams.

In this article, we introduce a new modeling approach that attempts to factor in the multiplicity of processing streams and versatility of tasks that are characteristic of the visual system. We show that requiring the outputs of many CNNs to serve as input to other CNNs downstream is sufficient for topography to emerge.

Reset networks

Reset networks are compositions of several neural networks - typically several levels of CNNs - whose outputs at one level are reshaped into a spatial input for the next level. They implement a sequence of neural spaces where networks performing similar computations end-up being neighbors, as do units that are selective to the same input.

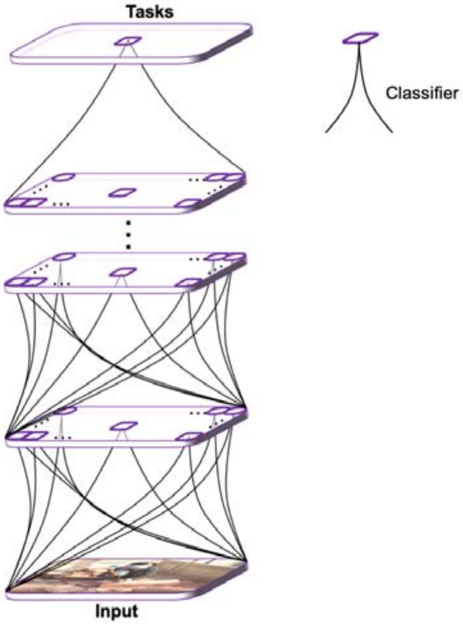

Figure 1. *A Reset network is a differentiable neural network system with an arbitrary number of levels, where each level itself consists of a spatial arrangement of deep neural networks.*

The general form of a Reset network is shown in Figure 1. It has an arbitrary depth of levels, each consisting of several networks operating in parallel on the same input. The next three requirements can be relaxed, but will be followed in the remainder of this article. At any level, all networks are independent processors: they do not share any weight parameters and do not project to each other laterally. All networks in a given grid also receive, as a common input, the entirety of the level below. The last level in a reset network is the only output level, where error signals for all k tasks are received.

Reset networks include in particular the family of depth 2 shown in Figure 2, where level 1 is obtained by reshaping and concatenating the outputs of nxn parallel networks into a single map, called "grid" hereafter, which then serves as input for a final network. We refer to such systems as Reset Networks of depth **2** and width **n**, or Reset(n).

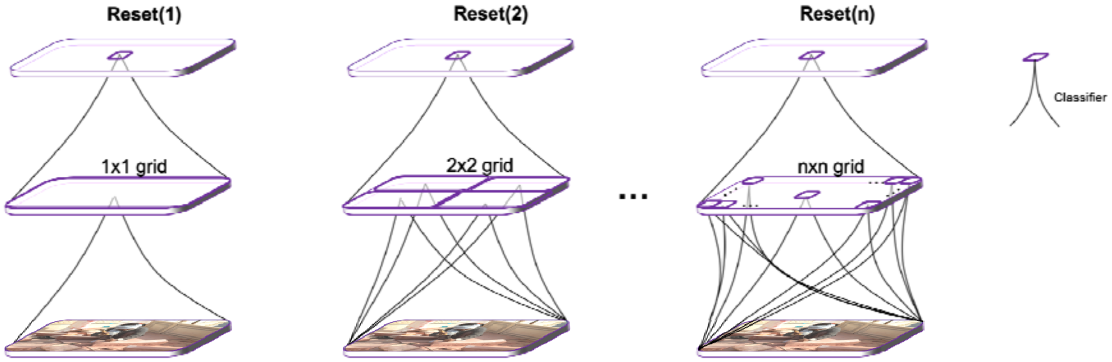

Figure 2. *A family of depth 2 Reset networks, with nxn intermediate grids for increasing n.*

The master network forces the grid of subnetworks underneath to organize in order to solve the task, distributing work in a way that creates topography.

We demonstrate that Reset networks can perform classification and regression at scale while also exhibiting emergent topographic organization. Our code is available on GitHub: https://github.com/THANNAGA/Reset-Networks.

Why are Reset networks relevant to cortical topography?

Cortical topography in the strict sense is the notion that "nearby neurons in the cortex have receptive fields at nearby locations in the world" [1]. When understood as applying also to local fields or voxels as well as to neurons, this is a widespread phenomenon in the brain, imaged throughout the visual cortex as well as in some associative areas.

Despite the architectural tension between CNNs and vOTC, recent innovative work has shown that categorical areas can indeed be simulated in topographic variants of deep CNNs, tCNN [2]. In tCNNs, topography is achieved by invoking a separate entity –dubbed “cortical tissue map”- assigning arbitrary locations on this map to units in the dense layer of the network, before introducing a loss regularizer that penalizes wiring length on the map during training. Since the mechanism realizing this mapping is unspecified, the ontological status of space in the model is problematic. Two different notions of space appear to exist that can contradict each other: the spatial layout of convolutional feature maps in the model, and the spatial dimensions of the cortical tissue map. This tension is not manifest in tCNN because cortical tissue maps are restricted to the upper dense layers of the model, where locality is lost. However, there is no reason why cortical tissue maps couldn’t also be invoked for the lower, convolutional levels of the network, with much less interpretability.

Results

Topography for numbers in parietal cortex

In parietal cortex, voxels selective for similar numbers are more likely to be contiguous, a phenomenon which has been partially explained as a planar diffusion process of number codes, due to an underlying locally and randomly connected network of cortical units [3]. This network, however, did not process real stimuli. As Figure 3 shows, a Reset Network with a single 8x8 grid, can be trained to map images of numbers onto number codes, and succeeds in reproducing topographic organization.

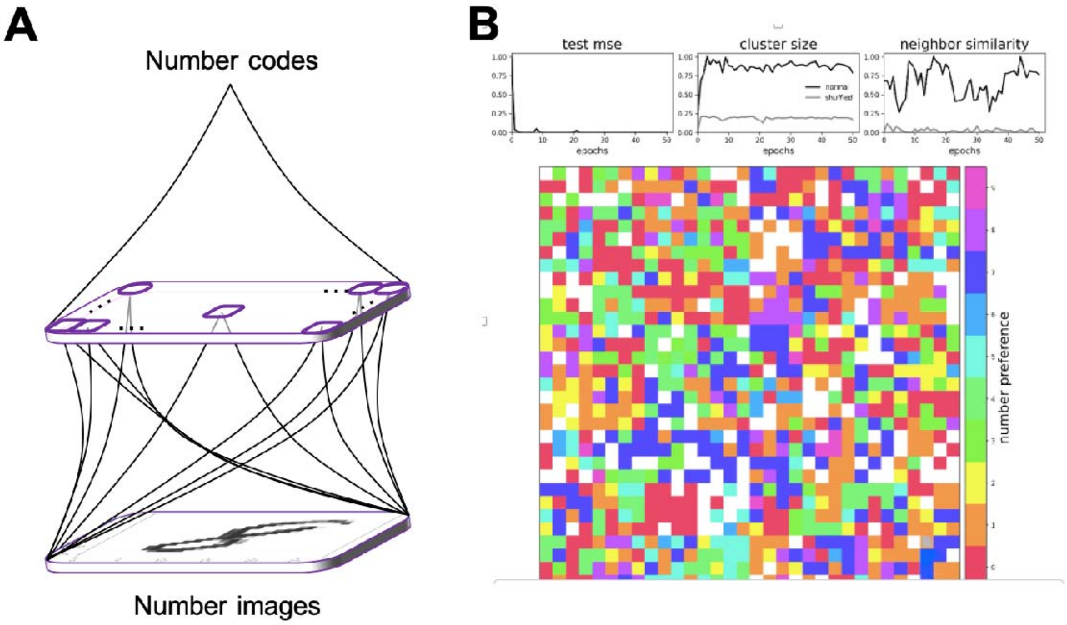

Figure 3. *(A) A depth 2 Reset network with 8x8 grid is trained to map images of numbers onto number codes. (B) Number preferences on the network grid show topography, quantified by cluster size (upper middle plot) and neighbor similarity (upper right plot), both significantly above the values for shuffled maps.*

Topography is clearly visible on the map of number preferences, and is quantified in the middle plot above, where it can also be seen to emerge quickly during training. Topography and neighborhood similarity (right plot) are both quite significantly above the levels obtained for the same selectivity maps, but shuffled. Also notable is the tendency of subnetworks to specialize for specific numbers, or numbers in the same ballpark.

Topography for categorical areas in ventral occipitotemporal cortex

In ventral occipitotemporal cortex, more than two decades of studies have established the presence of areas selective for various widespread visual categories, in particular faces, bodies, tools, houses, and words. While there is no shortage of computational models able to reproduce many characteristics of the visual system, including some of vOTC, only one [2] arguably achieves both topography and scale at the same time - with topography being problematic, requiring two different notions of space to coexist. By contrast, the way Reset networks achieve topography at scale is conceptually straightforward.

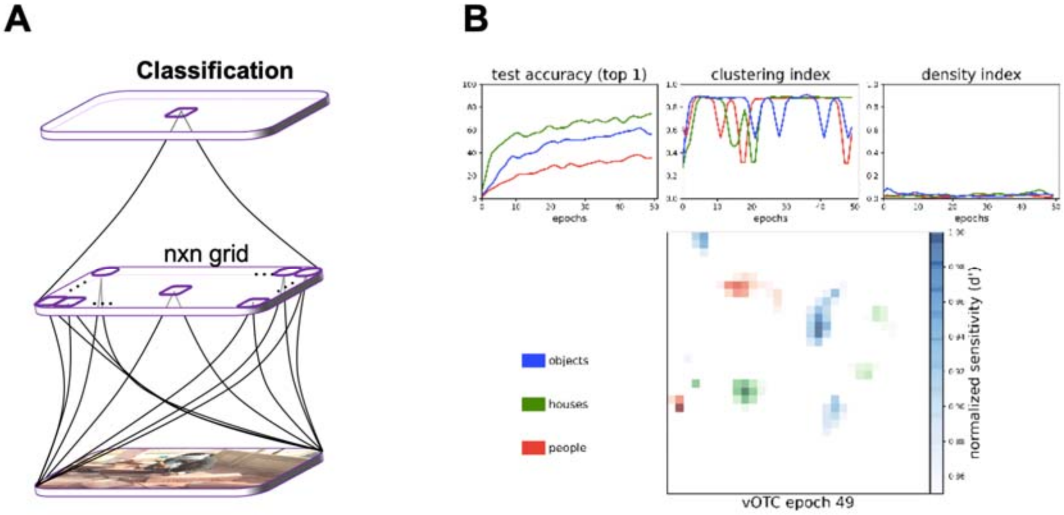

Figure 4. *(A) A depth 2 Reset network is trained on CIFAR-100. (B) Unit preferences on the network grid show clusters for objects, houses and people. These are quantified by means of a clustering index (upper middle plot) and density index (upper right plot).*

The left panel in Figure 4 shows a Reset network classifier trained on Cifar-100. The right panel shows category preferences on the grid after training. Only 3 categories are considered - objects, houses and people - which were obtained by aggregating the relevant Cifar-100 classes. Clustering is visible in the map, and quantified in the subplots above (although with different indicators as before for numerotopy).

VOTC topography and the Visual Word Form Area

A closely related topic is that of the so-called Visual Word Form Area, which, with the benefit of insight and despite its discoverers' best intent upon naming it, is neither visual (congenital blind subjects have it too), word-level specific (it is also active for individual letters), nor actually a single area (it appears to be organized in patches). But names have great inertia, and this one does convey well the idea of a localized region selective for stimuli related to words. While some efforts have gone into modeling the VWFA [4], currently no model can account for its specific place within the topography of vOTC. The network in Figure 6 describes what a Reset network of vOTC and the VWFA could look like.

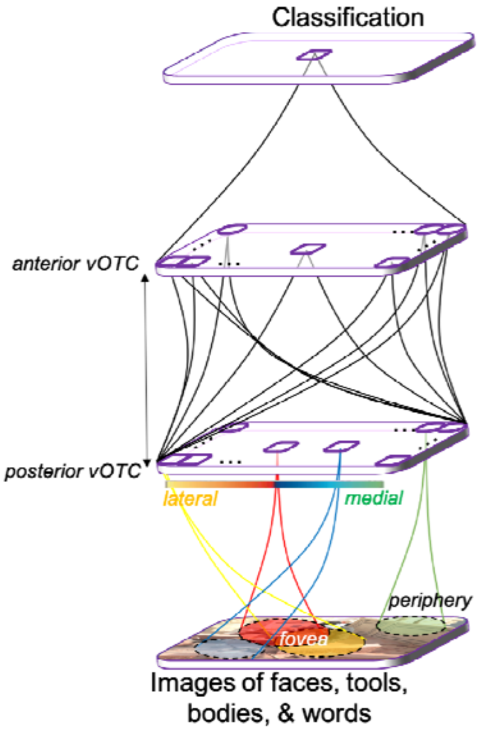

Figure 5. *A Reset network of depth 3 for vOTC and the VWFA. The network has 2 intermediate grids, standing for posterior and anterior vOTC. Networks in the posterior grid see different parts of the input depending on where they are: left-located ("lateral") networks on the grid receive input from the center of the image, whereas right-located ("medial") networks receive input from its periphery.*

First, this Reset network would have 2 intermediate grids, P and A, standing for the posterior and anterior axis in vOTC. This is not an innovation, but now Reset networks allow for something interesting to happen. In addition to the posterior-to-anterior gradient, we can capture a lateral-to-medial gradient by ensuring that networks in the P grid see different parts of the input depending on where they are: left-located (lateral) networks on the P grid would receive input from the center of the image, whereas right-located (medial) networks would receive input from the periphery. In other words, we build into the model a lateral-to-medial gradient in vOTC by exploiting its well-documented correspondence with center/periphery processing [5]. Such a relation cannot easily be built into a CNN, because of location invariance.

Discussion

Classification performance

We have showed that Reset networks can classify standard computer vision datasets such as CIFAR-100. However and as the figure below shows, at this stage their performance remains disappointing, only at best matching that of a single Resnet 20, while having many more parameters.

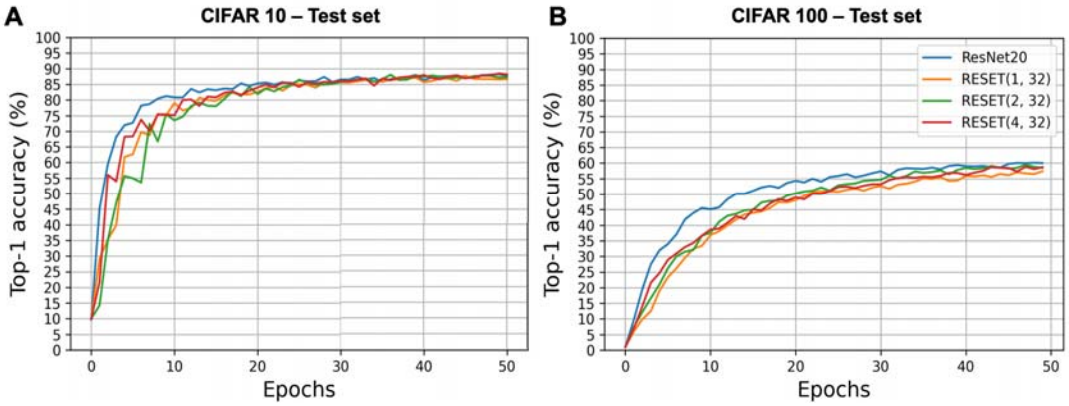

Figure 6. *Top-1 accuracy of Reset Networks on CIFAR-10 (A) and CIFAR-100 (B). Composing several CNN classifiers does not destroy classification abilities, though unsatisfyingly, Reset networks can currently at best only converge to the same performance as a single Resnet20. The notation Reset(n, 32) specifies that the outputs of nxn networks are reshaped into a map of 32x32 units.*

One reason for this could be that in our simulations, spatial resets between levels were always done by reshaping the subnetworks' outputs, which constitute an information bottleneck. Reshaping prior to the subnetwork's output, e.g. the dense layer or before, might be a more astute choice. We also observe that the full resources of the Reset network don't seem to be used: some subnetwork units are more active than others. This can be alleviated to some extent by using dropout, or another kind of regularization on the grid.

Regularization by auto-encoding

In the course of our investigations (not quantitatively reported here), we have observed that Reset networks that were based on smaller subnetworks than Reset20, performed much better when the second level had 2 networks: one that classified the input, and another that tried to reconstruct the input from the grid. Auto-encoding in this situation appears to act as an efficient regularizer for classification, forcing the error gradient to be distributed across the whole grid rather than to be drawn by one, or just a few subnetworks. Such regularization effects of auto-encoding have been reported before for standard classifiers [6].

Topography

Reset networks constitute a novel mechanism for topography to emerge in deep learning. We have presented solid evidence that they can reproduce at least two examples of topographic organization: in parietal cortex for numbers, and in ventral Occipitotemporal cortex for the so-called "categorical areas". A related point is that Reset networks provide a way to implement a cortical gradient, the mapping between foveal/peripheral input and lateral/medial in visual cortex, which is not easily captured within the standard assumptions of CNNs.

Adding networks when necessary for continual learning: the width and depth of Reset networks

The proposed approach aligns well with a view of neural development in which, as an alternative to recycling neural material, new resources can also be recruited in the system if needed. Learning a new task could require only to widen the system by adding a network at the current level, with different networks possibly trained on different tasks. If expertise from previously learned tasks is required, the system could be made deeper by reshaping network outputs at the current level and creating a new level. By better specifying the mechanisms of network growth without interference of functions, Reset networks can help investigate continual learning theories.

Conclusion

Reset networks show that topography must emerge in deep CNN classifiers, when these are composed with one another. In this view, the topographic cortex should not be modeled as a single classifier, however deep and richly organized, but as a sequence of levels of neural network classifiers. This rests on the idea that the cortex has the ability to compose networks if need be, and predicts that the outputs, or the late computational stages, of cortical classifiers are either spatially organized, or somehow reshaped spatially during the course of composition.

Acknowledgments

We thank Thibault Fouqueray for early and very stimulating discussions where this idea was seeded, and Florence Bouhali for ongoing discussions, original ideas and encouragements on Reset networks.

References

[1] Patel GH, Kaplan DM, Snyder LH. Topographic organization in the brain: searching for general principles. Trends Cogn Sci. 2014;18(7):351-363. doi:10.1016/j.tics.2014.03.008.

[2] Lee H, Margalit E, Jozwik KM, Cohen MA, Kanwisher N, Yamins DL, DiCarlo JJ. Topographic deep artificial neural networks reproduce the hallmarks of the primate inferior temporal cortex face processing network. 2020 bioRxiv.

[3] Hannagan T, Nieder A, Viswanathan P, Dehaene S. A random-matrix theory of the number sense. Phil. Trans. R. Soc. B. 2018;373:20170253. doi:10.1098/rstb.2017.0253

[4] Hannagan T, Agrawal A., Cohen L, Dehaene S. Emergence of a compositional neural code for written words: Recycling of a convolutional neural network for reading. Proceedings of the National Academy of Sciences Nov 2021, 118 (46) e2104779118; doi: 10.1073/pnas.2104779118.

[5] Op de Beeck HP, Pillet I, Ritchie JB. Factors Determining Where Category-Selective Areas Emerge in Visual Cortex. Trends Cogn Sci. 2019 Sep;23(9):784-797. doi: 10.1016/j.tics.2019.06.006.

[6] Le L, Patterson A, White M. Supervised autoencoders: Improving generalization performance with unsupervised regularizers. In Advances in Neural Information Processing Systems. 2018. 107–117.