Neural Population Control via Deep ANN Image Synthesis

Pouya Bashivan*, **Kohitij Kar***, **and James J. DiCarlo**
(bashivan,kohitij,dicarlo)@mit.edu
McGovern Institute for Brain Research and
Department of Brain and Cognitive Sciences
MIT
Cambridge, MA 02139, USA

Abstract

Specific deep feed-forward artificial neural networks (ANNs) constitute our current best understanding of the primate ventral visual stream and the core object recognition behavior it supports. Here we turn things around and ask: can we use these ANN models to synthesize images that *control* neural activity? We here test this control in cortical area V4 in two control goal settings. i) *Single Neuron State Control*: ”stretch” the maximal firing rate of any single neural site beyond its naturally occurring maximal rate. ii) *Population State Control*: independently and simultaneously control all neural sites in a *population*. Specifically, we aimed to push the recorded population into a ”one hot” state in which one neural site is active and all others are clamped at baseline. We report that, using ANN-driven image synthesis, we can drive the firing rates of most V4 neural sites beyond naturally occurring levels. And we report that V4 neural sites with overlapping receptive fields can be partly – but not yet perfectly – independently controlled. These results are the strongest test thus far of deep ANN models of the ventral stream, and they show how these models could be used for setting desired brain states.

Keywords: Neural control, Visual Cortex, Object Recognition, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), Image Synthesis

Introduction

It has been claimed that particular deep feed-forward artificial neural networks (ANNs) constitute a good – but not yet perfect – understanding of the primate ventral visual stream and the core object recognition behavior it supports. This claim is based on the finding that the internal neural representations of these particular ANNs are remarkably similar to the neural representations in mid-level (area V4) and high-level (area IT) regions of the ventral stream (D. L. K. Yamins, Hong, & Cadieu, 2013; D. L. Yamins et al., 2014; Khaligh-Razavi & Kriegeskorte, 2014), a finding that has recently been extended to neural representations in visual area V1 (Cadena et al., 2017) and to patterns of behavioral performance in core object recognition tasks (Rajalingham, Schmidt, & DiCarlo, 2015; Rajalingham et al., 2018).

However, at least two major potential limitations of this claim have been raised. First, because the evaluation images are randomly sampled from the same visual ”diet” as that used to set the models internal parameters (photograph and rendered object databases), it is unclear how far these models can extrapolate their brain predictions. Second, because the models are somewhat complex and have only been evaluated in terms of predictive similarity (above), it has been argued that they do not constitute an understanding of the ventral stream. A productive variant of this criticism is: what new things can these ANN models enable one to do?

Here we aimed to experimentally assess these potential limitations of ANN ventral stream models. Specifically, we used a deep ANN model to synthesize images that each are specifically targeted to control V4 neural firing activity in two settings. i) *Neural ”Stretch”*: synthesize images that ”stretch” the maximal firing rate of any single neural site well beyond its naturally occurring maximal rate. ii) *Neural Population State Control*: synthesize images to independently control every neural site in a small recorded population (75 V4 neural sites). Specifically, we tested such population control by aiming to set the population in an experimenter-chosen ”one hot” state in which one neural site is pushed to be highly active while all other nearby sites are simultaneously ”clamped” at their baseline activation level.

Methods

Electrophysiological Recordings in Macaques

We sampled and recorded neural sites across the macaque V4 cortex in the left and right hemisphere of two awake, behaving macaques, respectively. In each monkey, we implanted one chronic 96-electrode microelectrode array (Utah array), immediately anterior to the lunate sulcus (LS) and posterior to the inferior occipital sulcus (IOS), with the goal of targeting the central visual representation ($<5^\circ$ eccentricity, contralateral lower visual field). Each array sampled from $\sim$25 mm^2 of dorsal V4. Recording sites that yielded a significant visual drive, and high image rank-order response reliability ($r_{pearson} > 0.8$) across trials were considered for further analyses. In total, we recorded from 75 valid V4 sites which included 50 and 25 sites in the left and right hemisphere of monkey M and monkey N (shown as inset in Figure 1), respectively.

We do not assume that each V4 electrode was recording only the spikes of a single neuron. But we did require that the spiking responses obtained at each V4 electrode maintained stability in their image-wise ”fingerprint” between the day(s) that the mapping images were tested and the days that the Controller images were tested (see below). Specifically, we required an image-wise correlation of at least 0.8 tested on a set of 25 naturalistic images that were shown every day

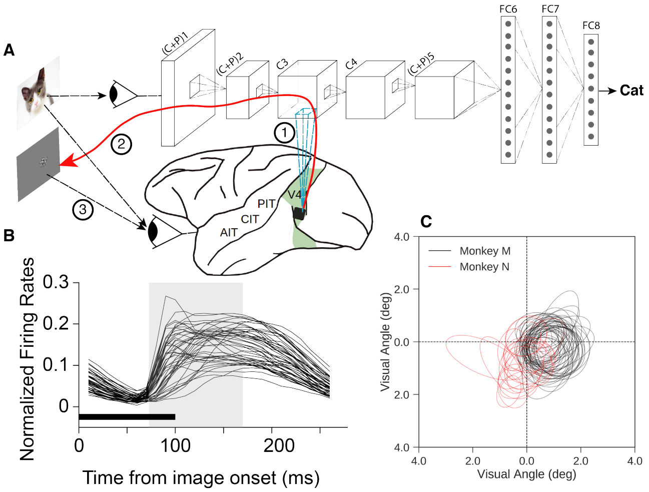

Figure 1: Overview of the synthesis procedure. A) The neural control experiments were done through four steps. (1) a fixed set of ANN features are mapped to the recorded set of V4 neural sites are used to create a predictive model of the activity of those neural sites. (2) That differentiable model is then used to synthesize ”controller” images for either single-site or population control. (3) These synthetic controller images are then shown to the animal to evaluate the degree of control over the neural sites. B) Response trajectories of multiple V4 neural sites to one example image (averaged over ∼40 repetitions of that image). Wide black line is the image presention time. Shaded area is the time window over which the activity level of each V4 neural site is computed (i.e. one value per neuron per image). C) Receptive fields of neural sites in monkey M (black) and Monkey N (red; see Methods).

(normalizer set). Each neuron’s firing rate in each recording session was normalized by subtracting the mean and dividing by the standard deviation of same neuron’s response to the normalizer set.

Passive Viewing Task

During the passive viewing task, monkeys fixated a white square dot (0.2°) for 300 ms to initiate a trial. We then presented a sequence of 5 to 7 images, each ON for 100 ms followed by a 100 ms gray blank screen. This was followed by a water reward and an inter-trial interval of 500 ms, followed by the next sequence. Trials were aborted if gaze was not held within ±1° of the central fixation dot during any point. To estimate the receptive fields (RF) of the neurons, we flashed 1°x1° white squares across the central 10° of the monkeys’ visual field and measured the corresponding neural responses.

Natural Images

We used a large set (N=640) of naturalistic images to measure the response of each recorded V4 neuron and every model V4 neuron to each of these images. These images each contained a three-dimensional rendered object instantiated at a random view overlaid on an unrelated natural image in the background, see (Majaj, Hong, Solomon, & DiCarlo, 2015) for details.

V4 encoding model

To use the ANN model to try to predict and control each recorded neural site (or neural population), the internal V4-like representation of the model must first be mapped to the specific set of recorded neurons. The assumptions behind this mapping are discussed elsewhere (D. L. Yamins & DiCarlo, 2016), but the key idea is that any good model of a ventral stream area must contain a set of artificial neurons (a.k.a. features) that, together, span the same visual encoding space as the brains population of neurons in that area (i.e. the model layer must match the brain area up to a linear mapping). To build this predictive map from model to brain, we started with a specific ANN model with locked parameters. (Here we used the Alexnet architecture trained on Imagenet (Krizhevsky, Sutskever, & Hinton, 2012) as we have previously found the feature space at the output of Conv-3 layer of Alexnet to be a good predictor V4 neural responses. We here refer to this as model ”V4”.) We used the responses of the 75 recorded V4 neurons and the responses of all the model ”V4” neurons to build a mapping from model to the brain. To do this, we used principal component regression (PCR). Specifically, we split the neural data into two equally sized sets along the images axis (train and test). The train set was used to optimize the parameters of the mapping function (PCR) and the test set was used to evaluate the prediction error of the mapping function.

To further reduce the mapping error we used a variant of the 2-stage convolutional mapping function proposed in (Klindt, Ecker, Euler, & Bethge, 2017). The resulting predictive model of V4 (ANN features plus linear mapping) is referred to as the *mapped v4 encoding model* and, by construction, it contains the same number of artificial V4 neurons as the number of recorded V4 neurons (50 and 25 neurons in monkeys M and N respectively).

Synthesized ”Controller” Images

Each artificial neuron in the *mapped V4 encoding model* (above) is a differentiable function of the pixel values $f : I^{w \times h \times c} \to \mathbb{R}^n$ that enables us to use the model to analyze the sensitivity of neurons to patterns in the pixels space. We then defined the synthesis operation as an optimization procedure during which images are synthesized to control the neurons firing patterns in the following two scenarios.

1. Stretch: We synthesize controller images that attempt to push each individual V4 neural site into a maximal activity state. To do so, we iteratively change the pixel values in the direction of gradients that maximizes the firing rate of the corresponding artificial V4 neuron. We repeated the procedure for each neuron using five different random start seeds, thereby generating five ”stretch” controller images for each V4 neural site.

2. One Hot Population: Similar to ”Stretch” scenario, except that here we constrain the optimization to change the

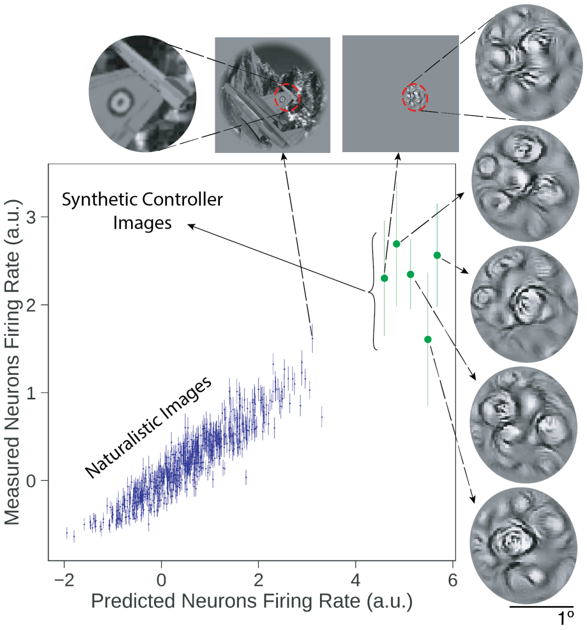

Figure 2: Results for an example successful "stretch" control test. Normalized activity level of the target V4 neural sites is shown for all of the naturalistic images (blue dots) and for its five synthetic "stretch" controller images (green dots; see Methods).

pixel values in a way that i) attempts to maximize FR of the target V4 neural site, and ii) attempts to maintain the predicted FR of all other recorded V4 neurons at their response baseline (here defined as their responses to pixel noise images that the optimization runs starts from). This procedure was performed once for each of the 75 neurons as the target neuron.

For each optimization run, we start from an image that consists of random pixel values drawn from a standard Normal distribution and optimize the objective function for a prespecified number of steps using gradient descent algorithm (steps=700). The procedure was considered to have failed when it was unable to find an image predicted to drive the target neural site beyond its predicted response to noise images. This procedure successfully maximized the objective function for most target neural sites ($\sim$95% for "stretch" and 65-70% for "OHP"; see Table 1).

Results

We recorded 75 neurons from area V4 in two monkeys and constructed a predictor model of the firing rate of each of those neurons (See Methods). We then used a synthesis procedure (see methods section) to generate "controller" images that each attempted to control the recorded V4 population. To test for successful control, we recorded from the same neural sites in V4 on subsequent days in response to the synthesized "controller images" (see Methods).

For the "Stretch" controller tests, we found that we could successfully drive $\sim$70% of neural sites beyond their maximal observed firing rates (measured over 640 naturalistic images; see Table 1). For the "One-hot population" controller tests, we generally were able to achieve reductions in the activity of the "off target" neural sites, even while maintaining a high response of the target neural site (see examples in Figure 4).

Table 1: Summary of results. A total of 45 and 17 reliable neural sites were measured in monkeys M and N respectively (see Methods). The first data column shows the fraction of control attempts that failed because the synthesis procedure failed (see Methods; no further neural recording was attempted in these cases). The second and third columns shows two assessments of control success. 95-Per = fraction of tested neural target sites for which at least one of the five controller images pushed the target neuron beyond 95% of its maximal natural rate. "Stretched" = fraction push beyond the maximal natural rate. Numbers in parentheses = number of target neural sites in each case.

Monkey	**Opt. Type**	**Failed Opt.**	**95-Per.***	**Stretched***
M	Stretch	4%(2)	89%(40)	64%(29)
	OHP	35%(16)	53%(24)	31%(14)
N	Stretch	6%(1)	82%(14)	70%(12)
	OHP	29%(5)	53%(9)	29%(5)

In both monkeys the synthesized images drove the firing rate of majority of neurons beyond the previously observed values. Since our measurements were spanning several days, we further removed neurons that their responses were not correlated across days from the analysis (n=5 in monkey M and n=7 in monkey N). Figure 2 shows the predicted and actual neural responses to these images as well as naturalistic images that were initially used to determine the mapping function. Images generated from different random seeds look perceptually similar while they all drive the target neuron very high. Table 1 summarizes the number of neurons for which this procedure successfully drives the neuron's response beyond previously observed values. In both monkeys, a large set of neurons were successfully driven to the 95-percentile of responses previously seen. Figure 3 shows several images generated for 5 different neurons.

Discussion

We report that, using an ANN-model of the ventral stream, we can drive the firing rates of most V4 neurons beyond naturally occurring levels. And we find evidence that this procedure can be extended to control the population (e.g. one hot states). We believe that these results are the strongest test thus far of deep ANN models of the ventral stream, and they show how these models might be used to set desired brain states.

Acknowledgments

This research was supported by Intelligence Advanced Research Projects Agency (IARPA), the MIT-IBM Watson AI Lab, US National Eye Institute grants R01-EY014970 (J.J.D.), and Office of Naval Research MURI-114407 (J.J.D).

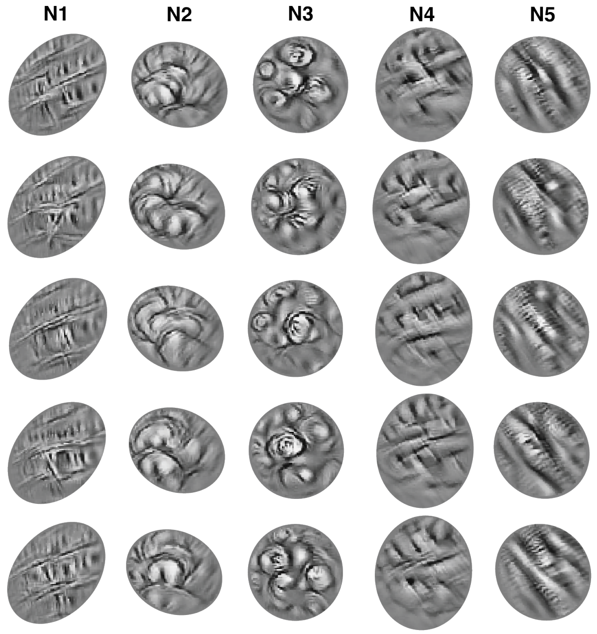

Figure 3: ”Stretch” controller images for five example V4 neural target sites (N1-N5) with overlapping receptive fields. Each row shows images generated using the same starting noise image, but optimized for each of the target sites. Note the perceptual similarity of the controller images for each site and the perceptual dissimilarity across different sites.

References

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., & Ecker, A. S. (2017). Deep convolutional models improve predictions of macaque V1 responses to natural images. *bioRxiv*, 201764.

Khaligh-Razavi, S. M., & Kriegeskorte, N. (2014). Deep Supervised, but Not Unsupervised, Models May Explain IT Cortical Representation. *PLoS Comp. Bio.*, *10*(11).

Klindt, D., Ecker, A. S., Euler, T., & Bethge, M. (2017). Neural system identification for large populations separating what and where. In (pp. 3509–3519).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*.

Majaj, N. J., Hong, H., Solomon, E. A., & DiCarlo, J. J. (2015). Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. *J. of Neuroscience*, *35*(39), 13402–13418.

Rajalingham, R., Issa, E. B., Bashivan, P., Kar, K., Schmidt, K., & DiCarlo, J. J. (2018). Large-scale, high-resolution comparison of the core visual object recognition behavior of humans, monkeys, and state-of-the-art deep artificial neural networks. *bioRxiv*, 240614.

Rajalingham, R., Schmidt, K., & DiCarlo, J. J. (2015). Comparison of Object Recognition Behavior in Human and Monkey. *J. of Neuroscience*, *35*(35), 12127–12136.

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neuroscience*, *19*(3), 356–365. doi: 10.1038/nn.4244

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS*, *111*(23), 8619–8624.

Yamins, D. L. K., Hong, H., & Cadieu, C. (2013). Hierarchical Modular Optimization of Convolutional Networks Achieves Representations Similar to Macaque IT and Human Ventral Stream. *NIPS*.

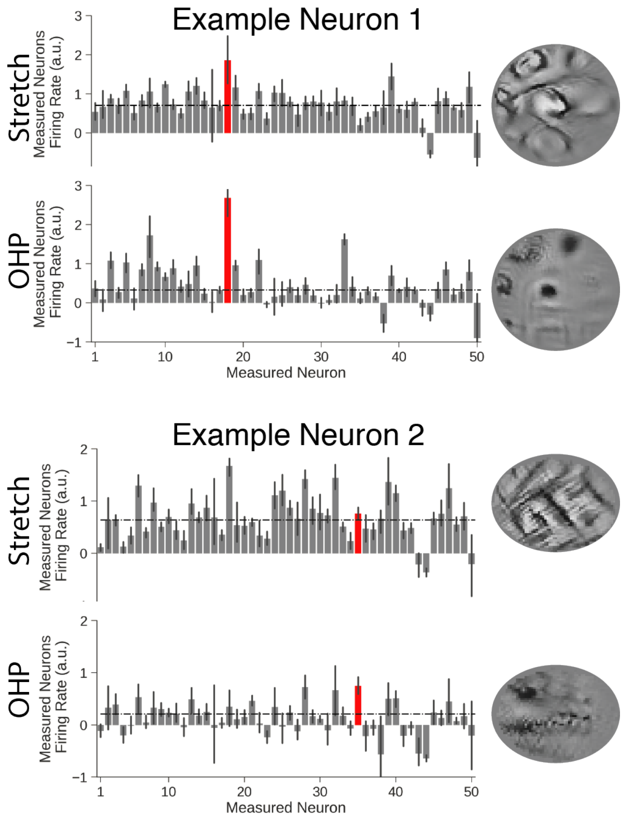

Figure 4: **Tests of Neural Population State Control.** Here we synthesized controller images that attempted to push the neural population into two different one hot states in which the target neural site (red) is active and all other recorded neural sites (gray) are at baseline (most of these neural sites have overlapping receptive fields). Neural activity states are shown for all V4 neural sites for controller images synthesized using the ”stretch” and ”one-hot population” (OHP) goals. Each neural site’s plotted activity level is normalized so that zero is the noise image activity level (our defined baseline), and 1.0 is the natural image maximal activation level. Compared with the ”stretch” controller images, we found that the OHP controller images were more selective in activating the target V4 site (the horizontal dashed line indicates the median ”off target” neural activity level)