Uncovering the Spatiotemporal Dynamics of Goal-driven Efficient Coding with a Brain-supervised Sparse coding Network

Bruce C. Hansen, Isabel S.H. Gephart, Victoria E. Gobo
(bchansen@colgate.edu; igephart@colgate.edu; vgobo@colgate.edu)
Department of Psychological & Brain Sciences, Neuroscience Program, Colgate University, Hamilton NY

Michelle R. Greene (mgreene2@bates.edu)
Neuroscience Program, Bates College, Lewiston ME

David J. Field (djf3@cornell.edu)
Department of Psychology, Cornell University, Ithaca NY

Abstract:

The early visual system is believed to build visual representations based on a dynamic code that is flexibly adapted to suit the behavioral goals of the observer. Efficient coding theories posit that redundant input information is compressed. However, the evidence in support of that process is largely descriptive. Here, we propose that information reduction can be achieved through task-relevant dynamic sparse-distributed coding. Using high-density EEG signals from humans engaged in different scene understanding tasks, we developed a brain-supervised sparse coding network to learn visual filters based on EEG variance over time. Stimuli were reconstructed based on filters that were sparsified according to the brain variance across time. The results show that information reduction via sparsification does not follow a simple 'many-to-few' operation, but instead differentially compresses and expands information over time in a task-dependent manner. This suggests that efficient coding may not result in a terminal state of compression, but instead operates to yield a goal-oriented state of dynamic equilibrium.

Keywords: Sparse-distributed coding; efficient coding; scene perception; electroencephalography (EEG)

Background

Recent advancements in systems neuroscience suggest that the early visual system does not represent real-world scenes in a static and deterministic manner (like a picture), but instead renders a time-varying code that is suited to support the behavioral goals of the observer (Hansen et al., 2021; Zhan et al., 2018). However, the processes that dictate how such coding takes place have only been loosely explored. One suggestion is that efficient coding is achieved through selective information reduction (Zhan et al., 2018), which closely follows earlier conjectures by Horace Barlow for recasting redundancy reduction as it relates to how the observer needs to use information (Barlow, 1961; 2001). In an elegant study, Zhan and colleagues (2018) provided compelling evidence for rapid task-related information reduction beginning at ~170ms. However, their account was largely descriptive (and limited to a single task and stimulus image) and therefore offered little insight into the neural processes that might support this form of information reduction.

The current study proposes that task relevant information reduction may be achieved through sparse-distributed coding, which has received recent support for carrying scene relevant information (Tang et al., 2018; Zhang et al., 2022). We therefore set out to explore the spatiotemporal dynamics of that sparse code as it relates to different task demands. To do this, we first developed a brain-supervised sparse coding network (BSSCN) to learn visual filters based on both the stimulus and task-related neural response variance. We then used brain-guided sparse reconstruction of the stimuli to assess the task-relevant sparse code over time.

Method

Participants (n = 24) viewed repeated presentations of 80 scenes while making cued assessments about either the presence of an object in the scene, or whether the scene afforded the observer to perform a particular action (both responses provided on 1-4 scale).

While participants performed each task, neural data were gathered via 128-channel EEG in a standard visual evoked potential (VEP) paradigm, whereby each image was presented (500 ms) 30 times (random order) and blocked by task for a total of 2 X 80 X 30 = 4800 trials split over four recording sessions (counter-balanced across task blocks).

Brain-supervised Sparse-coding Network

We based our network on the original sparse-distributed coding network (Olshausen & Field, 1996) that learns and sparsifies visual filters conditioned on scene patches, but augmented to sparsify according to

the variance across VEP signals recorded across scenes. The primary concept behind the original network can be conveyed through its cost function, stated formally below.

$$E = \sum_x \left[I(\vec{y}) - \sum_i a_i \phi_i(\vec{y}) \right]_x^2 + \lambda \sum_i S\left(\frac{a_i}{\sigma}\right)$$

The first term represents the sum of the squared error between a given image patch, $I(\vec{y})$, and the reconstruction of that patch obtained by taking the weighted sum of each filter, $\phi_i(\vec{y})$, and its activation coefficient, a_i. The network minimizes reconstruction error by finding a solution with a limited number of active filters with the second term (regularization term) whereby activation coefficients are passed through a cost function: $S = \ln[1 + (a_i/\sigma)^2]$, ($\sigma$ is a constant), with λ acting to control the amount of sparsification. Activation coefficients are sparsified across a given set of image patches via gradient descent with the filters being updated across subsequent image patch sets.

Brain supervision was incorporated during the sparsification of the activation coefficients:

$$a_i = a_i + \epsilon \left\{ \emptyset\varepsilon_i - \left[\lambda \left(S'\left(\frac{a_i}{\sigma}\right) \pm gdS'_i \right) \right] \right\}$$

Here, $\emptyset\varepsilon_i$ is the error that each filter introduced into the reconstruction, with S' being the first derivative of S and ϵ being the learning rate. The gdS'_i variable incorporates the brain supervision, achieved via weighted gradient descent: $gdS'_i = S'_i * S\rho_i * Err_i$. The Err_i variable is the residual error between the activation coefficients and the brain variance, with $S\rho_i$ being the brain sparsification function defined below.

$$S\rho_i = \frac{\ln\left[1 + \left(\frac{\rho_i}{r_{lm}}\right)^x\right]}{arg\,max(S\rho_i)}$$

The ρ_i variable is the correlation coefficient between the activation coefficients and the brain variance, x is an exponent that controls the extent of the sparsification, and r_{lm} is a dynamic threshold operator that scales with the improvement of the network.

Results

The BSSCN successfully learned filters based on the supervision signal. To demonstrate that the brain supervision does not result in random perturbations to the original sparse coding network, Figure 1 shows filters learned from a low spatial frequency (LSF) biased signal (left) or a high spatial frequency (HSF) biased signal (right).

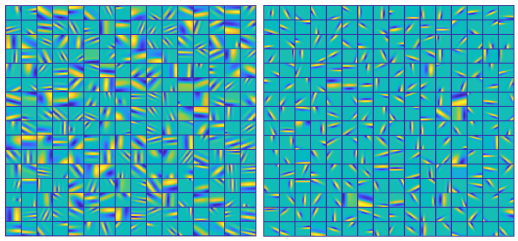

Figure 1. BSSCN concept proof.

Stimuli were reconstructed by filters learned at different time points and sparsified according to the corresponding brain variance. To examine task differences in sparsification over time, the reconstructions were first converted to 2-D filter count histograms (number of filters that contributed to prominent content at each pixel – examples for one image shown in Figure 2). The histograms were then fed through an algorithm that generated a single normalized sparsification metric (SM) for each image and task (0 = high sparsification; 1 = low sparsification).

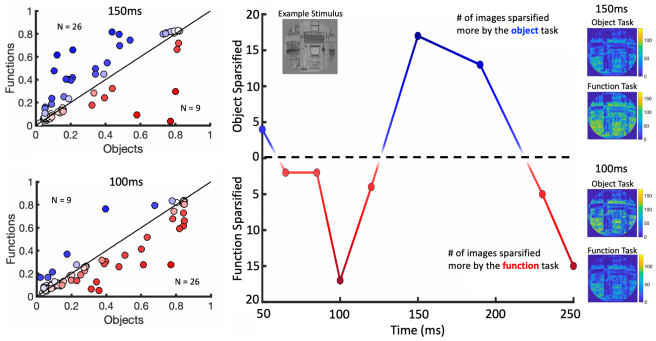

Figure 2. Left side: Scatter plots showing the object task SMs against the function task SMs at two time points. Each point represents one of the 80 images. Images below the reference line are sparsified more by the function task, with images above the reference line being more sparsified by the object task. Right side: summary plot showing the number of images that were more sparsified by the function task (red) or the object task (blue) at several time points.

The results of that analysis (Figure 2) show that information compression via sparse coding does not follow a simple high-to-low information content over time, but instead varies in a semi-periodic task-dependent manner. That is, for the tasks used here, there does not appear to be a terminal point of selective reduction, suggesting that efficient coding may operate to yield a goal-oriented state of dynamic equilibrium.

Acknowledgments

James S. McDonnell Foundation grant (220020430) to BCH; National Science Foundation grant (1736394) to BCH and MRG.

References

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. In *Sensory Communication*, MIT Press.

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems, 12,* 241-253.

Hansen, B.C., Greene, M.R., & Field, D.J. (2021). Dynamic electrode-to-image (DETI) mapping reveals the human brain's spatiotemporal code of visual information. *PLOS Computational Biology, 17,* e1009456.

Olshausen, B.A. & Field, D.J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607-609.

Tang, S., Zhang, Y., Li, Z., Li, M., Liu, F. et al. (2018). Large-scale two-photon imaging revealed super-sparse population codes in the V1 superficial layer of awake monkeys. *eLife, 7,* e33370.

Zhan, J., Ince, R.A.A., van Rijsbergen, N., & Schyns, P.G. (2018). Dynamic construction of reduced representations in the brain for perceptual decision making. *Current Biology, 29,* 219-326.

Zhang, Y., Bu, T., Zhang, J., Tang, S., Yu, Z. et al. (2022). Decoding pixel-level image features from two-photon calcium signals of macaque visual cortex. *Neural Computation, 34,* 1369-1397.