Predicting Complex Cell Behavior Using Deep Predictive Models

Zhuokun Ding

Baylor College of Medicine

One Baylor Plaza, Houston, TX 77030

zhuokund@bcm.edu

Lingyan Wang

Baylor College of Medicine

One Baylor Plaza, Houston, TX 77030

lingyan.wang@bcm.edu

Abstract

Seeking to understand the neuronal structure and computation in visual system, previous studies have devoted effort in different approaches including building the link between deep neural networks (DNN) and neuronal response pattern in primate brain. It has been found that a well-trained DNN count demonstrated high-level similarity to primate visual system in structural and hierarchical representation of visual stimuli through its layers, with incredible accuracy in object recognition and classification tasks. Studies have also shown that DNN could directly predict large population neuronal response with high accuracy and fidelity. However, it is not clear whether DNN could simulate the underlying features of the neurons in the visual system, i.e., orientation or phase invariance etc., or merely mimic the response with computational mechanisms that's independent from visual system. In this article, we trained a convolutional neural network (CNN) with a well-defined complex cell model and tested its responses towards novel visual stimuli. As a result, our CNN could resemble the complex cell's response with good accuracy and fidelity. Moreover, the model showed similar orientation tuning and phase invariance with the complex cell model, demonstrating the capability of neural networks to infer the innate features of real neurons. This finding indicates the possibility of using neural networks to predict features of neurons and largely increase the efficiency of electrophysiology studies in the future.

Introduction

Human visual system perceives the external world with incredible complexity and precision. The incredible functionality of visual system lies on the performance of single neurons, i.e., tuning and invariances properties, and hierarchical circuit structures. Therefore, numerous studies (Alkon et at., 2012) have been devoted to investigate single neuron characteristics as well as circuit level structures, in the effort of revealing computational mechanisms underlying visual functions.

One of the research branches is to use DNN to understand the underlying principles of visual system (Guo et al., 2016). It has been found that hierarchical layers in DNN resemble the layers in visual system in their preference to different scales of visual features, i.e., lower layers are mostly activated by contrasts, edge, or colors and higher layers towards more comprehensive shapes and figures (Voulouldimos et al., 2018). Moreover, with the evolution of algorithms and hardwares, object classification/detection accuracy performed by DNN is increasingly higher and reached (or even outperform) human level performance (Rewat et al., 2017).

However, despite the high performance achieved, one remaining question is how to interpret the incredibly complicated structure of neural networks. If visual system is a black box to us, this puzzle remains unsolved and we've created a different **black box** with neural networks. DNN might contribute to advancement in the field of computation vision like self-driving cars, its contribution to our understanding of human visual system is limited. Therefore, there has been recent studies focusing on building the link between cortical visual system and neural networks by attempting to control the neuronal activities of the visual systems with the knowledge we gain from the DNN models . Using neural network to model the behavior of visual system, Bashivan and colleague (2019) was able to use it to construct images either broadly or selectively activate populations of neurons. In another group, Ponce and colleague (2019) used a generative deep neural network and a genetic algorithm to evolve images guided by neuronal firing and they were able to construct images that activated neurons more than natural images. Most recently, to find sensory stimuli that optimally drive neurons, Walker and colleague (2019) used a closed-loop experimental paradigm which combines neural recordings on large population of neurons to in silico nonlinear response modeling. The well-trained deep-learning –based model was able to accurately predict the optimal visual stimuli for thousands of neurons in the mouse primary visual cortex.

Yet there is more to learn about the visual systems besides the optimal stimuli. The maxima point (optimal visual stimuli) of a function (the visual systems) is not the perfect summary of the overall computation of the function. We would like to learn about the whole function and be able to tell how the visual system responds to different stimuli. One simple step would be to summarize how neurons respond to stimuli changing on one certain dimension (e.g. orientation, phase, translation, etc.), which has been studied as tuning properties or invariances of neurons for a long history in the field of in vivo neuroscience. With the extra accessibility of the DNN models compared with the biological brain, we could potentially guide the in vivo neuroscience experiments with the novel tuning properties/invariances captured by the DNN models and thus largely increase efficiency and broaden possibilities in experiments. While in order to study the tuning properties or invariances of the visual system in the DNN models, there is still one concern to address in advance: if the DNN could predict neural responses, does that mean it could also provide information on computational features of neurons, for example, orientation tuning or phase invariance? Towards this direction, Ukita and colleague (2019) have been searching for invariances in CNN models of visual cortex neurons. Previously, another group (Cadena et al., 2019) succeeded in partially predicting the receptive fields of simulated simple and complex cells with a simple CNN model.

In this article, our research question is whether a well-trained CNN model could not only predict the receptive fields but also capture both the **invariances** and **tuning** properties of the complex cell model. We implemented a simple CNN model similar to the structure of LeNet and trained it to predict a complex cell's model on ImageNet images. Out network could resemble the complex cell's response as well as orientation tuning and phase invariance.

Data

In order to train a CNN model to predict the response of a complex cell with known computation, we generated our own stimulus-response dataset. We first implemented a perfect complex cell model which can produce the response of a perfect complex cell to an image within its receptive field (described in detail in the Model section). To mimic actual visual neuroscience experiments with rodents, we recorded the responses of the complex cell model to 5000 natural images from the imagenet database (Fig 1.). 5000 is about the maximum number of images experimenters could pack into one record session in rodent visual experiments.

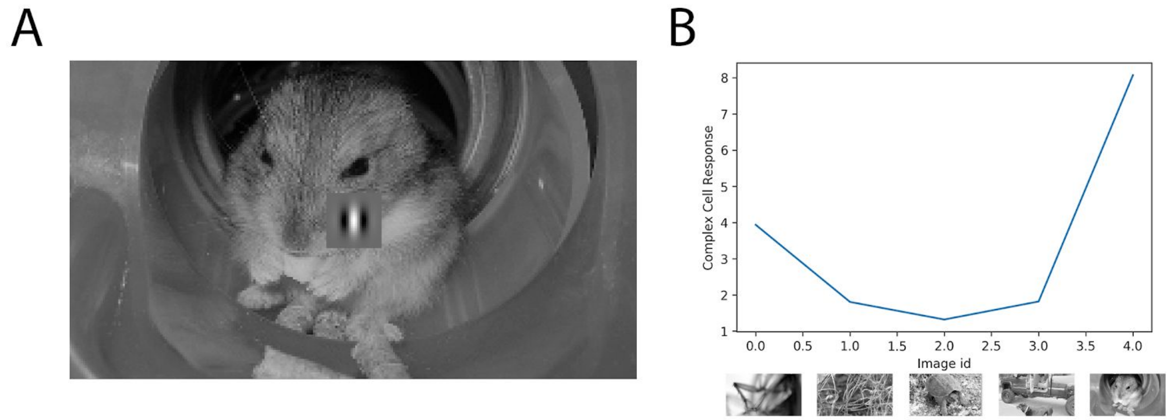

Figure 1. A). An example of natural images from ImageNet and the receptive field of the complex cell. B). Examples of complex neuron's response to different images

The 5000 natural images and the 5000 complex cell responses together make up our database. All natural images are preprocessed to be the same size (144 x 256) and the pixel amplitudes are normalized to be between 0 and 1. All responses are normalized to be between 0 and 1.

We see from the dataset that it holds a broad collection of visual patterns within the complex cell receptive field. The dataset provides a realistic approximation of the real neuronal recording data from one complex cell in mouse visual cortex during one experiment session. Our objectives are for our CNN model to: 1. predict the complex cell response to any natural image stimuli; 2. more importantly learn the orientation tuning and phase invariance implemented in the complex cell model. Since the second objective is not explicitly represented in the dataset and will not be part of the objective function, it could be difficult for the CNN models to learn.

Models

Complex Cell Model

We implemented the quadrature model similar to the motion-energy model (Gabbiani et al., 2017) as the model for a perfect complex cell. The response of the complex cell is obtained by squaring and summing the responses of two Gabor receptive fields 90 degrees out of phase, as defined in Eq. (1):

$$R_{cc} = R_{se}^2(t) + R_{so}^2(t) \qquad (1)$$

where,

$$R_{se}(t) = \int g_{se}(x)c(x,t)dx, \quad R_{so}(t) = \int g_{so}c(x,t)dx$$

and $c(x,t)$ is the stimulus, g_{se} and g_{so} is two Gabor filters with a 90-degree phase shift.

With such calculation, the model prefers gabor patches within the receptive fields at a specific orientation (orientation tuning) and sustains similar responses to gabor patches at different phases (phase invariance) (Figure 2). We aim to see if the orientation tuning and phase invariance existing in the complex cell model would be captured by the CNN model.

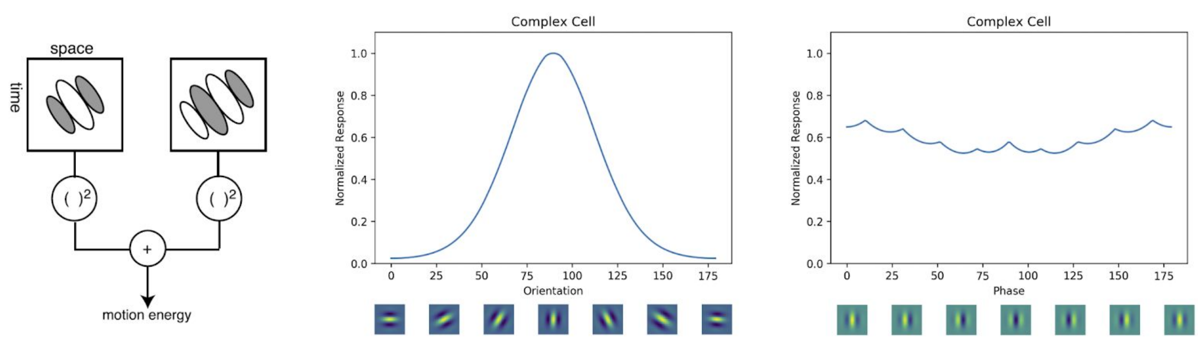

Figure 2. Complex cell model. Left: structure of complex cell computation. Middle: tuning property of complex cell to stimuli orientations. Right: tuning property of complex cell to stimuli phases

CNN model

We implemented a simple CNN model similar to the structure of LeNet (Figure 3). To adapt to the 144x256 input size, we increased the convolution kernel size to 9x9 for both the first and second convolutional layer. Since our data label is continuous values (complex cell responses) instead of the categorical labels LeNet originally designed for, we removed the softmax nonlinearity at the end of the net and replaced with a simple fully connected layer.

We chose a simple network structure since our task does not contain much nuisances and does not require disentanglement in a high dimensional space. We expect a simple neural network should have enough expression power to perform greatly on the task and we would like to avoid the lack of learnability of a complex net structure.

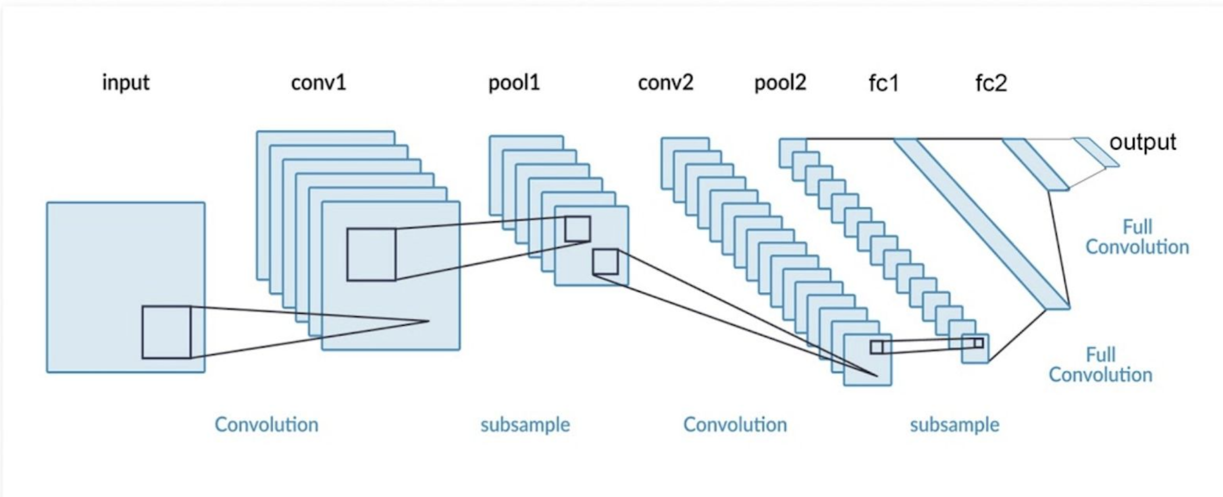

Figure 3. Structure of the convolutional neural network

Training

The dataset is split as a training set with 4900 stimulus-response pairs and a test set with 100 stimulus-response pairs. During training, we trained our model for 30 epoches. Each epoch contains 49 batches with a batchsize of 100 stimulus-response pairs. Since our task is similar to a regression, we chose the MSE loss function as the error function to optimize for. The model was trained using the SGD optimizer. During and after training, the model is evaluated by the correlation between the CNN responses and the complex cell response over 100 images in both training and testing set. After 30 epochs of training, the CNN model performed almost perfectly on the training set and reached about 0.5 correlation on the test set. The severe overfitting to the training set might be due to lack of regularization on the weight.

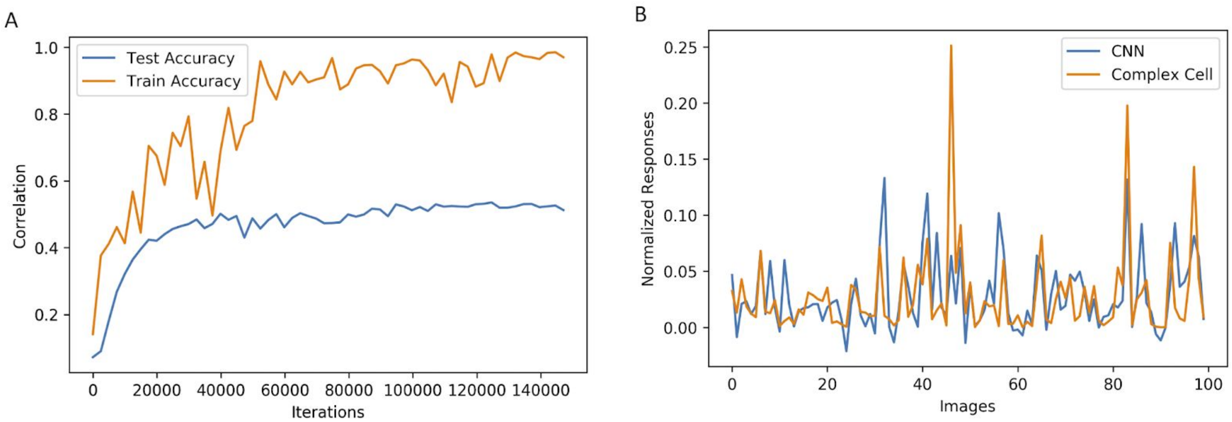

Figure 4. Model training. A: CNN training and testing accuracy with iteration number. B: CNN and complex cell's response to the test dataset.

Testing

To test if our model could capture the computation features (orientation tuning and phase invariance) of a complex cell, we generated Gabor patches as the same size of the receptive field of our complex cell model with different orientations and phases. The Gabor patches are shown to both the complex cell model and CNN at the location of the complex cell model receptive field. Surprisingly, even though our CNN model did not perform perfectly on the training set and it has never been trained on any gabor-like stimuli, it almost perfectly captured the orientation tuning of the complex cell model (Figure 5A). It also partially captured the phase invariance (Figure 5B.)

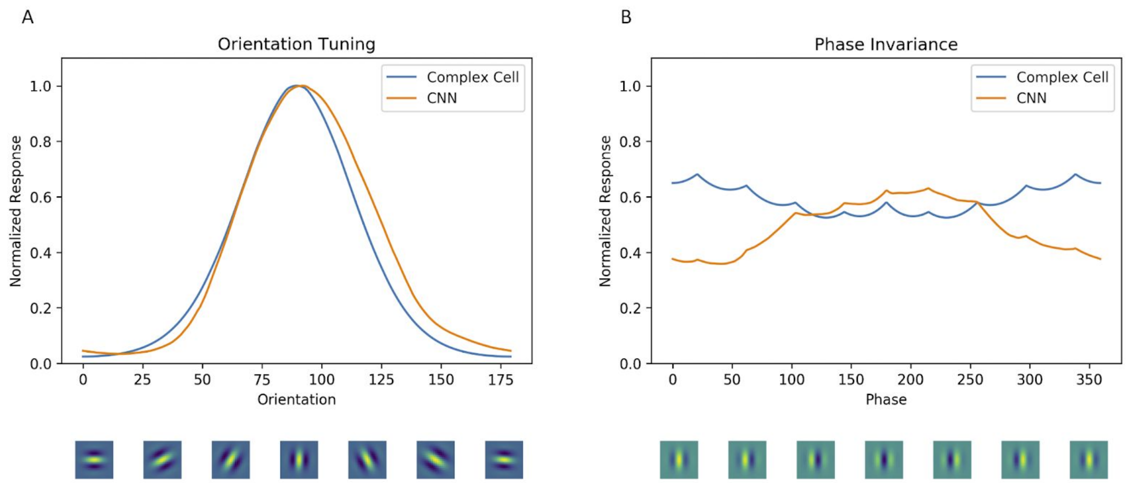

Figure 5. A: comparison between CNN and complex cell on tuning properties to stimuli orientations. B: comparison between CNN and complex cell on tuning properties to stimuli phase

Conclusion and Discussion

Using a simple CNN structure, our neural network could predict a complex cell model's response to novel visual images with an accuracy of around 0.5. Moreover, our CNN model almost perfectly captures the complex cell model's orientation tuning property and partially resembles the phase invariance.

As the neural network is innocent to tuning and invariance property before training, our results indicate that a well-trained neural network could provide or extract new information on inner structures of a neuron from its response patterns. Potentially, this could benefit the field of vision science in two ways. First, the inner structure of neural network could be used in future studies to infer the computational mechanisms of single neurons. Second, neural networks could be used to generate hypotheses, i.e., predict neurons' new properties, and thus largely increase the efficiency of in vivo electrophysiology research.

It remains unknown how the CNN model achieved the phase invariance and orientation tuning. Feature visualization techniques could be applied to the CNN model to further probe into the computation happening in the CNN. For example, we could ask how the invariance and tuning changes at different layers of the

CNN. Another challenging direction is to develop a framework to identify unknown invariance captured by the CNN model without any biased assumption.

References

Alkon, D. L. (2012). *Neural principles in vision*. Springer Science & Business Media.

Bashivan, P., Kar, K., & DiCarlo, J. J. (2019). Neural population control via deep image synthesis. Science, 364(6439), eaav9436.

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., & Ecker, A. S. (2019). Deep convolutional models improve predictions of macaque V1 responses to natural images. PLoS computational biology, 15(4), e1006897.

Carandini, M. (2006). What simple and complex cells compute. Journal of Physiology, 577(2), 463–466. https://doi.org/10.1113/jphysiol.2006.118976

Gabbiani, F., & Cox, S. J. (2017). Models of Simple and Complex Cells. In Mathematics for Neuroscientists (pp. 395–410). Elsevier. https://doi.org/10.1016/b978-0-12-801895-8.00023-3

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. Neurocomputing, 187, 27-48.

Ponce, C. R., Xiao, W., Schade, P. F., Hartmann, T. S., Kreiman, G., & Livingstone, M. S. (2019). Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. Cell, 177(4), 999-1009.

Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." Neural computation 29.9 (2017): 2352-2449.

Sinz, F. H., Ecker, A. S., Fahey, P. G., Walker, E. Y., Cobos, E., Froudarakis, E., ... Tolias, A. S. (2018). Stimulus domain transfer in recurrent models for large scale cortical population prediction on video. In Advances in Neural Information Processing Systems (Vol. 2018-Decem, pp. 7199–7210).

Ukita, J., Yoshida, T., & Ohki, K. (2019). Characterisation of nonlinear receptive fields of visual neurons by convolutional neural network. *Scientific reports*, *9*(1), 3791.

Walker, E. Y., Sinz, F. H., Froudarakis, E., Fahey, P. G., Muhammad, T., Ecker, A. S., ... Tolias, A. S. (2018). Inception in visual cortex: in vivo-silico loops reveal most exciting images. BioRxiv, 506956. https://doi.org/10.1101/506956

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018.