Bidirectional Predictive Coding

Gaspard Oliviers (gaspard.oliviers@bndu.ox.ac.uk)
MRC Brain Network Dynamics Unit
University of Oxford, UK

Mufeng Tang (mufeng.tang@bndu.ox.ac.uk)
MRC Brain Network Dynamics Unit
University of Oxford, UK

Rafal Bogacz (rafal.bogacz@bndu.ox.ac.uk)
MRC Brain Network Dynamics Unit
University of Oxford, UK

Abstract

Predictive coding networks (PCNs) offer a compelling model for cortical sensory processing, yet their traditional unidirectional formulations—focusing solely on top-down or bottom-up processing—limit their applicability across various tasks and fail to capture the sensory system's bidirectional nature. We introduce a bidirectional predictive coding model, leveraging a single neural network for both top-down and bottom-up processing. This approach not only maintains the biological plausibility of PCNs but also excels in generative and discriminative tasks simultaneously within a single training session. Our model presents a unified, mechanistic view of how the sensory system employs a bidirectional approach to achieve its functionalities.

Keywords: Predictive coding; Sensory processing; Discriminative and generative tasks

Introduction

Predictive coding (PC) is an influential computational framework for understanding cortical functions. Traditionally, it assumes that perceptual processing is a **top-down** process where higher levels of the brain predict sensory inputs and correct prediction errors by learning and inferring iteratively in a Bayesian manner (Rao & Ballard, 1999; Friston, 2005) (Fig. 1a). While top-down PC networks (PCNs) are effective in generative tasks like memory and image generation (Salvatori et al., 2021; Oliviers, Bogacz, & Meulemans, 2024), PC can also be formulated **bottom-up**, processing inputs from lower to higher levels of the cortex and excelling in discriminative tasks such as image classification (Whittington & Bogacz, 2017; Song et al., 2024) (Fig. 1b). However, these unidirectional PCNs struggle in cross-training scenarios (i.e., top-down PCNs for discriminative tasks, and vice versa) (Tscshantz, Millidge, Seth, & Buckley, 2023; Sun & Orchard, 2020) and fail to mirror the sensory system's bidirectional architecture and functional versatility (Lamme & Roelfsema, 2000; Siegel, Körding, & König, 2000; McMains & Kastner, 2011; Lange, Shivkumar, Chattoraj, & Haefner, 2023).

Our study introduces the **bidirectional predictive coding network** (bPCN), overcoming the limitations of unidirectional PCNs by training both top-down and bottom-up processes within a single network (Fig. 1c). This approach not only enhances performance across generative and discriminative tasks but also inherits classical PCNs' plausible local computation and Hebbian plasticity. Through bPCN, we offer a unified model of cortical functions, reflecting the sensory system's inherent bidirectionality.

Models

Classical PCNs minimize the sum of squared layer-wise errors, known as the energy function. However, the formulations of the energy functions are slightly different between the top-**d**own PCN (**d**PCN) (Rao & Ballard, 1999) and the bottom-**u**p PCN (**u**PCN) (Whittington & Bogacz, 2017). For an L-layer

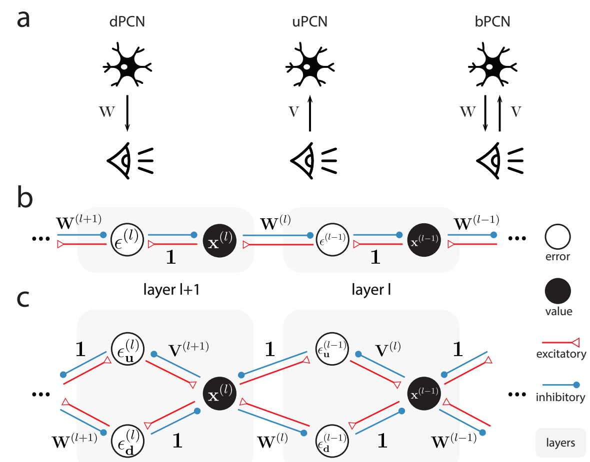

Figure 1: **a**. Directions of processing in dPCN, uPCN and bPCN. **b** & **c**: Neural implementation of dPCN and bPCN.

PCN, with a sensory layer $x^{(1)}$ and a topmost hidden layer $x^{(L)}$, the energy functions for dPCN and a uPCN are given by:

$$E_{\mathbf{d}} = \sum_{l=1}^{L-1} \frac{1}{2} \| x^{(l)} - \mathbf{W}^{(l+1)} f(x^{(l+1)}) \|_2^2 \tag{1}$$

$$E_{\mathbf{u}} = \sum_{l=1}^{L-1} \frac{1}{2} \| x^{(l+1)} - \mathbf{V}^{(l+1)} f(x^{(l)}) \|_2^2, \tag{2}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{V}^{(l)}$ are the top-down and bottom-up weights respectively, and f is a nonlinear function. It is known that dPCNs perform image generation well (Salvatori et al., 2021; Oliviers et al., 2024) and uPCNs excel in image classification (Whittington & Bogacz, 2017; Song et al., 2024). However, their performance degrades when cross-trained: when dPCNs are trained in classification tasks, i.e. when fixing $x^{(1)}$ to an image and running iterative inference on $x^{(L)}$ to obtain the correct labels, dPCNs fail to predict the labels with an accuracy similar to uPCNs (Tscshantz et al., 2023). Likewise, when uPCNs are tested by fixing $x^{(L)}$ to the labels and running iterative inference on $x^{(1)}$, the bottom layer struggles to converge to clear average images corresponding to the class labels (Sun & Orchard, 2020). Our bPCN strikes a balance between bottom-up and top-down processing, employing an energy function as follows:

$$E = E_{\mathbf{u}}/\sigma_{\mathbf{u}} + E_{\mathbf{d}}/\sigma_{\mathbf{d}}, \tag{3}$$

where $\sigma_{\mathbf{u}}$ and $\sigma_{\mathbf{d}}$ determine the weights given to the bottom-up and top-down energy components. The inference dynamics of a certain layer l can be derived following gradient descent on E as:

$$\begin{aligned} \dot{x}^{(l)} \propto & -\varepsilon_{\mathbf{d}}^{(l)} - \varepsilon_{\mathbf{u}}^{(l-1)} \\ & + f'(x^{(l)}) \odot \left((\mathbf{W}^{(l)})^\top \varepsilon_{\mathbf{d}}^{(l-1)} + (\mathbf{V}^{(l+1)})^\top \varepsilon_{\mathbf{u}}^{(l)} \right) \end{aligned} \tag{4}$$

where we define $\varepsilon_{\mathbf{d}}^{(l)} := (x^{(l)} - \mathbf{W}^{(l+1)} x^{(l+1)})/\sigma_{\mathbf{d}}$ and $\varepsilon_{\mathbf{u}}^{(l)} := (x^{(l+1)} - \mathbf{V}^{(l+1)} x^{(l)})/\sigma_{\mathbf{u}}$ i.e., the layer-wise down and up prediction errors, and f' the derivative of f, $\odot$ the element-wise

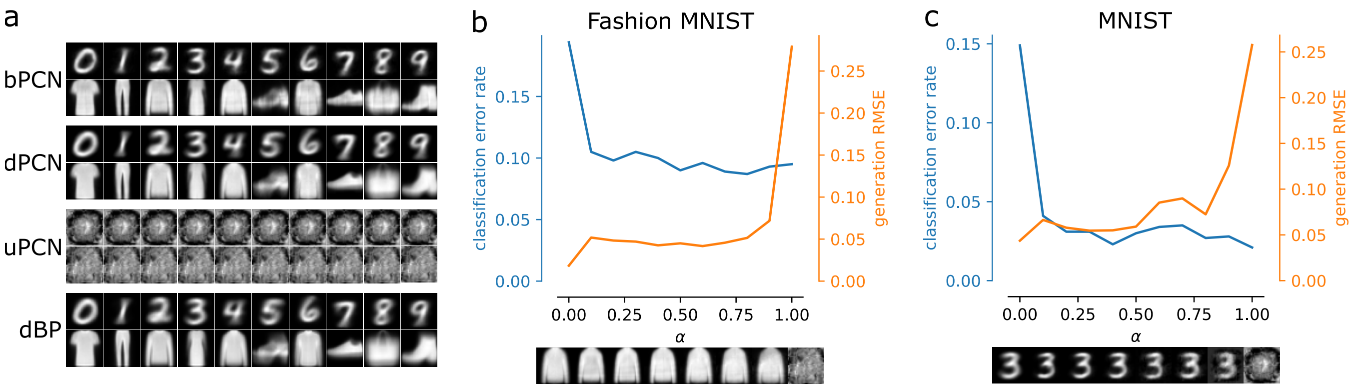

Figure 2: **Learning performance of bidirectional predictive coding network on MNIST and Fashion-MNIST**. **a**. Comparison of generated (iteratively inferred) images at $x^{(1)}$ when the $x^{(L)}$ is fixed at class labels. **b** & **c**. Classification and generation error of a bPCN as the balance between bottom-up and top-down streams is adjusted. The parameter α scales the relative weight of the streams with $1/\sigma_{\mathbf{d}} \propto (1-\alpha)/\sigma_{\mathbf{d,opt}}$ and $1/\sigma_{\mathbf{u}} \propto \alpha/\sigma_{\mathbf{u,opt}}$.

product. These computations can be implemented in a neural circuit shown in Fig. 1c, where each value neuron $x^{(l)}$ is connected to two error neurons, as opposed to classical PCNs where there is a one-to-one mapping between value and error neurons. Importantly, like in classical PCNs, the learning rules in our bPCN is Hebbian by gradient descent on E:

$$\Delta \mathbf{W}^{(l)} \propto \varepsilon_{\mathbf{d}}^{(l-1)} f(x^{(l)})^{\top}; \quad \Delta \mathbf{V}^{(l)} \propto \varepsilon_{\mathbf{u}}^{(l-1)} f(x^{(l-1)})^{\top}. \tag{5}$$

It is worth noting that our bPCN is not equivalent to (trivially) training a dPCN and a uPCN separately: the two streams share the same network of neurons, making bPCN a more efficient and integrated bidirectional model.

Results

We examine our bPCN in performing image generation and classification simultaneously. During training, $x^{(1)}$ is fixed to images and $x^{(L)}$ is fixed to the one-hot representation of labels. The model then performs 'constrained' inference (Eq. 4) and learning dynamics (Eq. 5), in a similar way to the uPCN in Whittington and Bogacz (2017). During image generation, we fix $x^{(L)}$ to the one-hot labels and perform iterative inference on $x^{(1)}$ (Eq. 4) to obtain the predicted images, which we then compare with the average image of each class; during classification, we fix $x^{(1)}$ to images and perform inference $\dot{x}^{(L)}$ to obtain the predicted class labels. The experiments are performed on both MNIST (LeCun, Cortes, & Burges, 2010) and fashion-MNIST (Xiao, Rasul, & Vollgraf, 2017) and we used a PCN with 2 hidden layers of 256 neurons for both datasets. We performed a grid search over possible combinations of $\sigma_{\mathbf{u}}$ and $\sigma_{\mathbf{d}}$ and other hyperparameters to obtain the (combined) highest classification accuracy and the lowest RMSE to the class average images. Our best-performing bPCN is shown in Table 1, where we compare it with fine-tuned dPCN and uPCN, as well as BP-trained models (uBP and dBP) for classification and class-conditional generation. It can be seen that bPCN excels in both classification and generation tasks, on par with its specialized counterparts, whereas dPCN and uPCN struggle when cross-trained. The performance is also comparable to models trained by BP in each task, where predicted labels and generated images are obtained by a forward pass bottom-up and top-down respectively, although there is no equivalent way of simultaneously training BP-based models on both tasks. In Fig. 2a we present visual demonstration of the class-conditioned generation.

Table 1: Comparison of **test** classification accuracy and conditional generation root mean squared error on the MNIST dataset.

	Accuracy (%)		RMSE	
Model	MNIST	F-MNIST	MNIST	F-MNIST
bPCN	**97.0**	**91.0**	**0.059**	**0.045**
uPCN	**97.9**	**90.5**	0.257	0.279
dPCN	85.1	80.6	**0.044**	**0.018**
uBP	97.1	89.9	-	-
dBP	-	-	0.048	0.013

In Fig. 2b and c we study the computational properties of the bPCN model. We vary a parameter α to control the relative weights of the top-down and bottom-up streams: $\alpha = 0$ and 1 correspond to models with only the top-down and bottom-up streams respectively, and $\alpha = 0.5$ corresponds to the optimal $\sigma_{\mathbf{u,opt}}$ and $\sigma_{\mathbf{d,opt}}$ from the grid search. In both datasets, the bPCN model has a high tolerance for different weightings of the streams, maintaining a high classification accuracy and image generation quality even when α is close to but not exactly 0 and 1.

Conclusion

In this work, we introduced bidirectional predictive coding networks, a model of top-down **and** bottom-up sensory processing in the cortex. Our bPCN model inherits the biologically plausible circuit implementation of classical unidirectional PCNs, and performs discriminative and generative tasks comparable to their specialized counterparts. Our model provides a mechanistic account of bidirectional processing in the sensory system underlying its various functions.

Acknowledgments

This work has been supported by MRC grant MC_UU_00003/1. MT is supported by the E.P. Abraham Scholarship in Chemical, Biological / Life and Medical Sciences.

References

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, *360*(1456), 815–836.

Lamme, V. A., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in neurosciences*, *23*(11), 571–579.

Lange, R. D., Shivkumar, S., Chattoraj, A., & Haefner, R. M. (2023). Bayesian encoding and decoding as distinct perspectives on neural coding. *Nature Neuroscience*, *26*(12), 2063–2072.

LeCun, Y., Cortes, C., & Burges, C. J. (2010). Mnist handwritten digit database. 2010. *URL http://yann. lecun. com/exdb/mnist*, *7*(23), 6.

McMains, S., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience*, *31*(2), 587–597.

Oliviers, G., Bogacz, R., & Meulemans, A. (2024). Learning probability distributions of sensory inputs with monte carlo predictive coding. *bioRxiv*, 2024–02.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, *2*(1), 79–87.

Salvatori, T., Song, Y., Hong, Y., Sha, L., Frieder, S., Xu, Z., ... Lukasiewicz, T. (2021). Associative memories via predictive coding. *Advances in Neural Information Processing Systems*, *34*.

Siegel, M., Körding, K. P., & König, P. (2000). Integrating top-down and bottom-up sensory processing by somato-dendritic interactions. *Journal of computational neuroscience*, *8*, 161–173.

Song, Y., Millidge, B., Salvatori, T., Lukasiewicz, T., Xu, Z., & Bogacz, R. (2024). Inferring neural activity before plasticity as a foundation for learning beyond backpropagation. *Nature Neuroscience*, 1–11.

Sun, W., & Orchard, J. (2020). A predictive-coding network that is both discriminative and generative. *Neural computation*, *32*(10), 1836–1862.

Tscshantz, A., Millidge, B., Seth, A. K., & Buckley, C. L. (2023). Hybrid predictive coding: Inferring, fast and slow. *PLoS Computational Biology*, *19*(8), e1011280.

Whittington, J. C., & Bogacz, R. (2017). An approximation of the error backpropagation algorithm in a predictive coding network with local hebbian synaptic plasticity. *Neural computation*, *29*(5), 1229–1262.

Xiao, H., Rasul, K., & Vollgraf, R. (2017). *Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms.*