Top-down Feedback Connections for Deep Neural Networks

Lab Rotation Report

by Johannes Rieke

(johannes.rieke@gmail.com)

30 March 2019

Humboldt University of Berlin

Bernstein Center for Computational Neuroscience Berlin

Supervisor: Prof. Dr. Matthew Larkum

Abstract

Top-down feedback connections are abundant in the neocortex and play a role in many behavioral tasks. Here, we integrate such feedback connections into deep neural networks. We test our feedback model on two tasks: First, we classify images of handwritten letters. We show that the feedback connections slightly improves the performance of our network compared to a feedforward net, apparently by adding more depth to the computation. Second, we design a sequential image classification task where we arrange images of handwritten letters in words. The goal is to classify each image in the sequence correctly. In this task, we train the feedback connections to transmit context information about the previously seen letters. We show that the model successfully learns to make use of this context information to improve its classification performance, especially on noisy images. Finally, we propose some other tasks (inspired by biological experiments) that our feedback net could be tested on.

1. Introduction

Recent progress in artificial intelligence has been driven mostly by deep neural networks (LeCun et al. 2015). These models are loosely based on the computational features of biological neurons. However, artificial neural networks lack one important aspect of signal processing in the brain: top-down feedback connections. These connections transmit information from hierarchically higher areas in the brain to hierarchically lower areas. They can be found everywhere in the neocortex and reach as far down as the lowest sensory areas (Callaway et al. 2004). Here, we aim to implement such feedback connections into deep neural networks.

While the concrete role of top-down feedback in the brain is debated, it was shown to play a role in various effects (Gilbert & Li 2013), such as contextual modulation of the visual field, task-dependent behavior, and different forms of attention. Recently, it was found that top-down feedback also plays a critical role in stimulus perception (Manita et al. 2015). While these effects are modulatory (i.e. feedback influences the feedforward computation), it was also proposed that feedback connections carry an error signal that is used in lower layers for credit assignment and learning (Whittington & Bogacz 2019, Richards & Lillicrap 2018).

In our work, we use neural networks with feedback connections in two settings: In the first experiment, we train a network to recognize images of handwritten letters. We find that networks with top-down feedback outperform those with feedforward connections, potentially by adding more depth to the network. In the second experiment, we train the network on a sequential image classification task where the feedback connections learn to provide contextual information about the sequence.

2. Model

We first take a fully-connected neural network with three layers and ReLU non-linearities. The output of layer l (with weight $W_{FF}^{(l)}$ and bias $b_{FF}^{(l)}$) is, therefore:

$$y^{(l)} = \sigma\left(x^{(l)} W_{FF}^{(l)\top} + b_{FF}^{(l)}\right)$$

To this feedforward network, we add top-down feedback connections. A feedback connection in our model is a standard linear layer (with weight $W_{FB}^{(l)}$ and bias $b_{FB}^{(l)}$), which takes the output of layer $l+1$ and computes the feedback input to layer l as:

$$x_{FB}^{(l,t)} = y^{(l+1,t-1)} W_{FB}^{(l)\top} + b_{FB}^{(l)}$$

This feedback is added to the feedforward input for layer l, so that the feedforward pass becomes:

$$y^{(l,t)} = \sigma\left(((1-\alpha)x^{(l,t)} + \alpha x_{FB}^{(l,t)})W_{FF}^{(l)\top} + b_{FF}^{(l)}\right)$$

α is a hyperparameter of the layer that controls the amount of feedback. On the output layer, we apply dropout (0.5) and the softmax activation. We do not add feedback connections from the output layer to the layer below, as this seemed to inhibit training in experiments.

Introducing the recurrent feedback connections adds a time dimension to the computation. Either, one can do multiple passes through the network with the same input (see experiment 1), or run the network on a sequence of samples, so that the feedback activation of one sample serves as a context signal for the next sample (see experiment 2).

3. Experiments

3.1 Experiment 1: Image Classification

Setup: In this experiment, we perform standard image classification (i.e. finding the class for a single image) with the EMNIST dataset (“letters” split; Cohen et al. 2017). This dataset is similar to the popular MNIST dataset but contains letters instead of digits. We use this dataset because it is more challenging than MNIST and makes experiment 2 more intuitive. We train the 3-layer feedback network (FB) and an equivalent feedforward network (FF; same architecture but no feedback connections) with stochastic gradient descent with momentum. We split 10 k samples from EMNIST’s training set for validation and perform a grid search for hyperparameters for 30 epochs (details given in table 1). For each image, we run a feedforward pass through the network, then a feedback pass and then another feedforward pass (which integrates the feedback). The output of this second feedforward pass is the classification.

Results: Table 1 shows the hyperparameter configurations and results for each network. The feedback net (FB) performs a bit better than the feedforward (FF) net on normal images (accuracy of 92.2 % vs. 91.2 %). This difference is larger when we test on images with added Gaussian noise (mean 0, standard deviation 64, scaling factor 2). Here, the FB net wins by a considerable margin (51.2 % vs. 39.7 %). Apparently, it is able to extract more meaningful features. One could argue that the FB net is simply better because it has more parameters. To test for this, we train a feedforward net with twice as many hidden neurons (1000 vs. 500), which has approximately the same number of parameters as the FB net. However, as we see in table 1, this network performs similarly as the original FF net.

Discussion: We see that the feedforward connections allow the network to extract more meaningful features, which boost its accuracy on image classification. This result is in alignment with two recent papers: Spoerer et al. (2017) use feedback connections in convolutional neural networks (CNNs). They show superior performance on recognition of partially occluded images compared to feedforward nets. Nayebi et al. (2018) constructed novel CNN architectures with lateral and feedback connections, which can outperform deep

Table 1: Hyperparameters and results for single image classification (experiment 1). Hyperparameters were chosen via grid search across the parameter values in the first row.

	Number of parameters	Learning rate (0.005, 0.01, 0.02, 0.04)	Momentum (0.3, 0.5, 0.7)	Batch size (64, 128)	Alpha (0.3, 0.5, 0.7)	Accuracy	Accuracy with noise
FF net	405 526	0.02	0.7	128	-	91.2 %	39.7 %
FF net (large)	811 026	0.04	0.5	128	-	91.8 %	40.2 %
FB net	798 310	0.02	0.3	64	0.3	92.2 %	51.2 %

ResNets while requiring fewer layers and parameters. As was argued in the latter paper, top-down feedback seems to add more depth to the network. This might be critical for the visual system in the brain as it consists of a small number of areas (which can be seen as layers in the neural network view). Top-down feedback connections might help the visual system to reach the same performance as deep neural networks, which usually consist of hundreds of layers.

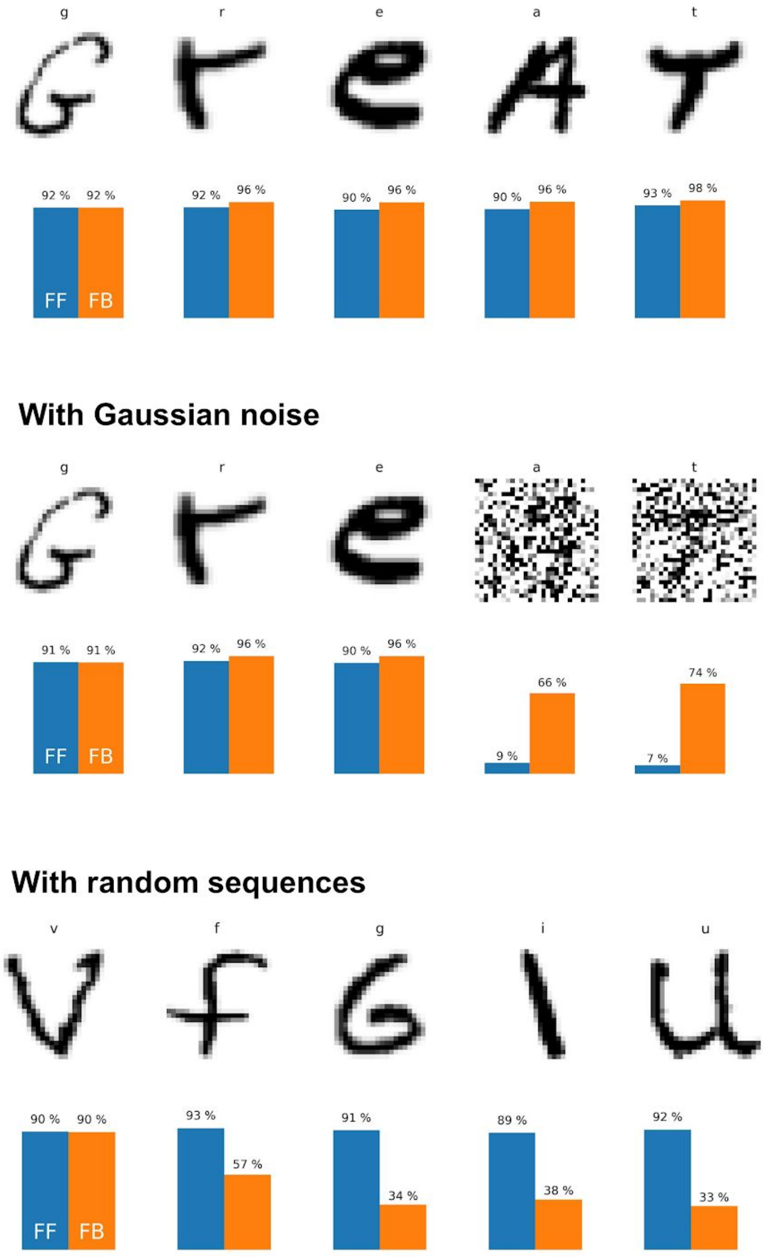
3.2 Experiment 2: Sequential Image Classification

Setup: In the second experiment, we use our feedback model for sequential image classification. We construct sequences of EMNIST images based on 50 common English words with five letters[1]. Fig. 1 shows an example sequence. The goal for our network is to classify each image in the sequence correctly. For this, we run the first image through the network and classify it, run a feedback pass, run the next image through the network and classify it, etc. This way, we hope that the feedback connections learn to transmit a context signal about the previous image(s), which can help the network with the classification task. The feedforward and feedback connections of the model are trained separately: First, we train the feedforward connections exactly as in experiment 1 (standard image classification with EMNIST, hyperparameters as FF net in table 1). Then, we freeze the feedforward weights and train the feedback weights on the image sequences with stochastic gradient descent (learning rate 0.00001, momentum 0.5). We set alpha to 0.5 and apply dropout of 0.5 before the output layer.

Results: Fig. 1 shows classification accuracies for different positions in the sequence (averaged over the test set). As in experiment 1, we compare the results of the feedback net (FB) against an equivalent feedforward net (FF). On normal sequences (fig. 1, top), the FF net achieves roughly the same accuracy for each position, while the FB net improves during the sequence. Apparently, the FB net successfully learned to make use of the sequential information through its feedback connections. Next, we add Gaussian noise (mean 0, std 64, scaling factor 4) to the last two images (fig. 1, middle). While the FF net fails to recognize the images, the FB net makes use of the sequential information and can recognize most of the

[1] Scraped from http://www.thefreedictionary.com/5-letter-words.htm

noisy images. However, we want to note that this works only on the sequences that the FB net learned. When we test the network on random sequences (fig. 1, bottom), the FB net cannot make sense of these unknown sequences and fails to recognize most images.



With Gaussian noise

With random sequences

Figure 1: Classification accuracy per sequence position (experiment 2) for normal (top), noisy (middle) and random (bottom) sequences. FF = feedforward net, FB = feedback net. Image sequences are individual samples from the test set.

Discussion: In this experiment, we show that top-down feedback connections can successfully learn to provide context information for the feedforward computation. Obviously, our results could well be achieved with “standard” recurrent neural networks (e.g. LSTM), which are often used for sequential tasks. However, these networks need to be trained from scratch on sequential data, while our feedback connections are only modulatory. Also, the kind of parameterized feedback connections we use here are biologically more plausible, as they resemble the hierarchical architecture of connections in the cortex. It would be interesting to see how feedback connections in the brain behave during sequential tasks and if preventing top-down feedback impairs performance on such tasks.

4. Conclusion

In this work, we explored the use of top-down feedback connections in deep neural networks. We showed that feedback connections can improve standard image classification as well as provide a context signal during sequential classification. Our experiments show 1) how top-down feedback connections might be used for processing in the neocortex and 2) how such connections can be used to explore new use cases in machine learning. While we tested only on simple multi-layer perceptrons with one hidden layer, our architecture can be easily scaled to deeper networks, potentially with convolutions and feedback connections across multiple layers (similar to Nayebi et al. 2018).

Given that feedback seems to play a vital role for many functions in the brain (Gilbert & Li 2013), our model could be tested on a variety of tasks. We briefly outline two ideas:

- Multi-task learning: In the visual system, the response of low-level neurons changes depending on the task that is being performed, which is mediated by top-down feedback (Gilbert & Li 2013, Bagur et al. 2018). Taking inspiration from this, train a network with feedback connections on two different tasks. The feedback connections can influence the feedforward computation based on the task being performed.

- Multi-modal learning: Sensory systems in the brain (e.g. visual, auditory) influence each other through feedback connections (Petro et al. 2017, Ghazanfar & Schroeder 2006). To classify multi-modal input (e.g. video with sound), one could think of feedback connections between different parts of the network, so that the processing of the video can influence the processing of the sound on a low level.

Finally, we want to raise some general comments on this study and the aim of bringing together neuroscience and machine learning. In both fields, we face vastly different intuitions: Machine learning systems are trained on huge datasets and specific tasks completely from scratch. In neuroscience, experiments involve few samples and are (necessarily) influenced by existing knowledge in the brain. In our eyes, these differences make it difficult to compare machine learning and neuroscience, or transfer knowledge between both areas. To overcome these difficulties, it seems important to 1) focus on neuroscience experiments that test very specific functions of the brain, so that their results can be directly compared to the capabilities of machine learning systems and 2) spend more time using transfer and multi-modal learning in machine learning systems, so that they become more capable of modeling the highly complex and diverse nature of the neocortex.

References

Bagur, S., Averseng, M., Elgueda, D., David, S., Shamma, S., Boubenec, Y., … Ostojic, S. (2018). Go/No-Go task engagement enhances population representation of target stimuli in primary auditory cortex. Nature Communications, 9(2529). https://doi.org/10.1038/s41467-018-04839-9

Callaway, E. M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. Neural Networks, 17(5–6), 625–632. https://doi.org/10.1016/j.neunet.2004.04.004

Cohen, G., Afshar, S., Tapson, J., & Van Schaik, A. (2017). EMNIST: Extending MNIST to handwritten letters. Proceedings of the International Joint Conference on Neural Networks. https://doi.org/10.1109/IJCNN.2017.7966217

Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? Trends in Cognitive Sciences, 10(6), 278–285. https://doi.org/10.1016/j.tics.2006.04.008

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. Nature Reviews Neuroscience, 14(5), 350–363. https://doi.org/10.1038/nrn3476

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Manita, S., Suzuki, T., Homma, C., Matsumoto, T., Odagawa, M., Yamada, K., … Murayama, M. (2015). A Top-Down Cortical Circuit for Accurate Sensory Article A Top-Down Cortical Circuit for Accurate Sensory Perception. Neuron, 86(5), 1304–1316. https://doi.org/10.1016/j.neuron.2015.05.006

Nayebi, A., Bear, D., Kubilius, J., Kar, K., Ganguli, S., Sussillo, D., … Yamins, D. L. K. (2018). Task-Driven Convolutional Recurrent Models of the Visual System, (NeurIPS). https://doi.org/arXiv:1807.00053v2

Petro, L. S., Paton, A. T., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. Philosophical Transactions of the Royal Society B: Biological Sciences, 372(20160104). https://doi.org/10.1016/j.cub.2014.04.020

Richards, B. A., & Lillicrap, T. P. (2018). Dendritic solutions to the credit assignment problem. Current Opinion in Neurobiology, 54(September), 28–36. https://doi.org/10.1016/j.conb.2018.08.003

Spoerer, C. J., McClure, P., & Kriegeskorte, N. (2017). Recurrent convolutional neural networks: A better model of biological object recognition. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.01551

Takahashi, N., Oertner, T. G., Hegemann, P., & Larkum, M. E. (2016). Active cortical dendrites modulate perception. Science, 354(6319), 1159–1165. https://doi.org/10.1126/science.aah6066

Whittington, J. C. R., & Bogacz, R. (2019). Theories of Error Back-Propagation in the Brain. Trends in Cognitive Sciences. https://doi.org/10.1016/j.tics.2018.12.005