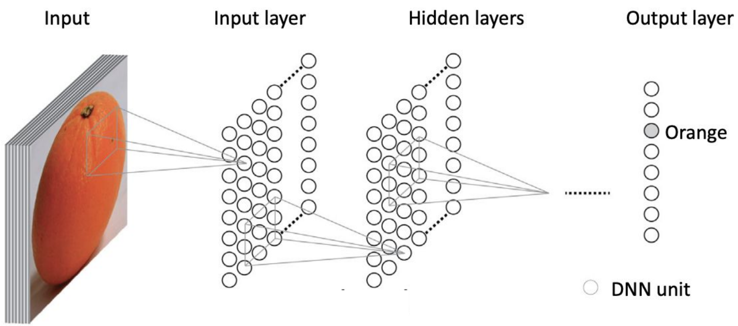
Investigating Deep Neural Networks (DNNs) Through Learning Rule-Specific Representational Profiles

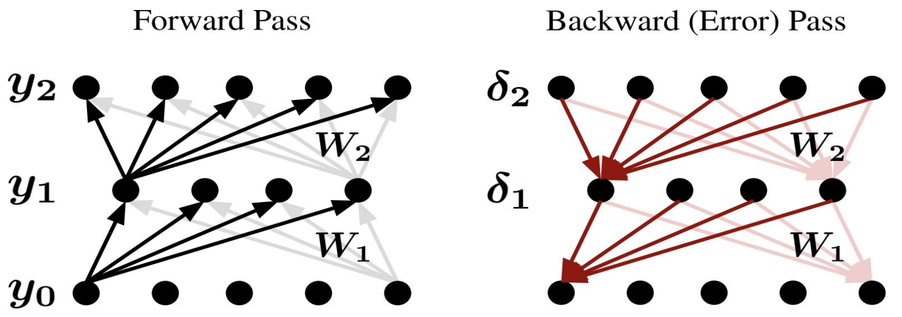
By: Bati Yilmaz, Omid Amiratashani, Chengcheng Du, Ghanendra Singh
Pod name: Appreciative Nemesia
Group name: µRepresent

neuromatch academy

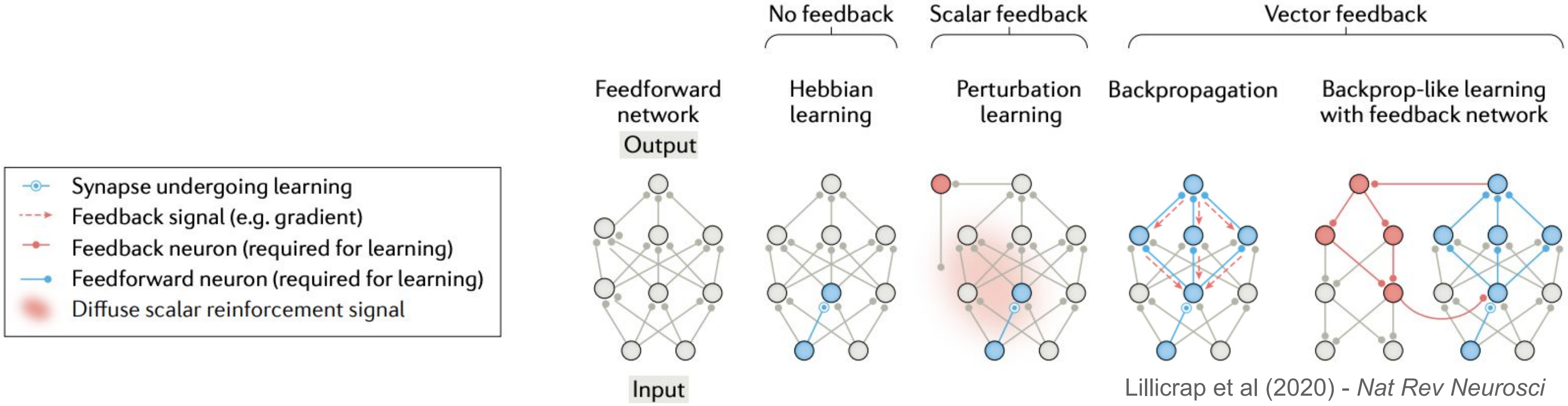
Introduction



Cichy and Kaiser (2019) - *Trends Cogn Sci*



Ahmad et al (2020) - *arXiv*



Lillicrap et al (2020) - *Nat Rev Neurosci*

Main Question

How and to what extent
DNNs having the same architecture, trained on the same dataset but using different learning rules exhibit unique representational profiles?

Side Quest

Specifically, **the comparison of an artificial and a bio-plausible learning rule: Backpropagation (BP) and Feedback Alignment (FA)**, respectively.

Methodology

Utilized Pytorch and BioTorch (Sanfiz and Akrout, 2021) library to implement FA.

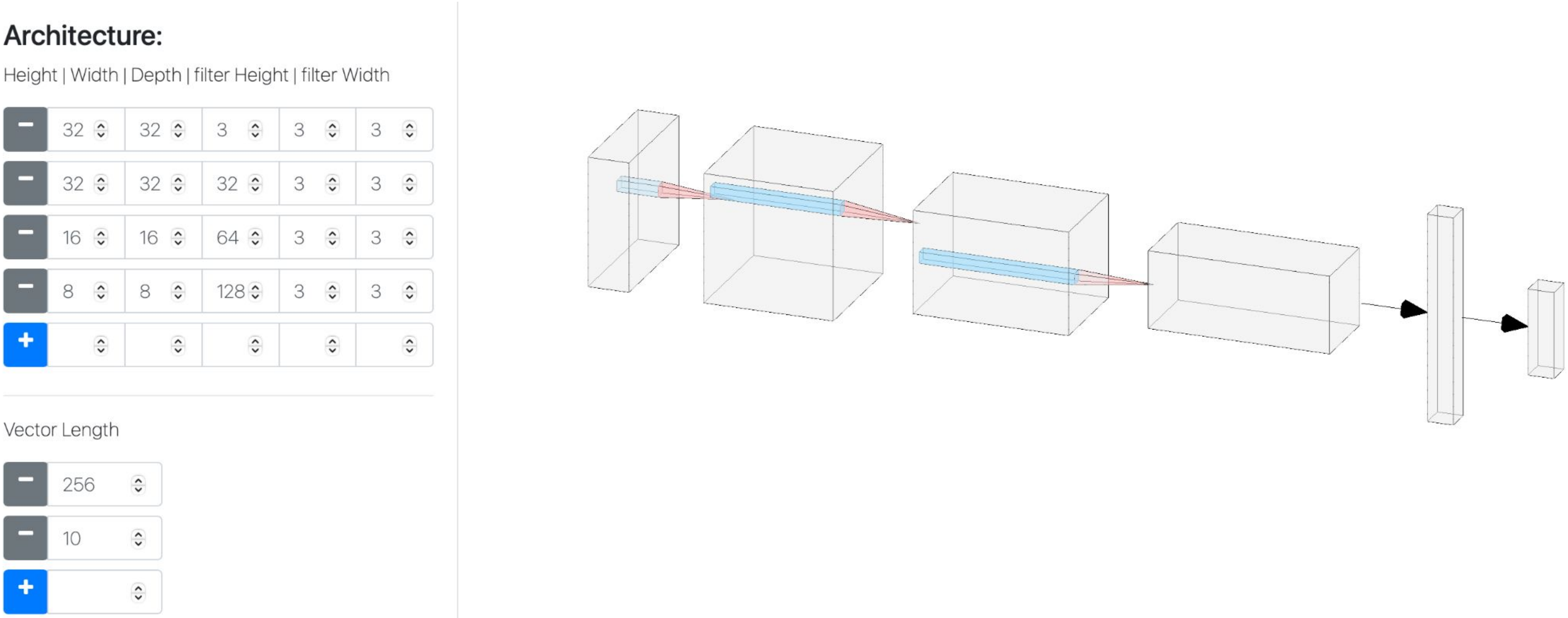
We used the same parameters:

- Cross Entropy Loss
- Adam Optimizer
- Learning Rate 0.0001
- 100 epochs

We Used CIFAR-10 dataset.

Credits to Dennis Layh in the memes channel

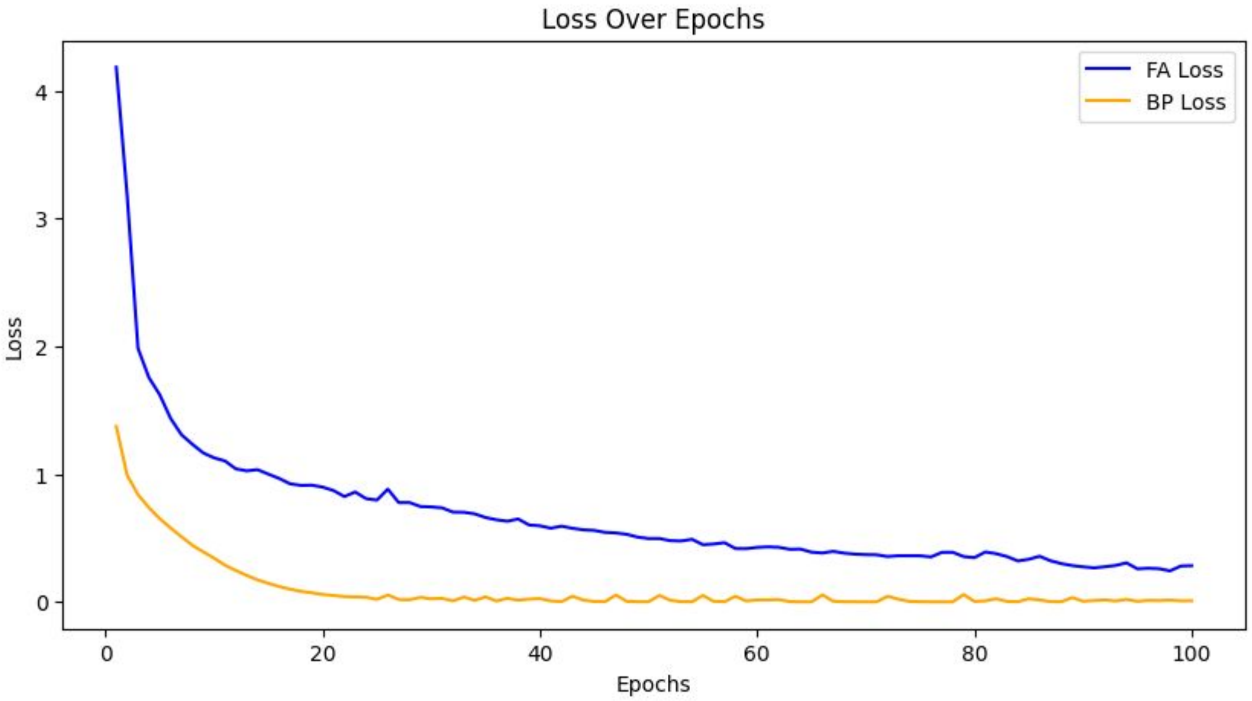
Same CNN Architecture



Made with https://alexlenail.me/NN-SVG/AlexNet.html

Results

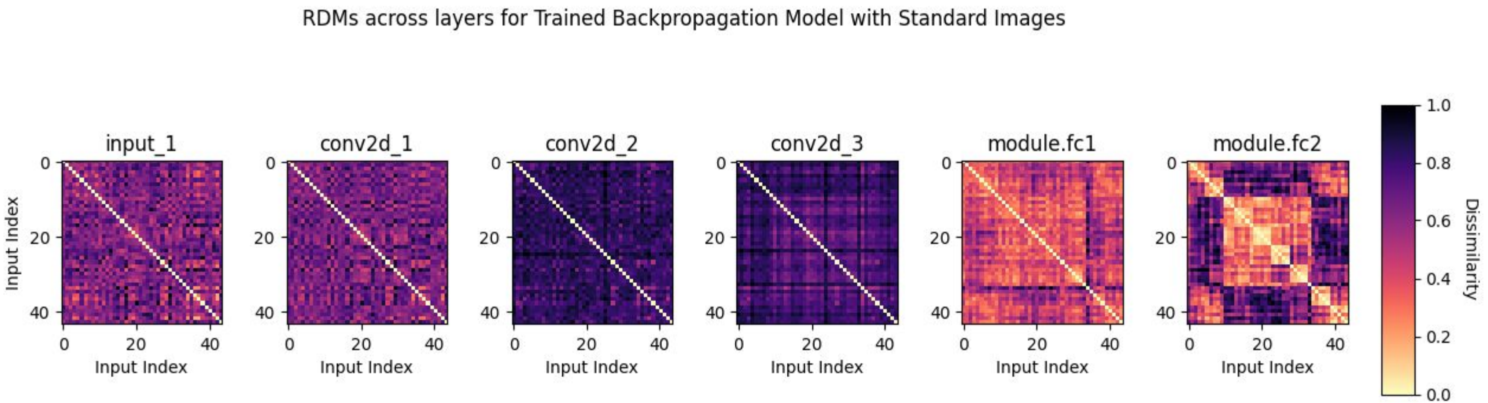
Losses across (FA vs BP)



Results

RDMs BP

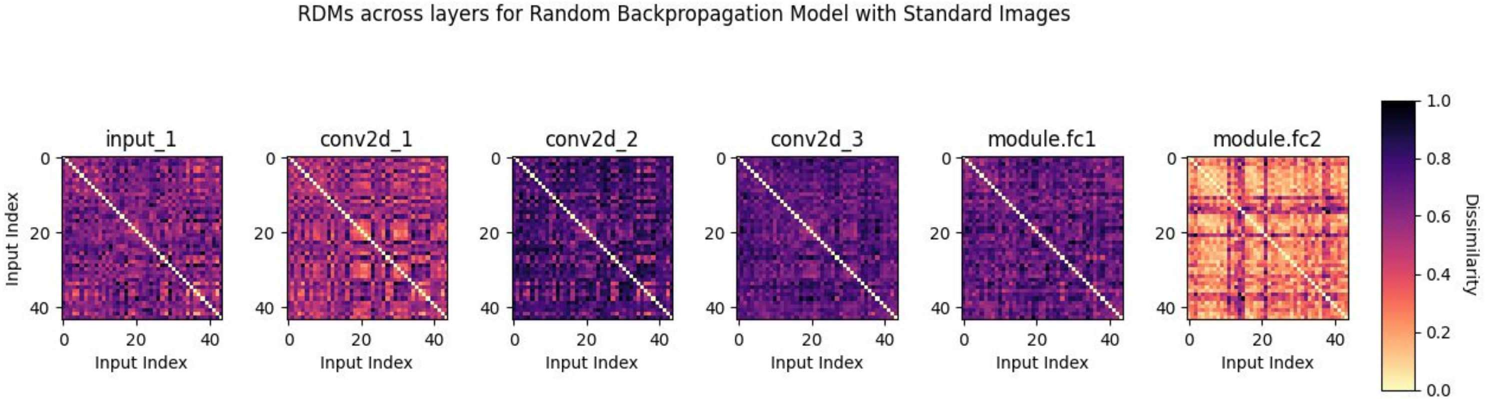
Trained BP



Accuracy

72.31%

Random BP

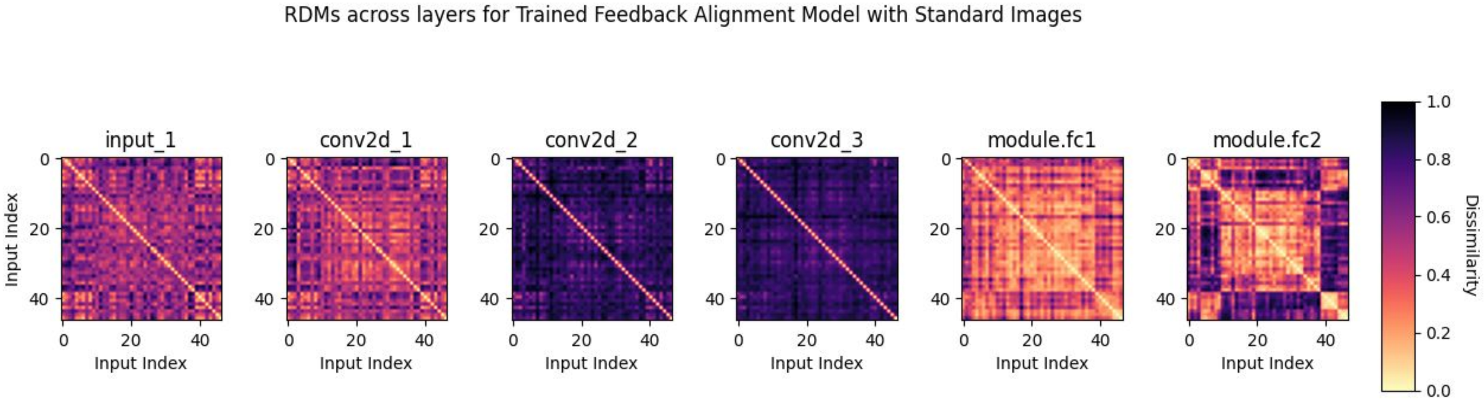


RDM: Representational Dissimilarity Matrix

Results

RDMs FA

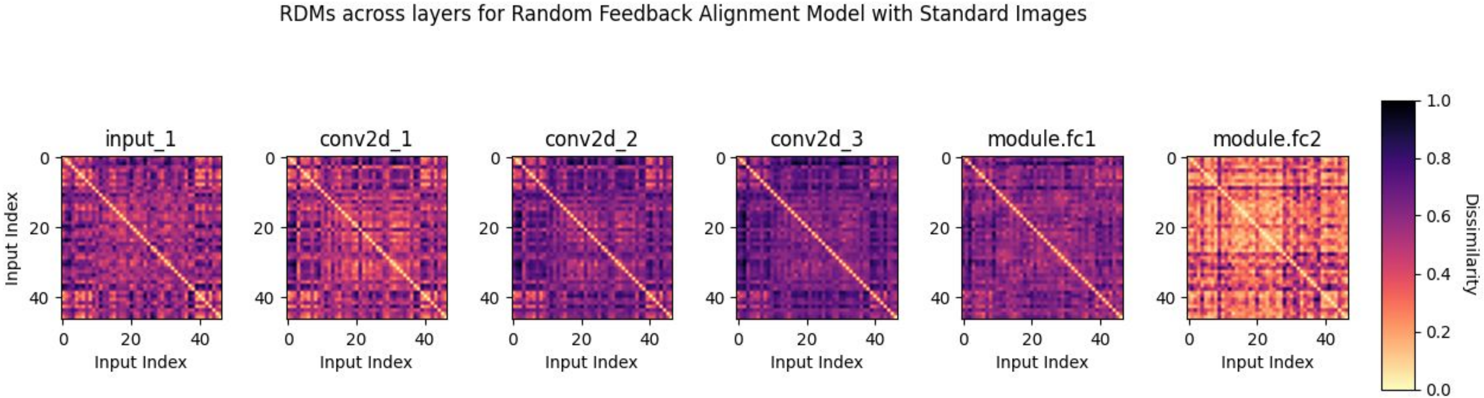
Trained FA



Accuracy

62.86%

Random FA



RDM: Representational Dissimilarity Matrix

Conclusion / Discussion

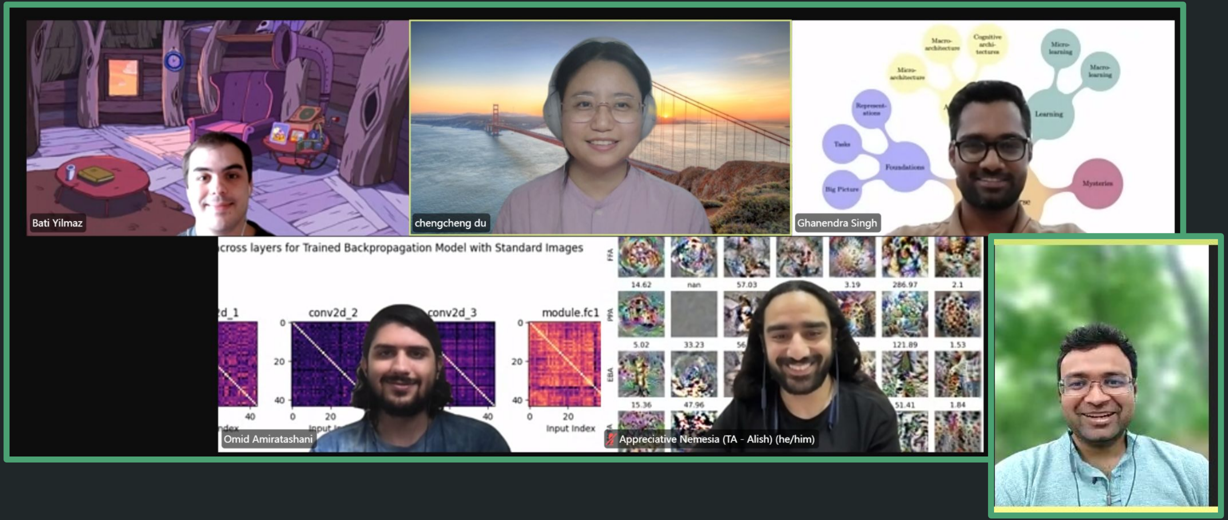
1. **Representations** of the same category of clean images appear remarkably similar in **deeper layers**, exhibiting a block diagonal structure, while **earlier layers** captures more **general** and granular visual features in DNN.

2. **Backpropagation** achieves higher accuracy than feedback alignment as it uses **precise gradient information** to update weights, ensuring more effective and optimal learning of the network parameters.

3. **Feedback alignment** shows competitive performance on **simpler tasks** but falls behind with more complex ones like CIFAR-10, which are crucial for understanding (biologically plausible) learning processes in human brain with insights from neuroscience. (Kolen and Pollack, 1994; Lillicrap et al., 2016; Sanfiz and Akrout, 2021)

References

- Ahmad N, Ambrogioni L, van Gerven MA (2020) Overcoming the weight transport problem via spike-timing-dependent weight inference. arXiv: 2003.03988.
- Cichy RM, Kaiser D (2019) Deep neural networks as scientific models. Trends Cogn Sci 23:305–317.
- Kolen JF, Pollack JB (1994) Backpropagation without weight transport. In: Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN’94), pp 1375–1380 vol.3.
- Lillicrap TP, Cownden D, Tweed DB, Akerman CJ (2016) Random synaptic feedback weights support error backpropagation for deep learning. Nat Commun 7:13276.
- Lillicrap TP, Santoro A, Marris L, Akerman CJ, Hinton G (2020) Backpropagation and the brain. Nat Rev Neurosci 21:335–346.
- Mehrer J, Spoerer CJ, Kriegeskorte N, Kietzmann TC (2020) Individual differences among deep neural network models. Nat Commun 11:5725.
- Sanfiz AJ, Akrout M (2021) Benchmarking the accuracy and robustness of feedback alignment algorithms. arXiv: 2108.13446.



Aakash Agrawal

Thank you for your attention!

And a special thanks to our TA, Alish Dipani, our project TA, Aakash Agrawal!

Also the Kolen-Pollack algorithm (Kolen and Pollack, 1994), which proposed a very similar bio-plausible learning rule idea before.

neuromatch academy

General Discussion / Q&A

Do you have any questions or remarks?