LIGHT-IN-THE-LOOP: USING A PHOTONICS CO-PROCESSOR FOR SCALABLE TRAINING OF NEURAL NETWORKS

Julien Launay, Iacopo Poli, Kilian Müller, Igor Carron, Laurent Daudet, Florent Krzakala, Sylvain Gigan
Contact: firstname@lighton.ai
Scaling deep learning: more compute is all you need

New co-design philosophy: hardware for learning beyond backpropagation

🔥 Neural networks are growing larger... 175 billion trainable parameters in GPT-3! [2]
✅ Staggering correlation between size and performance. [1]
👏 Enabling larger models will benefit deep learning.
😊 How can we scale training hardware to such large models?

Neural networks are growing larger...

Staggering correlation between size and performance.

Enabling larger models will benefit deep learning.

How can we scale training hardware to such large models?
LightOn OPU: the first at scale photonic co-processor in the cloud

On the hardware side...

💡 Multiple light scattering through a diffusive medium provides massively parallel random projections! \(^3\)

📚 Leveraging holography to retrieve a linear random projection.

1Mx300k random projection in 1.25ms with less than 40W... with expected x10 performance every 2 years.

up to 2 trillions parameters per RP, equivalent to ~1 petaOPS with 2TB cache memory in a non-von Neumann architecture.

On the training side...

💡 Direct Feedback Alignment (DFA)... scales to modern deep learning tasks and architectures. \(^4\)

Replace BP updates, \[ \delta W_j = - [(W_{i+1}^T \delta a_{i+1}) \odot f'_j(a_i)] h_{i-1}^T \]

with a random projection of the error, \[ \delta W_i = - [(B_i^T e) \odot f'_i(a_i)] h_{i-1}^T. \]

🎉 parallelized backward pass

Photonic training demonstrated on MNIST: more coming soon, holographic photonic core OPU pre-release end of 2020.
Light-in-the-loop: poster overview

slides

5–7  Training neural networks with Direct Feedback Alignment

8–9  Holography and photonic for linear random projections

10  Light-in-the-loop: photonic training of neural networks
Training neural networks the usual way: backpropagation of the error (BP)

Training a neural network is a credit assignment problem: find which neurons are to blame for the error at the output.

Backpropagation is the canonical training method, and assigns blame precisely to each neuron.

But doing so brings practical limitations (e.g. no parallelization of the backward pass) and is not biologically plausible.

Can we train neural networks differently?

alternative methods can combine computational and biological motivations
Training neural networks with... **direct feedback alignment (DFA)**

**Biologically-inspired (weight transport problem):** uses **random weights** to deliver feedback from global loss.

**Puts a single random projection at the cornerstone of training and enables parallelization of the backward pass.**

\[
\delta W_{i}^{BP} = - \left[ (W_{i+1}^T \delta a_{i+1}) \odot f_i'(a_i) \right] h_i^{T-1} \quad \delta W_{i}^{DFA} = - \left[ (B_i^T e) \odot f_i'(a_i) \right] h_i^{T-1}
\]

**Mathematical intuition:** \( (B_i^T e)^T (W_{i+1}^T \delta a_{i+1}) > 0 \Rightarrow \) DFA update within 90° of BP

(going roughly in the right direction is enough)

can be enforced by tuning \( W_{i+1} \): **learned alignment** of the forward weights.
**DFA scales to modern deep learning tasks and architectures**

👍 Is DFA as versatile as BP as a training method?\(^5\)

✅ Graph Convolutional Neural Networks  
✅ Neural scene modeling with NeRF  
✅ Massive hybrid Transformer architectures

![t-SNE of graph embeddings learned with DFA.](image1)  
![Synthetic 3D scene learned with SOTA methods.](image2)  
![DFA and BP can be mixed to accelerate the training of large architectures with limited accuracy cost.](image3)

😊 One caveat: DFA doesn't work on convolutional layers (yet!) \(^7, 8\)

.drawText(position=2, text='Further research work is needed to improve theory of DFA and build principled training methods. BP has its own tricks: dropout, batch normalization, etc.

 génértext(position=2, text='Hybrid strategies (DFA+BP) can work out of the box in large architectures.')
LightOn OPU: a photonic co-processor for random projections

💡 Leveraging photonics brings a number of advantages:

- Massively parallel processing, with the entire projection computed at once;
- Very high dimensional input and output, with easy scaling;
- Energy-efficient hardware, as the computation is mostly passive.

(OPUs have already been demonstrated in a diverse set of use cases:

- Molecular dynamics studies \[^9\]
- Theoretical analysis of NNs \[^10\]
- Reinforcement learning \[^11\]

Anomaly detection on SARS-CoV-2 glycoprotein.  
Recovering the double descent curve.  
Playing PacMan with model-free RL.

Easy-to-use, the photonics are abstracted away: opu.transform1d(x) in Python; can use Numpy/PyTorch arrays.

LightOn: the first and only photonic machine learning co-processor available in the cloud now!
Going linear with an **holographic photonic core**

🤔 Current OPUs deliver a **non-linear** random projection, $|B\mathbf{x}|^2$ not suitable for all applications.

🚀 We leverage **holography** to recover a linear operation from non-linear measurements.

✨ The magic: **technology stack remains identical**, enabling fast iterations.

💭 **Massive potential** for optical linear random projections:

- **Randomized linear algebra**
- Localized sketching to compress large data streams.
- **Optical training**
- Optical Direct Feedback Alignment to train neural networks.

Pre-release of **holographic photonic core OPU in the cloud** end of 2020.
Light-in-the-loop: photonic training of neural networks

💡 Implement the random projection of DFA optically:

A single optical random projection to train the entire network: we take slices for each layer

(operation performed with OPU)

(operation performed with GPU)

Agnostic to neural network architecture: can be widely applied, beyond largest architectures in deep learning.

Demonstrated on MNIST, with scaling to other tasks and architectures coming soon.

95.8% accuracy vs 97.7% on GPU for considered architecture

🎉 The first time a neural network is trained with light-in-the-loop!
Conclusion and outlooks

LightOn: OPUs are the first and only photonic machine learning co-processor available in the cloud now!
More information at cloud.lighton.ai, including on our research program.

🎉 The first time a neural network is trained with light-in-the-loop:

- We leverage learning beyond backpropagation to enable the use of advanced photonic hardware;
- Our accelerator is architecture-agnostic and scales to layers comprising millions of parameters.

🗓 Pre-release of holographic photonic core OPU in the cloud end of 2020.

Interested in knowing more about our technology? Check-out our white paper at lighton.ai!
References


see https://medium.com/@LightOnIO