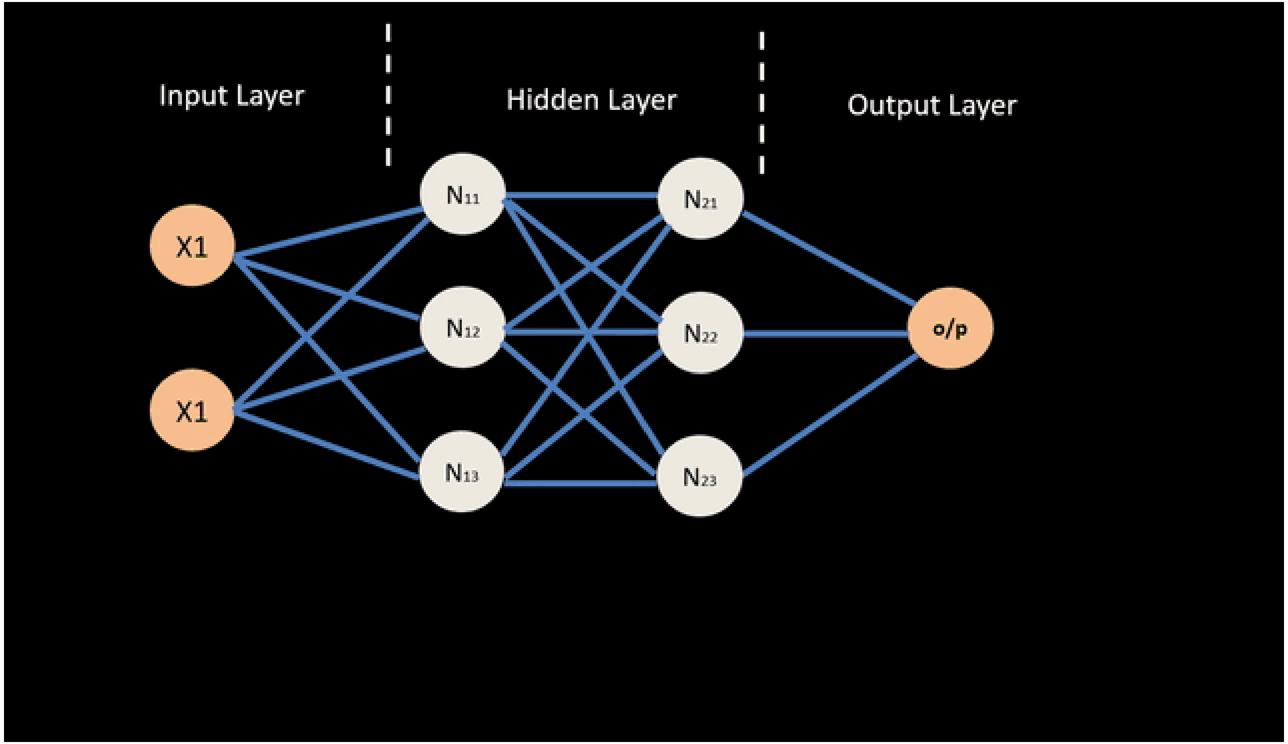
Lecture 11-a

Unconventional Algorithms and Hardware for Neural Networks

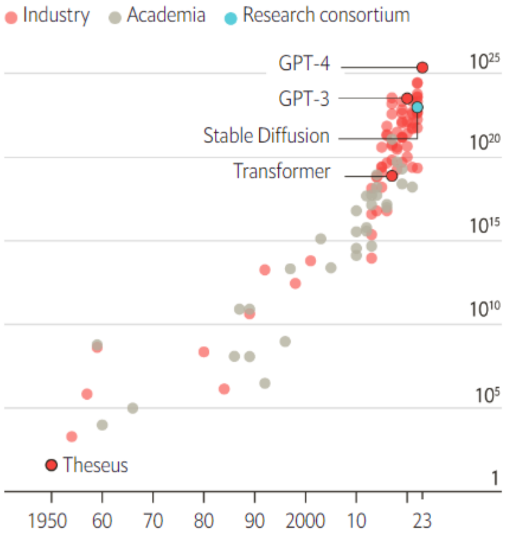
Artificial Neural Networks



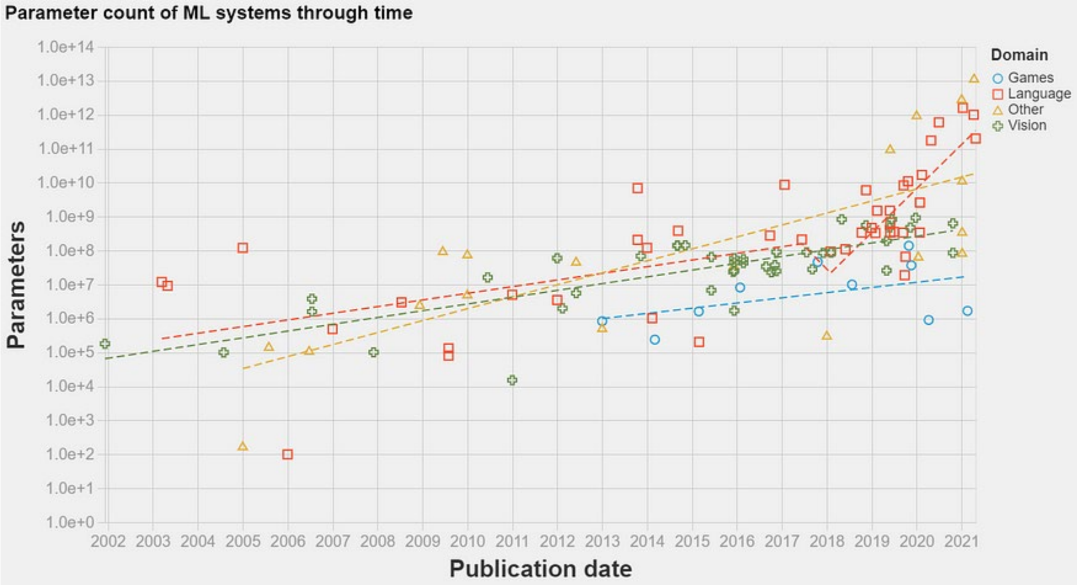
Ilker Oguz, LO and LAPD, EPFL

Artificial Neural Networks

Computing power used in training AI systems
Selected systems, floating-point operations, log scale



Sources: Sevilla et al., 2023; Our World in Data

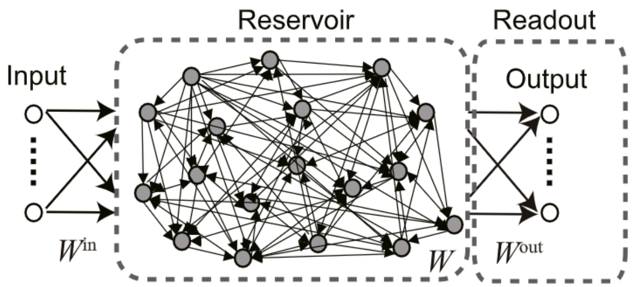


Ilker Oguz, LO and LAPD, EPFL

Reservoir Computing

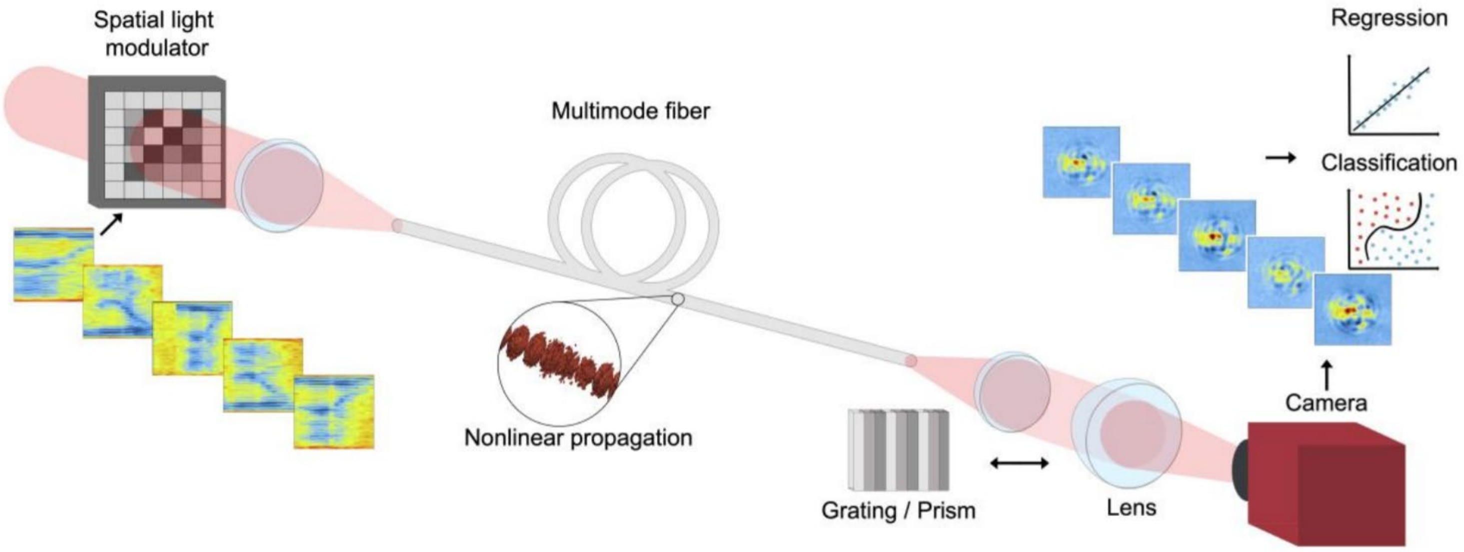
Ilker Oguz, LO and LAPD, EPFL

- Consists of high-dimensional, fixed, non-linear connections for transforming data, only output weights are optimized.
- Especially useful for fast operation and low training cost
- Can be realized by **high dimensional** and **nonlinear** physical systems.



(a) Conventional RC

Computing with Spatiotemporal Nonlinearities of MMFs

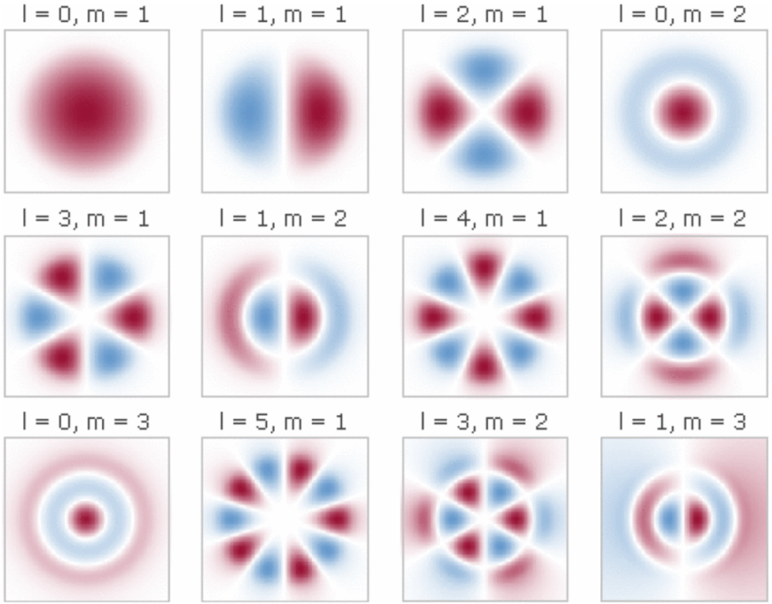


Teğin, U., Yıldırım, M., Oğuz, İ., Moser, C. & Psaltis, D. Scalable optical learning operator. *Nat. Comput. Sci. 2021 18* **1**, 542–549 (2021).

Computing with Spatiotemporal Nonlinearities of MMFs

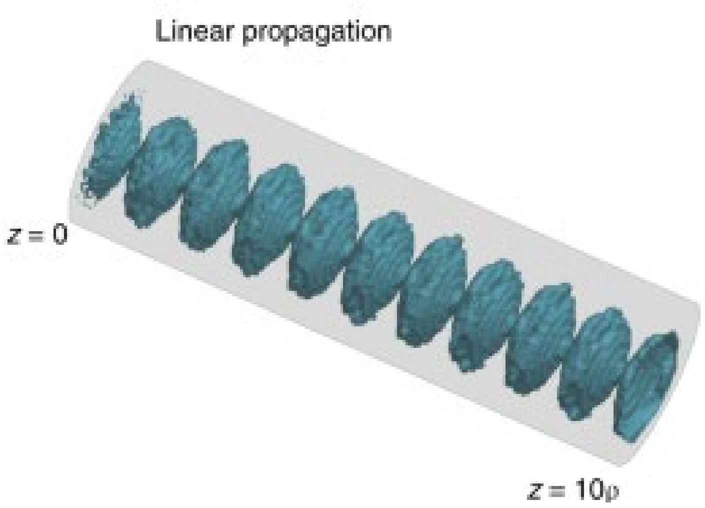
- In multimode fibers light propagate in discrete channels, depending on their properties MMFs can support up to millions of channels:
  $E\,(\mathrm{x}, \mathrm{y}, \omega) = \sum_{n}^{N} \mathrm{A}_n F_n(\mathrm{x}, \mathrm{y}, \omega)$

- At high intensities modes start to couple each other due to light-matter interactions.

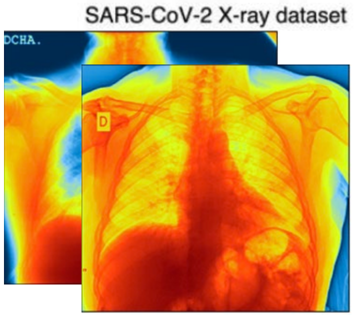


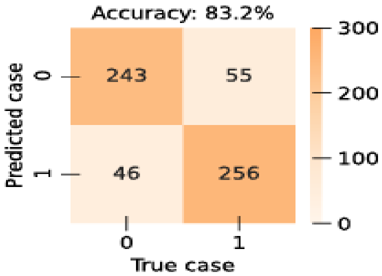
Computing with Spatiotemporal Nonlinearities of MMFs

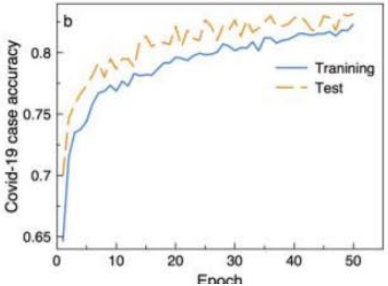
Ilker Oguz, LO and LAPD, EPFL

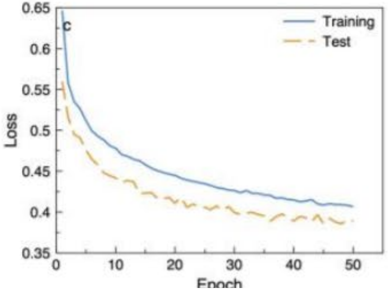


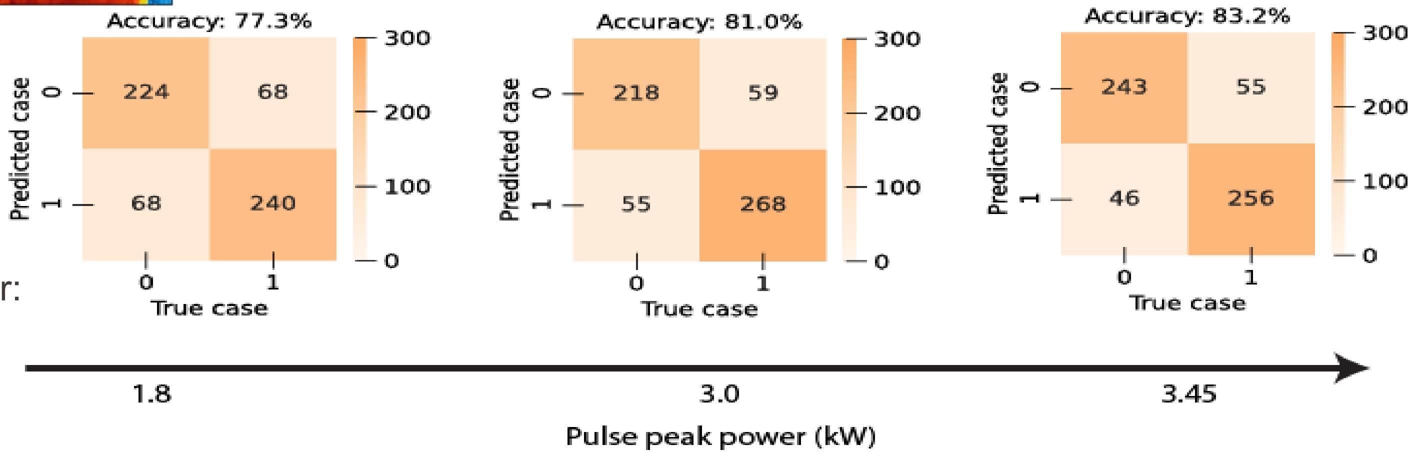
Computing with Spatiotemporal Nonlinearities of MMFs







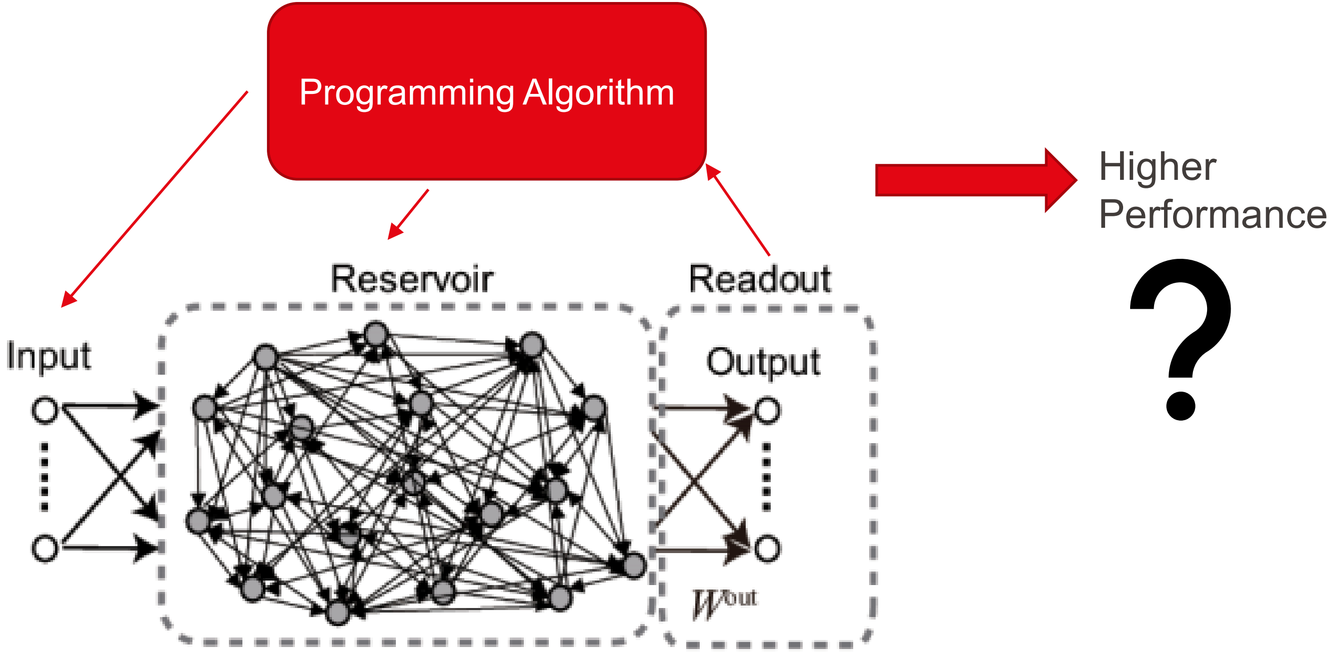




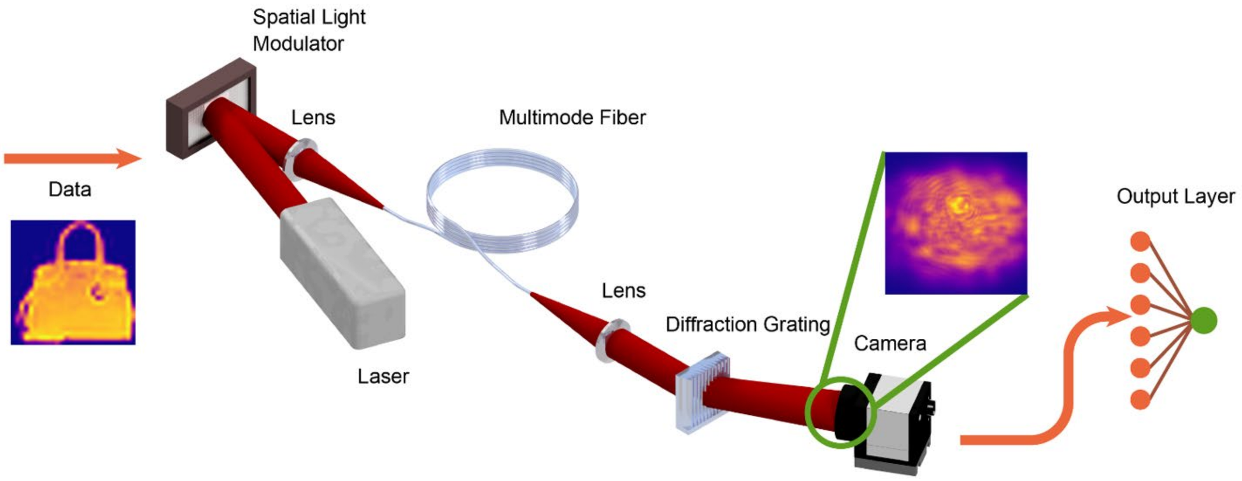
W/out fiber: 77%

Ilker Oguz, LO and LAPD, EPFL

Programming Propagation inside MMFs



Programming Propagation inside MMFs



Programming Propagation inside MMFs

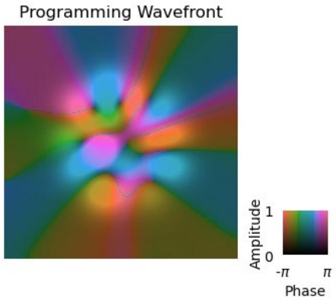
Programming Parameters

Mode Shape Coefficients

Experimental Parameters

- Intensity
- Diffraction Angle of Blazed Grating
- Pattern Displacement
- Defocus

$$WF(x,y) = \sum_{n=1}^{N} (a_{2n-1} + i * a_{2n}) * F_n(x,y)$$



Optimization of Programming Parameters

$f^{\sim}(x, w') \rightarrow y$

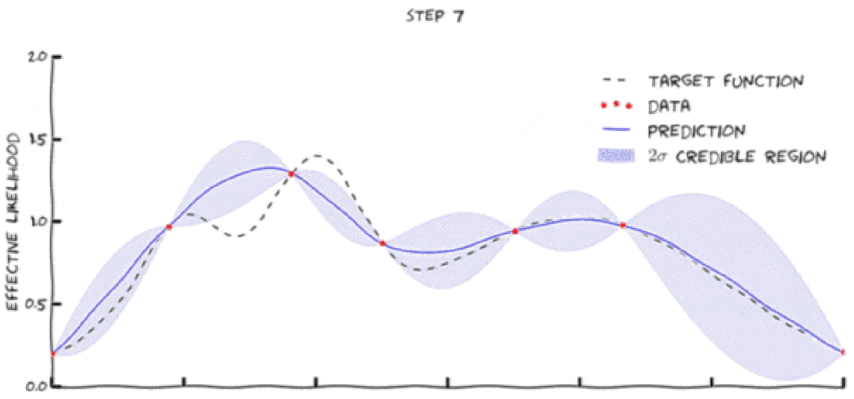
Surrogate Model: Simple function(RBFs) to fit to experimental results

New Sampling Points for Exploring/ Exploiting a Solution

Outputs from Experiment for Refining the Surrogate Model

Experiment: Unknown function, costly to evaluate, tens of parameters

$f(x, w) \rightarrow y$



Forward-Forward Algorithm

The Forward-Forward Algorithm: Some Preliminary Investigations

Geoffrey Hinton
Google Brain
geoffhinton@google.com

Training with FF:
for i in layerCount:

1- optimize W_{ij} such that goodness is high for positive samples and low for negative samples:

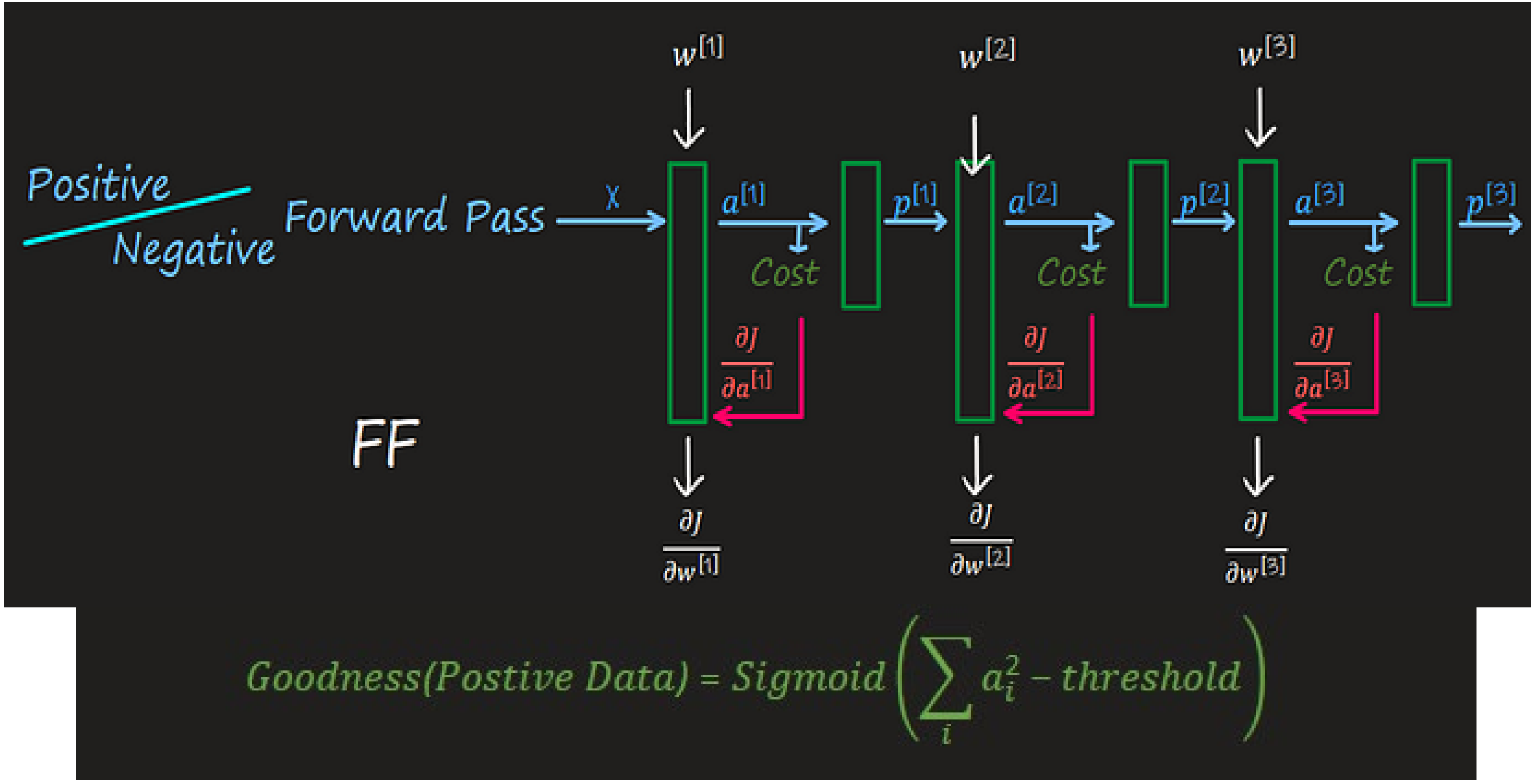
$p(\,positive\,) = \sigma\left(\sum_j y_j^2 - \theta\right)$, where $y_j = f(\sum_i^N W_{ij} x_i)$

2- Normalize $\sum_j y_j^2$'s for each sample before passing it to the next layer

EPFL

Forward-Forward Algorithm

Ilker Oguz, LO and LAPD, EPFL

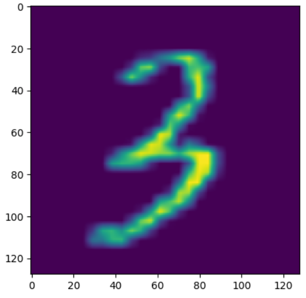


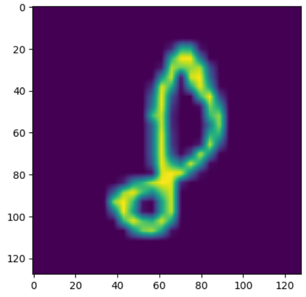
Forward-Forward Algorithm

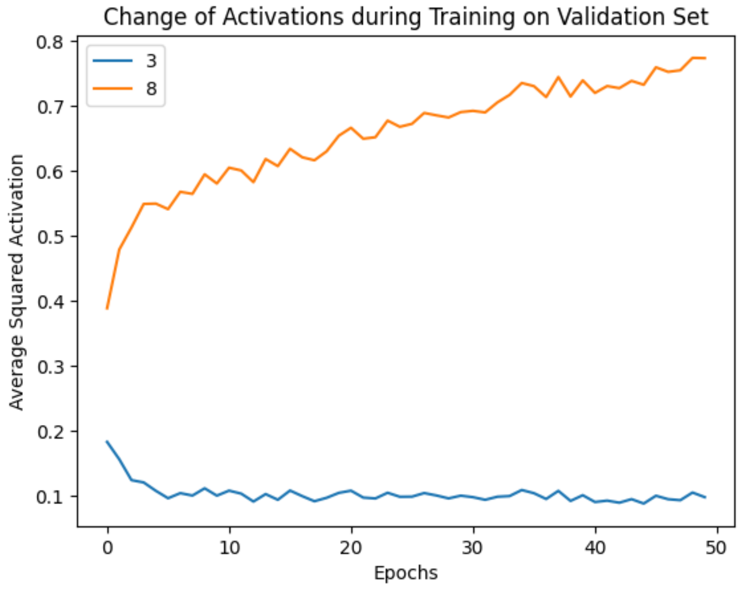
- The connections between layers can be unknown transforms ➔ Can be used by very-low power analog computing devices
- Each layer is updated at a time ➔ No need to save all activations ➔ Less memory consumption
- Biologically plausible
- Can be used by very-low power analog computing devices

Forward-Forward Algorithm on MNIST

- 8’s are labelled as positive and 3’s are labelled as negative data
- Fully connected layer with 1024 neurons are trained FFA
- Thresholding activations: 92.7 %
- Backprop training: 96.5 %



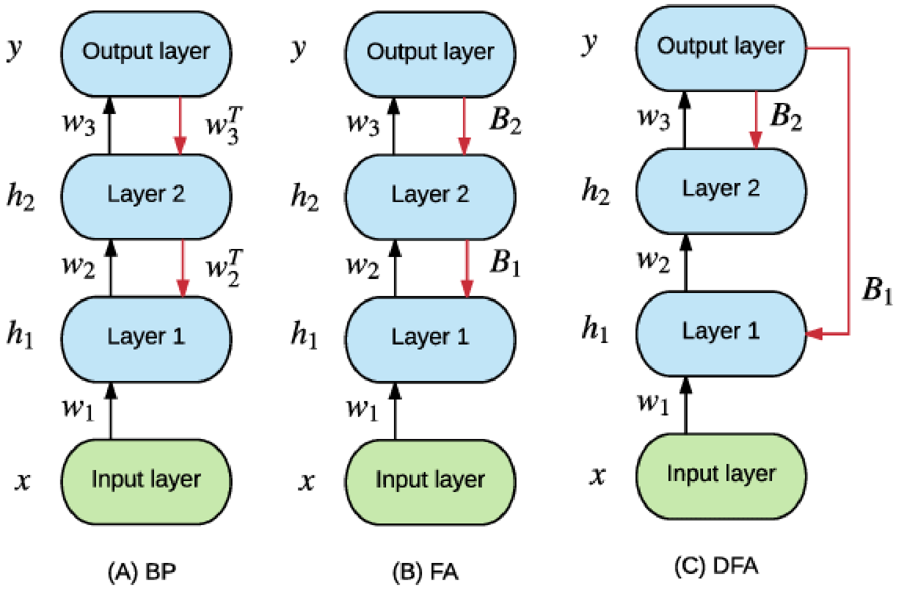




Ilker Oguz, LO and LAPD, EPFL

Feedback Alignment

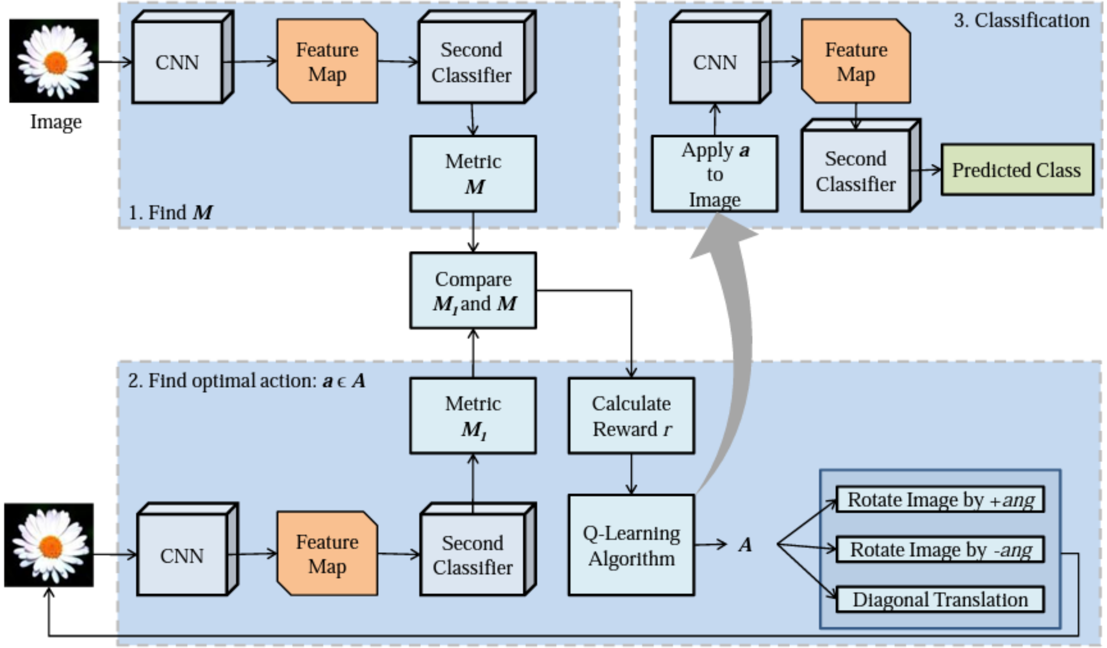
- Instead of backpropagating errors with same weights as the forward pass, feedback alignment uses random backward connection.
- When w and B matrices are from similar distribution, convergence is observed.



Error Rate

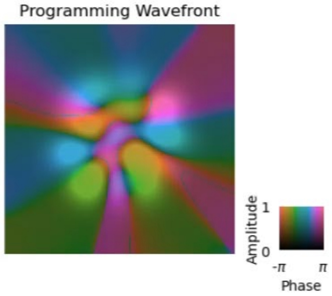
	Dataset	
	MNIST	Fashion MNIST
BP	0.91	9.20
FA	1.7	13.06
DFA	1.61	12.81

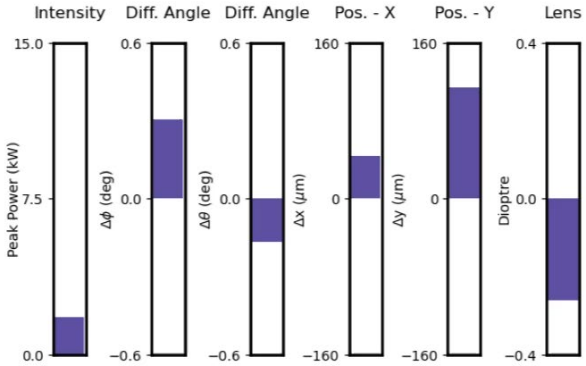
An Alternative: Reinforcement Learning

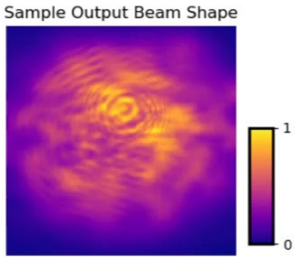


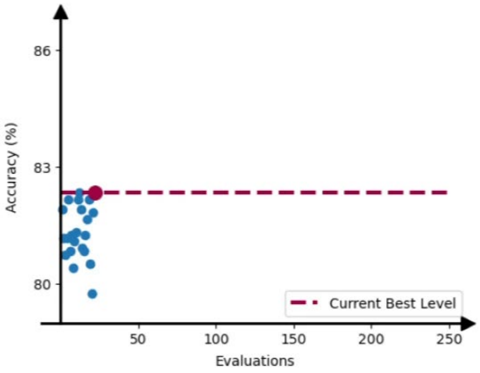
Hafiz, Abdul Mueed. "Image Classification by Reinforcement Learning With Two-State Q-Learning." *Handbook of Intelligent Computing and Optimization for Sustainable Development* (2022): 171-181.

Programming Propagation for MNIST-Fashion





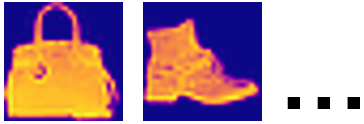




Ilker Oguz, LO and LAPD, EPFL

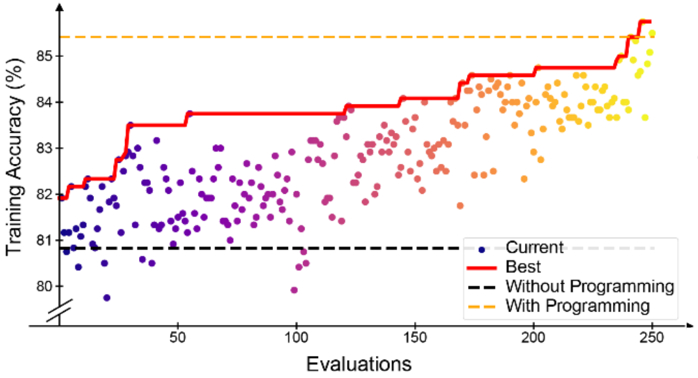
Programming Propagation for MNIST-Fashion

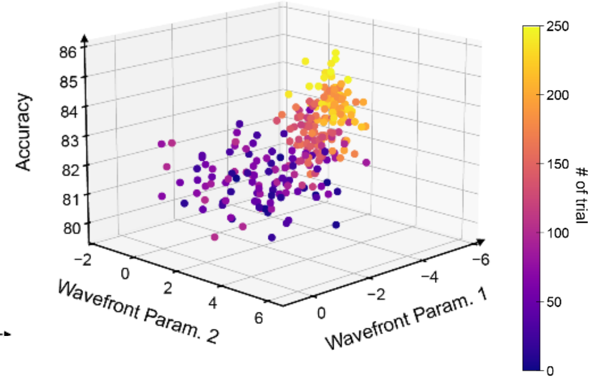
1200 training and 300 test samples

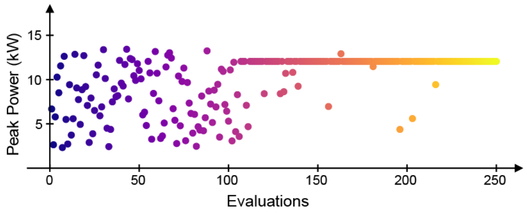


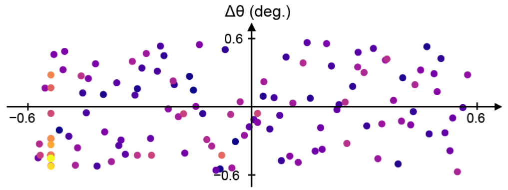
Bag Boot

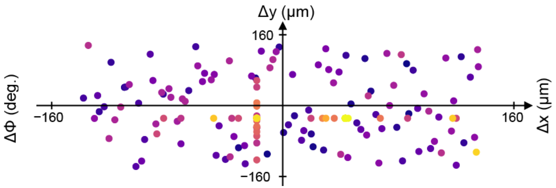
$\Delta Acc = \ +5.0\%$









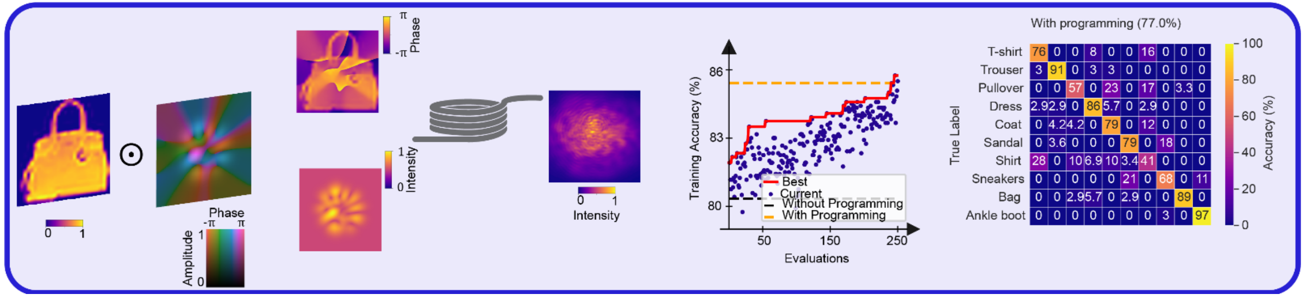


Ilker Oguz, LO and LAPD, EPFL

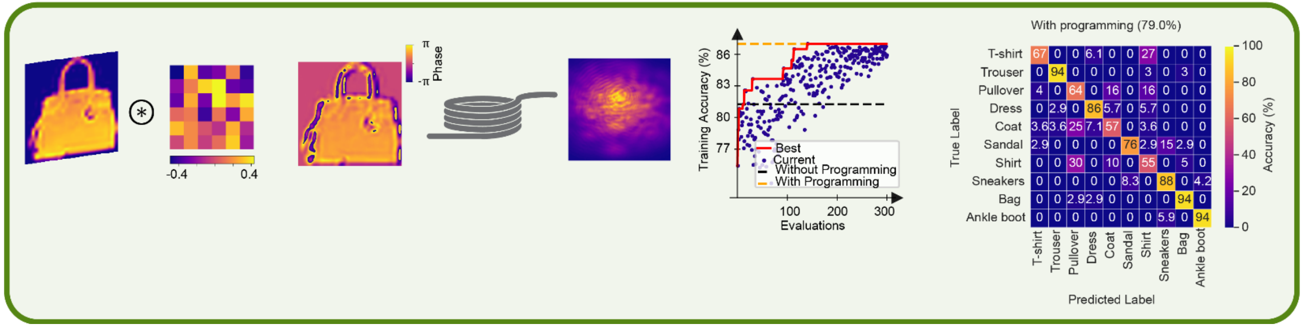
EPFL

Programming Propagation for MNIST-Fashion

Complex Field Control

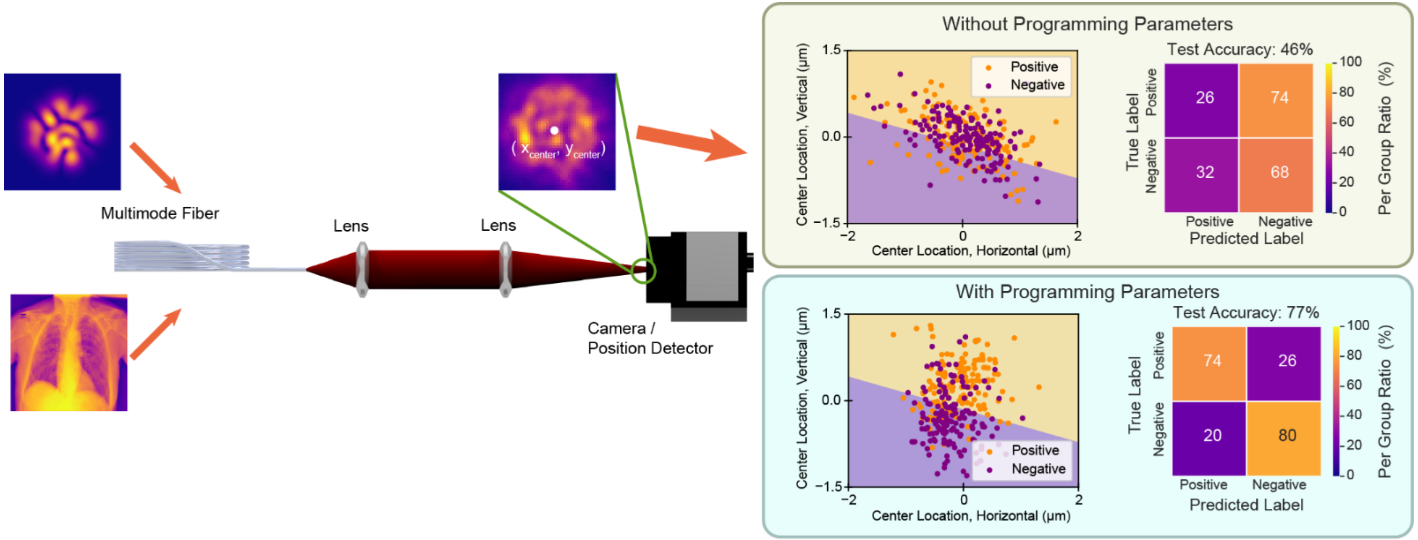


Convolution with Kernel



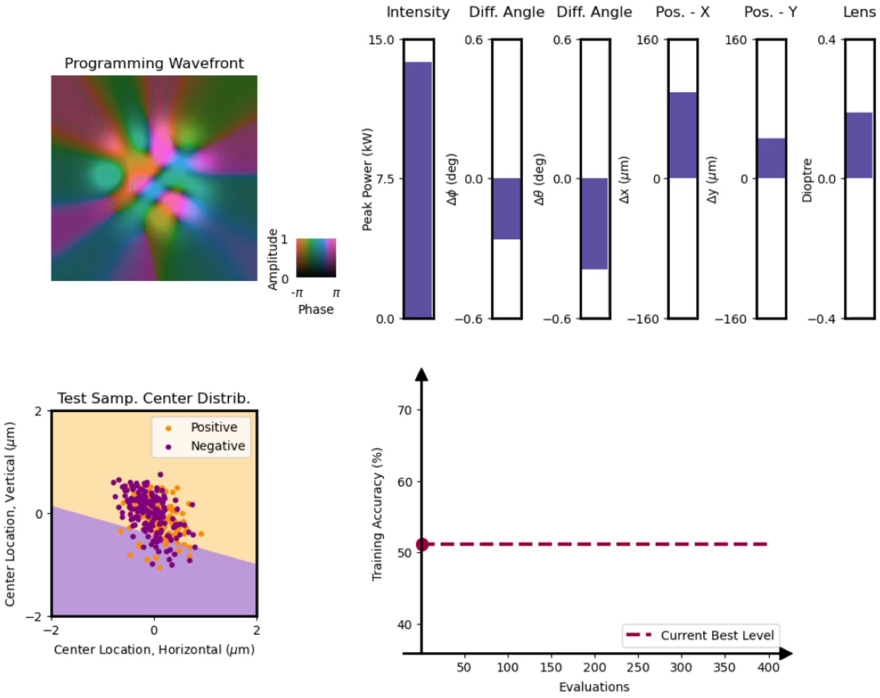
Ilker Oguz, LO and LAPD, EPFL

Programming Propagation for All-Optical Classification



Ilker Oguz, LO and LAPD, EPFL

Programming Propagation for All-Optical Classification



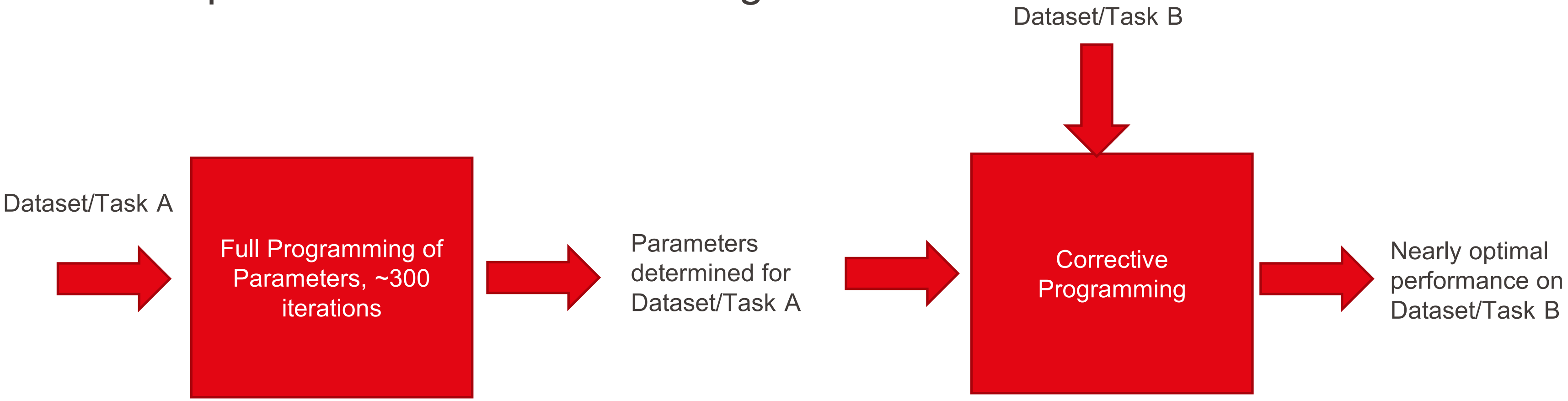
Ilker Oguz, LO and LAPD, EPFL

Programming Propagation for All-Optical Classification

Network Structure	Total Number of Parameters	Operations per Sample on Digital Computer (FLOP)	Accuracy on Melanoma dataset (%)	Accuracy on COVID-19 dataset (%)
LeNet-5	82826	1175640	64.9	74.6
MMF + classification with output location (with programming)	55	2029	61.3	77.0

Transferring Programming Parameters between Different Tasks

- Current method requires ~300 iterations over the dataset for programming to converge.
- For 1500 samples at 50 images per second, full programming corresponds to ~3 hours of training.



Ilker Oguz, LO and LAPD, EPFL

Transferring Programming Parameters between Different Tasks

CelebA - Gender

CelebA – Young/ Not Young

Full P
Para

Test Acc

readout
ngle
l pass,
version

Test Acc:
69.0% ⇒ 76.0%

- Training from scratch : 76.3%

Comparison with GPU-based NNs

Ilker Oguz, LO and LAPD, EPFL

Digital

Network Structure	Total Number of Parameters	Operations per Sample on Digital Computer (FLOP)	Test Accuracy on Age Task	Test Accuracy on Gender Task
LeNet-5	~82k	~1.2M	63.0	75.2
7-layer Convolutional NN	~410k	~65M	65.3	80.1
MMF + linear output layer	2026	4050	59.0	69.0
Programmed MMF for Age Task + linear output layer	2078	6075	67.0	76.0
Programmed MMF for Gender Task + linear output	2078	6075	64.7	76.3

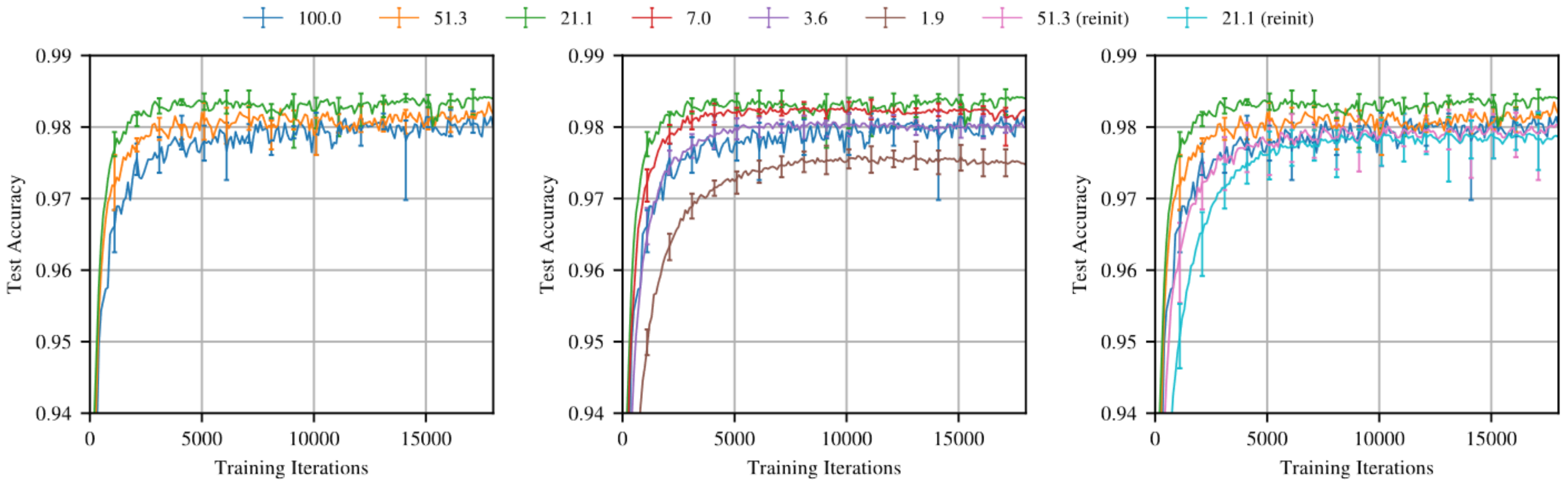
Optical + Digital

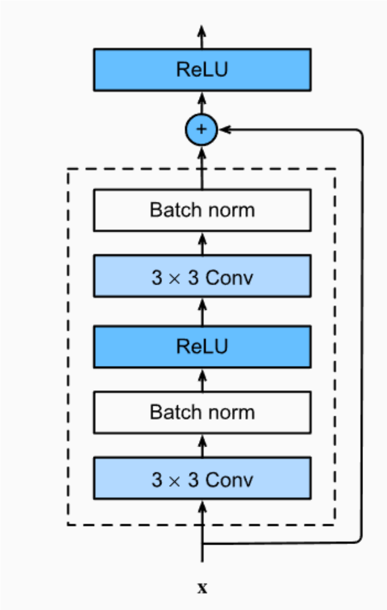
Published as a conference paper at ICLR 2019

THE LOTTERY TICKET HYPOTHESIS: FINDING SPARSE, TRAINABLE NEURAL NETWORKS

Jonathan Frankle
MIT CSAIL

Michael Carbin
MIT CSAIL





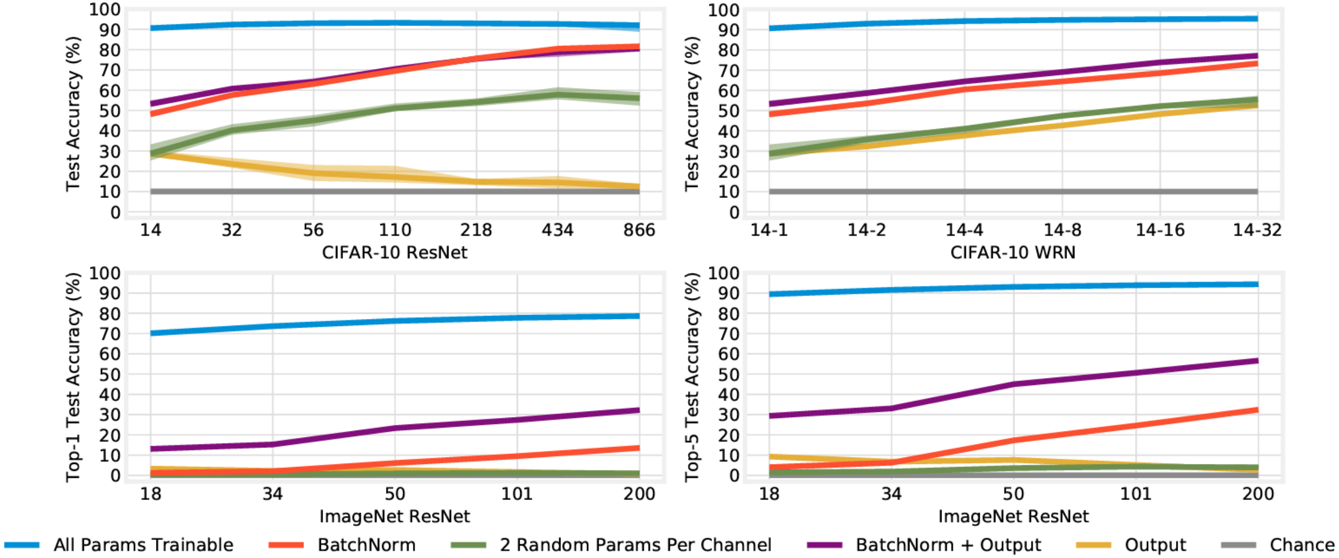
Published as a conference paper at ICLR 2021

TRAINING BATCHNORM AND ONLY BATCHNORM: ON THE EXPRESSIVE POWER OF RANDOM FEATURES IN CNNS

Jonathan Frankle*
MIT CSAIL
jfrankle@mit.edu

David J. Schwab
CUNY Graduate Center, ITS
Facebook AI Research
dschwab@fb.com

Ari S. Morcos
Facebook AI Research
arimorcos@fb.com



Ilker Oguz, LO and LAPD, EPFL