Convolutional Deep Optical Learning Devices and Architectures

Kelvin H. Wagner and Sean McComb

Optoelectronics Center, Dept ECE, University of Colorado, Boulder, CO 80309 kelvin@colorado.edu

Summary. A coherent-optical implementation of a Rectifying Linear Unit (ReLu) as an interferometric phase-sensitive bidiectional switch along with convolutional layers using lenslet arrays and multiplexed Fourier holograms allows the efficient implemention of Deep Neural Networks (DNN) in a self-aligning multilayer adaptive-holographic optical neural network.

Deep learning has emerged as the leading approach to a variety of cognitive tasks.[1, 2] These computationally-demanding deep neural networks (DNN) with vast numbers of hidden units and many layers were not previously trainable due to numerous local minima of the back-propagation training procedure and small derivatives of the sigmoid nonlinear activation function for large neuron inputs which leads to small gradients of the weights and slow convergence.[3] Recent improvements in neuron functionality, training algorithms, network architectures, and GPU computational powers have made deep learning practical for networks with up to about $10^3 - 10^4$ inputs. For even larger problems there may be an opportunity for an optical hardware implementation of an analog DNN co-processor for accelerating both training and processing.

Progress in liquid crystals, VLSI, and nonlinear optics may make the implementation of a deep optical neural networks a compelling alternative to simulation on GPUs. So even though the inevitable progress of digital electronics has made many TFLOP neural network simulations routine, deep convolutional neural networks will continue to overrun these capabilities, and vastly increased throughputs may be achieved with an optically implemented neural network, albeit at a lower analog precision. As a nominal benchmark, GoogLeNet operates on images up to $244 \times 244 \times 3$, and with a GPU accelerator can achieve frame rates of 75 mini-batches per second, but learning runs can take a week or more. Much larger image resolution is possible with optically implemented DNN using current Spatial Light Modulators (SLMs) up to 4k×4k. And SLM frame rates of 5 KHz are currently available with MHz rate liquid crystal under development promising future input devices and optical neuron arrays that allow for orders of magnitude improvement in both image size and training speed.

For large inputs to sigmoidal neurons the derivative of their nonlinear activation is very small, effectively shutting off the back propagation of error through them to previous layers. This inhibits learning, especially in very deep architectures. DNN have migrated from sigmoidal nonlinear neurons to Rectifying Linear Units (ReLU), shown in Fig. 1, which are a simple nonlinear activation function whose derivative is zero (or small) below threshold and unity above threshold, so the back propagating error never gets blocked for large inputs to the neuron. Such DNN of ReLU hidden units is trainable with back prop, making learning of complex cognitive tasks achievable even in very deep networks.[1]

The optical implementation of such ReLU neurons for both forward nonlinear activation and gated back propagation of the error is simplified compared to sigmoidal nonlinearity since only two transmissive values are required and for a threshold at 0 bidirectional ReLU operation is automatic. Dropout during training can be incorporated by randomly gating off the optical ReLU neurons during training in order to improve the network robustness and generalization. An optically controlled transmissive switch based on the sign of the forward propagating presynaptically summed neuron input that goes from an opaque

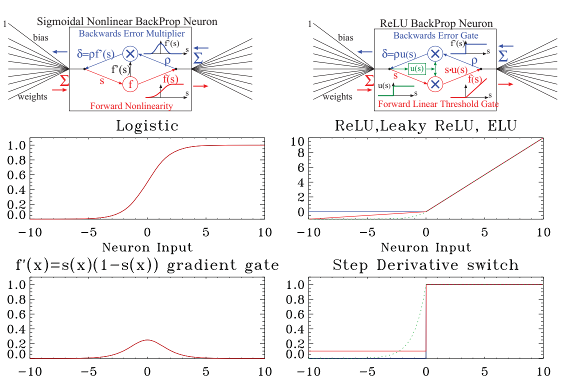

Figure 1: Sigmoidal vs Rectifying Linear Units (ReLU). Bidirectional operation, nonlinearity and derivative. A thresholded bidirectional switch implements the ReLU.

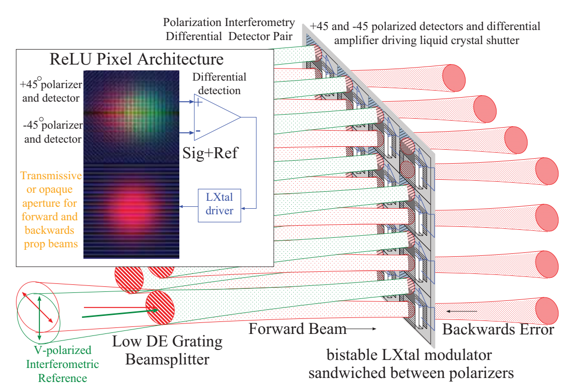

Figure 2: Coherent optical implementation of rectifying linear unit using polarization interferometry and a bidirectional liquid crystal on silicon switch that can be fabricated in large arrays.

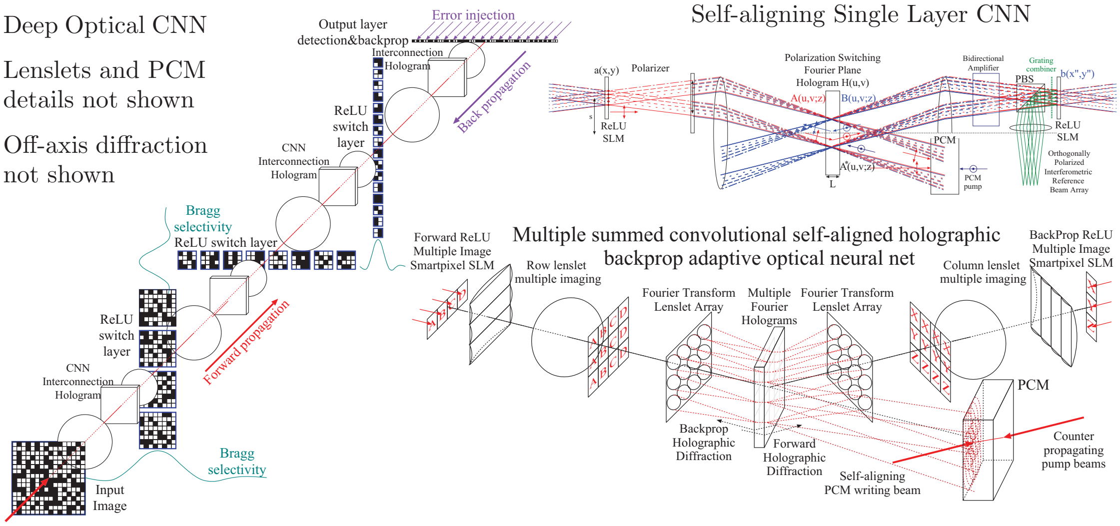

Figure 3: Multilayer convolutional neural net with each layer a matrix-vector-multiplier of image convolvers.

to transmissive state as the neuron input goes from negative to positive implements both the forward and backwards propagation functionalities. Amplitude sign encoded with $(0, \pi)$ phase can be demodulated with interferometry, for example by splitting the forward propagating signal into 2 copies, one for modulation and propagation to subsequent layers and one for detection in the neuron. Polarization interferometry using a cross-polarized reference beam and a pair of orthogonally-polarized differential detectors can be used to set the state of the neuron transmission, as illustrated in Fig. 2.

An optical architecture for a deep convolutional neural network loosely based on a previous self-aligning holographic back-propagation architecture is illustrated in Fig. 3.[4] An input image is Fourier transformed and multiplied by a spatial-frequency multiplexed array of holographically-encoded convolutional-filter transfer functions and this is Fourier transformed to produce a y multiplexed array of M output image convolutions. A 2:1 demagnifying optical system gives a lower pixel count output image, operating like the max-pooling used in deep learning and successively decreases the resolution of each subsequent feature space, thereby increasing the neuron receptive fields at each stage. Each pixel of the smart-pixel optical neuron array detects the multiple convolution outputs operates as an optical ReLU in this diagram. The next stage computes an x-multiplexed array of M^2 lower-resolution convolutionaly-filtered images diffracted in y by the thick dynamic hologram, which is thin enough to act as a Fourier-plane convolutional filter but thick enough to eliminate unwanted partial convolutions through Bragg selectivity along y. Successive multi-image convolutional stages with resolution-decreasing pooling and computing even more convolutional feature spaces are followed by fully interconnected stages using dynamic volume holograms thick enough to operate as Bragg-matched vector-matrix multipliers. Backwards propagating errors injected at the network output are diffracted off each hologram back towards the neurons where they propagate backwards through the ReLU pixels currently in the transmissive state. The backwards propagating error interferes with the phase-conjugates of the forward propagating beams within the hologram volume to update the holographic weights in each layer, completing the shift-invariant outer-product learning. The phase conjugated signals must be blocked from further propagation after recording the hologram with appropriate polarization filtering or time gating.[4] In this fashion, an all-optical self-aligning deep-learning back-propagation algorithm can potentially be implemented at speeds far in excess of that possible with digital simulations.

The requirements of deep convolutional neural networks seem ideally suited for an all optical implementation utilizing adaptive multiplexed Fourier holograms, and this is enabled by a coherent optical rectifying linear unit that is more readily implemiented than a sigmoidal back-propagation neuron.

This work is supported by the US National Science Foundation (NSF ECCS) grant number 1810508.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, p. 85, 2015.

[3] S. Hochreiter, "The vanishing gradient problem during learning," *Int J Uncertainty*, vol. 6(2), p. 107, 1998.

[4] K. Wagner and D. Psaltis, "Multilayer optical learning networks," *Appl. Opt.*, vol. 26, pp. 5061–5076, 1987.