ClassificationCam: How can optics be used in convolutional neural networks? (not final)

Anonymous ECCV submission

Paper ID ***

Abstract. The abstract should summarize the contents of the paper. LNCS guidelines indicate it should be at least 70 and at most 150 words. It should be set in 9-point font size and should be inset 1.0 cm from the right and left margins. Paper page limit is 14 pages, excluding references.

Keywords: We would like to encourage you to list your keywords within the abstract section

1 Introduction

Deep neural networks have found success in a wide variety of applications, ranging from computer vision to natural language processing to game playing [1]. Convolutional neural networks (CNNs), capitalizing on the spatial invariance of certain properties of images, have been especially popular in computer vision problems such as image classification, image segmentation, and even image generation [2,3,4]. As performance on a breadth of tasks has improved to a remarkable level, the number of parameters and connections in these networks has grown dramatically, and the power and memory requirements to train and use these networks have increased correspondingly.

While the training phase of learning parameter weights is often considered the slow stage, large models also demand significant energy during inference due to millions of repeated memory references and matrix multiplications. For example, the final version of Google DeepMind's AlphaGo in [5] used 40 search threads, 48 CPUs, and 8 GPU to play a game of Go. Live imaging and sensing applications face the additional challenge of power-hungry sensors and high bandwidth transfer of data to feed into the downstream computer vision algorithms [6]. For these reasons, it remains difficult for embedded systems such as mobile vision, autonomous vehicles and robots, and wireless smart sensors to deploy CNNs due to stringent constraints on power and bandwidth.

Optical computing has been tantalizing for its high bandwidth and inherently parallel processing, potentially at the speed of light. Furthermore, certain linear transformations can be performed in free-space or on a photonic chip with minimal to no power consumption, e.g. a lens can take a Fourier transform "for free" [7,8]. Nonlinear operations could also be addressed optically, drawing on passive nonlinear materials or devices whose refractive indices or transmission states are dependent on optical input [9,10]. An optimizable and scalable set of optical configurations that preserves these advantages and serves as a framework for building optical CNNs would be of interest to computer vision, robotics, machine learning, and optics communities. Optical implementation could also have the potential to expand beyond traditional operations of CNNs, potentially by harnessing wave optics and quantum optics in new ways.

We take initial steps toward this broader goal from a computational imaging approach, integrating image acquisition with computation via co-design of optics and algorithms. By pushing one or more layers of a CNN into the optics, we can reduce the workload of the electronic processor when performing inference with a CNN. Imaging systems are often characterized by their point spread function (PSF), which describes how a single point source of light propagates through the system. Hence, for a simple linear and space-invariant system, the image recorded at the output is the convolution of the original object with the system PSF [8]. This built-in convolution motivated us to explore how we could use optics to replace one or more of the layers in a CNN.

In this paper, we propose a toolbox of optical building blocks that could be used to implement common neural network layers. To evaluate these components, we build a simulation framework for testing a few variations of optical CNNs with the relevant physical constraints, including learned optical correlators, hybrid optoelectronic CNNs, and fully optical CNNs. We train these networks to perform image classification on a few different datasets (MNIST, GoogleQuickdraw, or CIFAR-10), and we compare the simulated ONN accuracy against the unconstrained computer implementation of the same network structure. To demonstrate the validity of our simulations, we build a hybrid optoelectronic two-layer network with an optical convolutional layer and electronic fully connected layer for CIFAR-10 classification. We compare performance with the same inference performed on the computer, with and without the simulated physical constraints of an optical setup.

Overview of limitations. While the proposed ONN architectures offer lower power inference on classification tasks, the physical image formation imposes several constraints on the CNN architecture, including nonnegative signal and weights when using incoherent light, no bias, limited set of nonlinearities, etc. We will discuss in more detail in the paper how much each of these constraints limit the performance of our system. Here we demonstrate proof-of-concept with bulk optics and free-space propagation, which is not necessarily practical or scalable to commercial applications. However, photonic integrated circuits could significantly help in both these regards [11,12,13]. Combination of these next-generation large-scale photonic circuits with compressed deep learning models could provide a potential route for high performance ONNs.

2 Related Work

Efficient convolutional neural networks. Since our work is motivated by the potential of optics to increase the effiiciency of CNN applications, we first review algorithms and electronic hardware also designed to address this challenge.

Pruning, trained quantization, huffman encoding, and altered architectural design have been successfully used to compress CNN models, preserving AlexNet-level accuracy on ImageNet even with 510× less memory usage and 50× fewer parameters [14,15]. On the hardware front, there are now specialized processing units for deep learning, such as TrueNorth, Movidius's USB-based neural compute stick (NCS), and Google's tensor processing unit (TPU). All of these are complementary to our approach, which still requires offline training to optimize the optical components. Other inference-focused efforts aimed at embedded vision applications have tried to incorporate a portion of the image processing on the sensor chip, eliminating or reducing the need to shuttle full image data to a processor. Analog circuitry has been used to detect edges and orientations, to perform wavelet or discrete cosine transforms, and even to execute layers of a CNN [16,17]. These approaches still rely on electronic computation on the image sensor chip, whereas our goal is to push more of the computation into optical hardware that requires no power input.

Computational cameras. Optical computation is attractive because it offers inherent parallelism and high interconnectivity, both of which are encountered when passing signals through neural networks. In the computational imaging community, many system designs already exploit the physical propagation of light through custom optics to encode information about a scene that would be lost in a standard 2D image capture. Computational cameras have been created to record depth, light fields, light transport, and more with a toolbox including coded apertures, lenses and lenslets, active illumination, and wavefront shapers [18,19,20,21,22]. In this work, we propose a computational imaging system modeled after a CNN that assists in performing classification of input images. We begin by learning an optical correlator consisting of a single convolutional layer that essentially performs template matching on images, as has been explored for optical target detection and tracking [23,24], and then expand beyond a single matched filter in hybrid optoelectronic and fully optical designs.

Optical neural networks. The concept of an optical neural network (ONN) captured the attention of many in the late 1980s to mid-1990s, primarily due to the capability of optics to perform the expensive matrix multiply of a fully connected layer. In 1985, an optoelectronic implementation of the Hopfield model, a basic model of a recurrent neural network, was created with 1D LED array input signals and a binary transmission mask [25]. This model divided the weight matrix into two parts, positive and negative, and required electronics for subtraction of the two parts and signal thresholding. Psaltis et al. further explored the potential of dynamic photorefractive crystals to store neural network weights, which could allow for optical backpropagation-based learning in ONNs [26]. Meanwhile, the optoelectronic network of a Hopfield model was extended to 2D signals by partitioning the pixels of a liquid crystal television to store an array of smaller 2D patterns [27]. Furthermore, an optical thresholding perceptron was implemented with liquid crystal light valves (LCLV), which disposed of the need to convert between optical and electronic signals between layers [28]. We draw on some of these insights for the design of our optical CNN. A more extensive overview of the varied implementations of ONNs can be found in [29].

Despite the progress in this area, as neural networks fell out of the spotlight, the demand for ONNs also waned. However, with the resurgence of CNNs that are far more powerful and computationally expensive than before, there is renewed interest in optical computing [1]. Recent works that connect efforts of the last century to modern hardware include a two-layer fully connected neural network based on programmable photonic circuits [13] and a recurrent neural network with DMD-based weights [30]. However, none of the ONNs mentioned previously involve convolutional layers, which have become essential in computer vision applications. The ASP Vision system approaches the task of designing a hybrid optoelectronic CNN, using angle sensitive pixels to approximate the first convolutional layer of a typical CNN, but it is limited to a fixed set of convolution kernels [31]. Our goal is to design a system with optimizable optical elements to demonstrate low-power inference by a custom optical or optoelectronic CNN.

3 ONN Toolbox, pp. 5-7

In this section we present building blocks for an optically implemented CNN. In this rest of this paper we focus on the convolutional layer, but we will briefly discuss other layers here too.

- Convolutional layer: Small kernel, tiled kernels, single large kernel (Fig. 1)
- Other layers that we didnt test in detail:
  Optical nonlinearities
  Max/avg pooling with spectral pooling
  Fully connected layers

4 Training, pp. 7

Here we talk about how we train the ONN offline in Tensorflow. PSF optimization followed by phase mask optimization OR end-to-end phase mask optimization.

5 Simulations, pp. 8-10

We use simulations to better understand the performance of optical CNNs. Fig. 2: diagram of possible ONN models. Table 1: Results.

- Introduce toy classification problem(s?), discuss constraints.
- Learned optical correlator (single optical conv. layer)
  Here we are able to use end-to-end learning.
  Possible figure with learned phase mask and PSF.

[1] Fathom Computing (`fathomcomputing.com`), Lightelligence (`lightelligence.ai`), Optalysys (`optalysys.com`)

- Hybrid optoelectronic (one optical conv. layer)
  Talk about the dual channel positive and negative weights
  Grayscale
  With color filters Vincent?
  Fig. 3: Hybrid ONN phase masks and PSFs.
- All-optical convolutional neural network
  Doesnt fully work, but can discuss some results

6 Optical Prototype, pp. 10-11

Implement the hybrid optoelectronic two-layer neural network. Goal is to show that the hybrid ONN can perform on par with the electronic ONN, with the same number of layers, and better than the electronic ONN with one fewer layer.
Fig. 4: optical setup
Fig. 5: actual PSF and sample images
Table. 2: results

7 Discussion, pp. 12-13

- Not straightforward to generalize first optical conv. layer to multiple optical layers
- Discuss importance of negative weights Vincent?
- Instead of trying to replicate a CNN exactly, could take advantage of optical transformations that aren't as practical in computations. For example, we use a 4f system for convolution, but this requires two extra lenses. Perhaps a single custom learned optical element can be used instead.
- In the future, exploit other properties of light (polarization, phase)
- Specifically, coherent light and holography, photonics

8 Conclusion, p. 14

Important step towards optical CNNs. We hope this will inspire more research in the area.

References

1. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553) (2015) 436–444
2. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. (2014) 2672–2680
4. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2015) 3431–3440
5. Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al.: Mastering the game of go with deep neural networks and tree search. nature **529**(7587) (2016) 484–489
6. LiKamWa, R., Priyantha, B., Philipose, M., Zhong, L., Bahl, P.: Energy characterization and optimization of image sensing toward continuous mobile vision. In: Proceeding of the 11th annual international conference on Mobile systems, applications, and services, ACM (2013) 69–82
7. Yang, L., Zhang, L., Ji, R.: On-chip optical matrix-vector multiplier for parallel computation. In: SPIE. Volume 10. (2013) 004932
8. Goodman, J.: Introduction to fourier optics. (2008)
9. Gibbs, H.: Optical bistability: controlling light with light. Elsevier (2012)
10. Christodoulides, D.N., Khoo, I.C., Salamo, G.J., Stegeman, G.I., Van Stryland, E.W.: Nonlinear refraction and absorption: mechanisms and magnitudes. Advances in Optics and Photonics **2**(1) (2010) 60–200
11. Sun, J., Timurdogan, E., Yaacobi, A., Hosseini, E.S., Watts, M.R.: Large-scale nanophotonic phased array. Nature **493**(7431) (2013) 195–199
12. Rechtsman, M.C., Zeuner, J.M., Plotnik, Y., Lumer, Y., Segev, M., Szameit, A.: Photonic floquet topological insulators. In: Lasers and Electro-Optics (CLEO), 2013 Conference on, IEEE (2013) 1–2
13. Shen, Y., Harris, N.C., Skirlo, S., Englund, D., Soljačić, M.: Deep learning with coherent nanophotonic circuits. In: Photonics Society Summer Topical Meeting Series (SUM), 2017 IEEE, IEEE (2017) 189–190
14. Han, S., Mao, H., Dally, W.J.: Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. arXiv preprint arXiv:1510.00149 (2015)
15. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360 (2016)
16. Gruev, V., Etienne-Cummings, R.: Implementation of steerable spatiotemporal image filters on the focal plane. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing **49**(4) (2002) 233–244
17. LiKamWa, R., Hou, Y., Gao, J., Polansky, M., Zhong, L.: Redeye: analog convnet image sensor architecture for continuous mobile vision. In: ACM SIGARCH Computer Architecture News. Volume 44., IEEE Press (2016) 255–266
18. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Computer Science Technical Report CSTR **2**(11) (2005) 1–11

19. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. ACM transactions on graphics (TOG) **26**(3) (2007) 70
20. McGuire, M., Matusik, W., Pfister, H., Chen, B., Hughes, J.F., Nayar, S.K.: Optical splitting trees for high-precision monocular imaging. IEEE Computer Graphics and Applications **27**(2) (2007)
21. O'Toole, M., Kutulakos, K.N.: Optical computing for fast light transport analysis. ACM Trans. Graph. **29**(6) (2010) 164–1
22. Chang, J., Kauvar, I., Hu, X., Wetzstein, G.: Variable aperture light field photography: overcoming the diffraction-limited spatio-angular resolution tradeoff. In: Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on, IEEE (2016) 3737–3745
23. Manzur, T., Zeller, J., Serati, S.: Optical correlator based target detection, recognition, classification, and tracking. Applied optics **51**(21) (2012) 4976–4983
24. Javidi, B., Li, J., Tang, Q.: Optical implementation of neural networks for face recognition by the use of nonlinear joint transform correlators. Applied optics **34**(20) (1995) 3950–3962
25. Farhat, N.H., Psaltis, D., Prata, A., Paek, E.: Optical implementation of the hopfield model. Applied optics **24**(10) (1985) 1469–1475
26. Psaltis, D., Brady, D., Wagner, K.: Adaptive optical networks using photorefractive crystals. Applied Optics **27**(9) (1988) 1752–1759
27. Lu, T., Wu, S., Xu, X., Francis, T.: Two-dimensional programmable optical neural network. Applied optics **28**(22) (1989) 4908–4913
28. Saxena, I., Fiesler, E.: Adaptive multilayer optical neural network with optical thresholding. Optical Engineering **34**(8) (1995) 2435–2440
29. Denz, C.: Optical neural networks. Springer Science & Business Media (2013)
30. Bueno, J., Maktoobi, S., Froehly, L., Fischer, I., Jacquot, M., Larger, L., Brunner, D.: Reinforcement learning in a large scale photonic recurrent neural network. arXiv preprint arXiv:1711.05133 (2017)
31. Chen, H.G., Jayasuriya, S., Yang, J., Stephen, J., Sivaramakrishnan, S., Veeraraghavan, A., Molnar, A.: Asp vision: Optically computing the first layer of convolutional neural networks using angle sensitive pixels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 903–912