Exploring YOLO Inference using Digital–Analog Hybrid Photonic Processor

Xiansong Meng[(1,*)], Kwangwoong Kim[(2)], Po Dong[(3)], Hao Hu[(1,*)], and Deming Kong[(1,*)]

[(1)] DTU Electro, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
* {xiame, huhao, dmkon}@dtu.dk
[(2)] Nokia Bell Labs, New Providence, NJ 07974, USA
[(3)] II-VI Incorporated, 48800 Milmont Dr., Fremont, CA 94538, USA

Abstract We explore application of a high-precision digital-analog hybrid optical MVM processor in a real-world pretrained YOLO NN model and achieve high confidence scores in object detection replacing the first convolution layer with pre-trained weights, while simulation shows analog-only optical computing fails under realistic signal-to-noise ratios. ©2024 The Author(s)

Introduction

Photonic computing offers a promising alternative to traditional microelectronics beyond Moore's law limits[1]. In optical neural networks (ONNs), the core computation relies on matrix-vector multiplications (MVMs), typically implemented via analog methods where data is carried by continuous variables such as light intensity[2]–[4]. Although these systems provide high speed and energy efficiency, their precision is fundamentally limited, often to 4-bit resolution[5],[6]. Integration with electronics also requires high-speed, high-resolution digital-to-analogue conversions (DACs) and analogue-to-digital conversions (ADCs), increasing system complexity and power consumption. These issues mainly arise from the limited signal-to-noise ratio (SNR) inherent in analog signal representation. To address these limitations, we have recently proposed a digital-analog hybrid MVM processor for optical neural networks and shown its high numerical precision[7]. However, the image processing and handwritten digit recognition tasks in the work were relatively simple and considered as toy models that can achieve high accuracy even with low SNRs.

In this paper, we explore the practical application of our hybrid optical processor (HOP) processor in a real-world 106-layer You Look Only Once (YOLO) neural network model to demonstrate the importance of, and evaluate the benefits of, high numerical precision and strong noise tolerance in an optical MVM processor. We run the first layer of the YOLO model with 32 convolution kernels using our HOP processor in an experiment, and achieve an average pixel error rate (PER) of 2.8×10^{-2} for all 32 feature images at an SNR of 14 dB, which eventually leads to high confidence scores in vehicle detections. Complementary simulations show analog optical MVM would degrade the performance, especially for detecting distant objects, due to limited numerical precision. If all convolution layers are replaced with a reasonable achievable SNR range, the analog optical processing would fail completely even with the detection of nearby objects; while our HOP processor can maintain a high confidence score. To the best of our knowledge, YOLO model has not yet been attempted by optical MVM processors.

The architecture of the HOP scheme

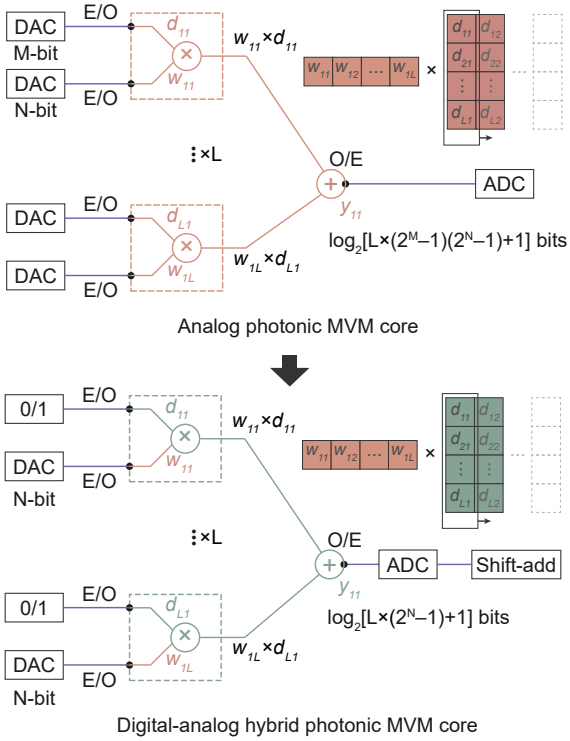

Fig. 1: Photonic MVM architecture comparison, the analog and the digital-analog hybrid cores. The conventional analog photonic MVM core utilizes analog signals for both the input d and the weight w. In contrast, the proposed digital-analog hybrid photonic MVM core employs digital signals for the input d, thereby relaxing the requirements on signal format converters.

Analog optical matrix multiplication uses photonic multiplication cores, as illustrated in Fig.1.

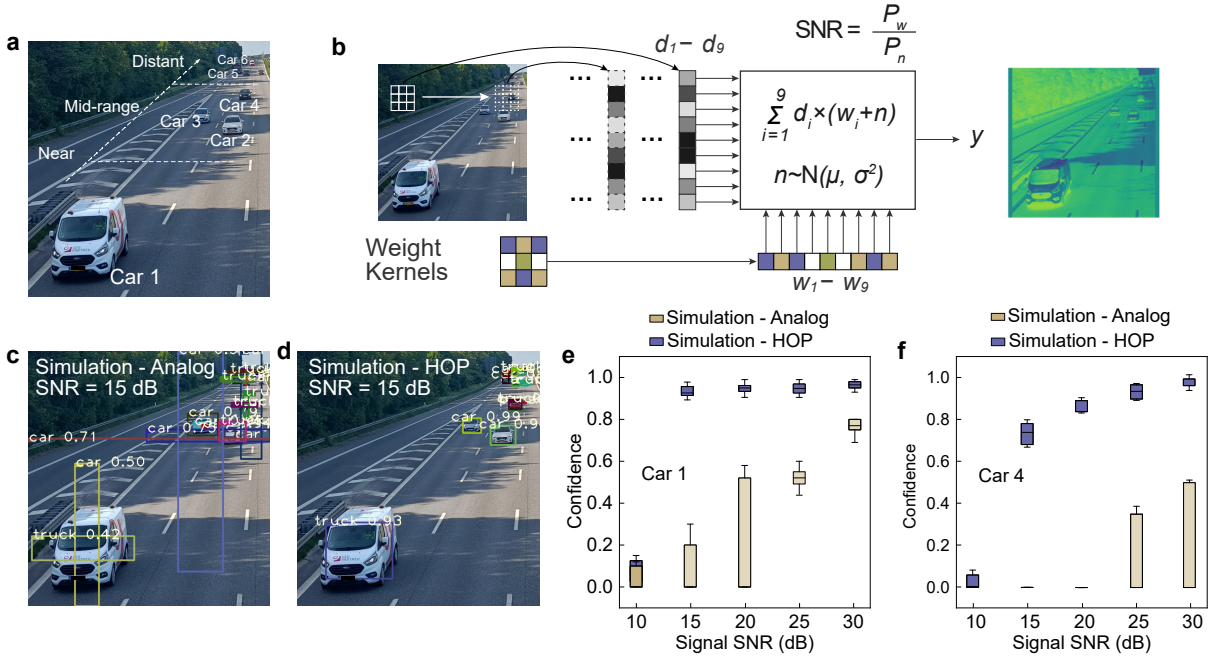

Fig. 2: YOLO object detection under simulation using the HOP scheme. a, Input image containing labeled vehicles arranged from nearest to farthest. **b,** Simulation setup. An image is convolved with the weight kernels for edge detection. To evaluate noise tolerance, additive white Gaussian noise is applied to the weights, and the output distributions are analyzed for both analog and hybrid photonic computing systems.**c,d,** Simulated detection results at an SNR of 15 dB using the analog and HOP schemes, respectively. **e,f,** Comparison of detection accuracy under various SNRs for the nearby vehicle (Car 1) and a distant vehicle (Car 4) using both schemes.

The inputs d and w are generated by M-bit and N-bit DACs, then converted into optical signals, multiplied, and converted back to the electrical domain. To accomplish a vector-vector multiplication with L elements, the overall number of signal levels becomes $L \times (2^M - 1) \times (2^N - 1)$, which requires an ADC resolution of $\log_2 \left[L \times (2^M - 1) \times (2^N - 1) \right]$ if full-precision is to be preserved for the calculation. On the other hand, the proposed HOP scheme uses digital optical inputs in binary words, while the weights remain analog. Consequently, as shown in Fig. 1, the overall number of signal levels for the same vector-vector multiplication becomes $L \times (2^N - 1)$, which reduces the ADC resolution requirement to $\log_2 \left[L \times (2^N - 1) \right]$ bits.

In addition, the HOP processor removes the need for high-speed DACs for the input **D**, thus lowers the cost, reduces power consumption, and improves the compatibility with microelectronics. From an encoding perspective, analog ONNs rely solely on amplitude, while our HOP processor uses both amplitude and time. This extended encoding improves the Euclidean distance between signal states, making the system more robust to noise and improving numerical precision.

The application with real-world YOLO model

To evaluate the feasibility of the HOP and its relevance to real-world tasks, we applied it to a YOLO v3 object detection model. This network contains 106 convolutional layers with nonlinear activations. In our implementation, we focus on the linear convolution parts, while activations are computed digitally.

We first conducted simulations by introducing additive white Gaussian noise (AWGN) to the weights of each convolutional layer, following the procedure outlined in Fig. 2b. Ten trials were run for each SNR value. The test image (size 416×416), captured by a smartphone, includes six vehicles labeled Car 1 to Car 6, representing varying distances: Car 1 as nearby, Cars 2–4 as mid-range, and Cars 5–6 as distant as shown in Fig. 2a. The YOLO model outputs object locations with associated confidence scores.

Simulation results in Fig. 2c and d at 15 dB SNR reveal that the analog scheme fails to detect some nearby and distant cars accurately, whereas the HOP identifies all six vehicles with high confidence. To analyze the effect of SNR, we examine the detection accuracy of Car 1 under different noise levels in Fig. 2e. As SNR increases, both schemes show improvement, but the HOP consistently yields higher confidence scores, especially in low SNR conditions. The analog method also exhibits larger performance fluctuations. Similar observations hold for Car 4 in Fig. 2f, a distant vehicle. When SNR exceeds 15 dB, HOP maintains over 70% confidence, while the analog system performs poorly below 20 dB and becomes unstable between 25–30 dB.

These findings demonstrate the need for sufficient numerical precision when applying MVM in real-world neural networks. In autonomous driving contexts, higher detection accuracy for distant vehicles could translate into faster reaction time and

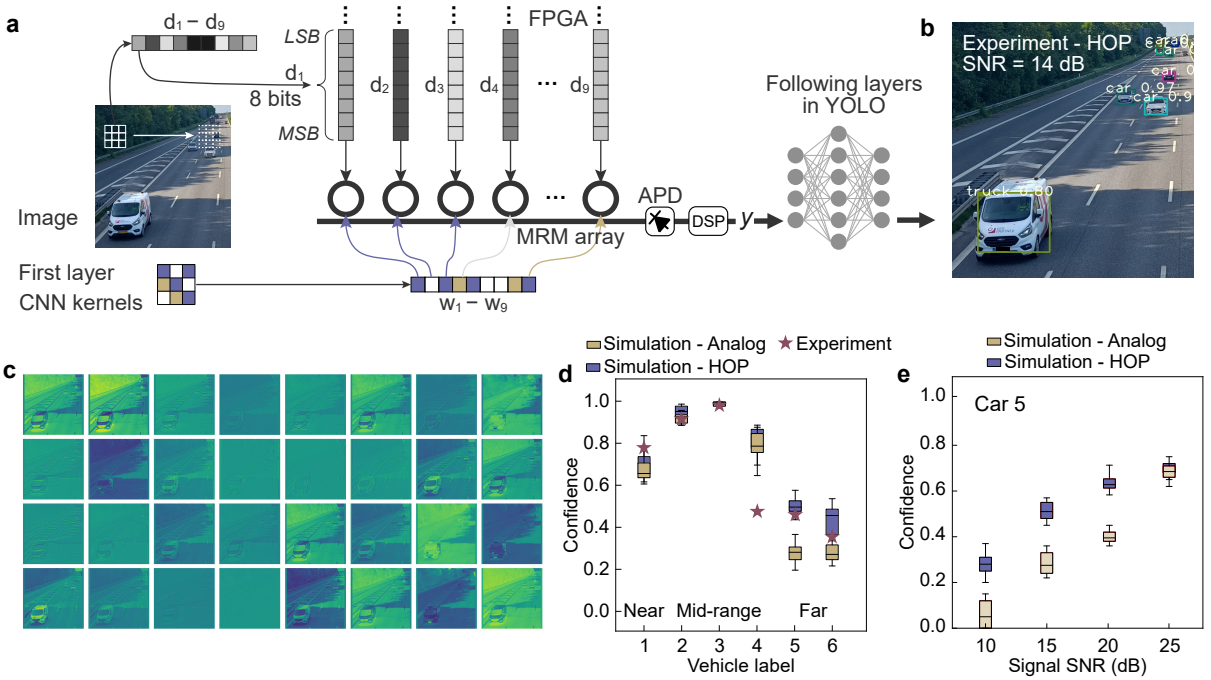

Fig. 3: YOLO object detection task with experiment implementation. a, Experiment setup. The inputs are loaded as binary words into the high-speed modulation ports of nine MRMs. The convolution kernel is reshaped into a 1×9 vector and applied via microheater-based bias control to the same MRM array. The output is sampled by the avalanche photodiode (APD) and processed using digital signal processing to obtain the convolution result y, which is then fed into the subsequent layer of the YOLO model to compute the final confidence score. **b,** Detection output from the HOP system, where the first convolutional layer is executed experimentally and the remaining layers are computed offline. **c,** Reconstructed feature maps from the experimentally implemented first layer. **d,** Confidence score comparison for Cars 1–6 between simulation and experiment, where the first layer is replaced with either the analog or HOP scheme. **e,** Simulated confidence scores for Car 5 under a range of SNRs.

improved safety, highlighting the potential benefits of the HOP scheme.

We also experimentally validated the system using our photonic integrated circuit (PIC) chip. Due to hardware limitations, only the first 32 convolution operations of the YOLO network were implemented optically, with the rest handled offline as shown in Fig. 3a. The data is input through the FPGA board's GPIO port at a rate of 300 MHz. The output is sampled by the APD and processed using digital signal processing to obtain the convolution result y, which is then fed into the subsequent layer of the YOLO model to compute the final confidence score. The experiment achieved 14 dB SNR and an average PER of 2.8×10^{-2}. With 8-bit inputs, weights, and outputs from our processor, the low PER confirms the high numerical precision of our HOP processor. All vehicles were correctly detected in Fig. 3b. The reconstructed feature in Fig. 3c maps from the first layer reveal distinct object boundaries.

Additionally, we performed a simulation replacing only the first convolutional layer with either the HOP or analog scheme shown in Fig. 3d. For nearby and mid-range objects, both schemes performed similarly. However, for distant objects like Car 5 and Car 6, HOP showed clear advantage: for Car 5, confidence reached 49% in simulation and 46% in experiment, compared to only 28% using analog. We further analyzed how SNR affects Car 5 detection shown in Fig. 3e. The HOP maintained better performance between 10–20 dB SNR, confirming its robustness under noise.

In conclusion, even replacing only the first layer with realistic HOP hardware leads to better final detection confidence for distant targets, reaffirming the critical role of numerical precision in optical neural computation.

Conclusion

We have explored the application of a digital-analog hybrid optical MVM processor to demonstrate the importance and necessity of sufficient numerical precision and noise tolerance in real-world neural network models. We have experimentally demonstrated high confidence scores in object detection in the YOLO model, by running its first convolution layer on our processor. We have also compared the performance between our HOP processor and analog optical MVM processors in complementary simulations, where the analog optical processor shows significant performance degradation if the first layer is run, or completely fail if all layers are replaced. This confirms the benefit of sufficient numerical precision and noise tolerance provided by our HOP processor, making it potentially more feasible for real-world neural network inference tasks.

Acknowledgements

D.K. acknowledges the research grant (VIL53077) of the Young Investigator Program (DONN) from the VILLUM FONDEN, and the EIC Pathfinder project DOLORES (101130178) from the European Innovation Council (EIC). H.H. acknowledges the research grant (VIL34320) of the Synergy Program (Searchlight) from the VILLUM FONDEN.

References

[1] M. A. Nahmias, T. F. De Lima, A. N. Tait, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Photonic multiply-accumulate operations for neural networks", *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–18, 2019.

[2] Y. Shen, N. C. Harris, S. Skirlo, *et al.*, "Deep learning with coherent nanophotonic circuits", *Nature photonics*, vol. 11, no. 7, pp. 441–446, 2017.

[3] X. Xu, M. Tan, B. Corcoran, *et al.*, "11 tops photonic convolutional accelerator for optical neural networks", *Nature*, vol. 589, no. 7840, pp. 44–51, 2021.

[4] J. Feldmann, N. Youngblood, M. Karpov, *et al.*, "Parallel convolutional processing using an integrated photonic tensor core", *Nature*, vol. 589, no. 7840, pp. 52–58, 2021, ISSN: 0028-0836. DOI: 10.1038/s41586-020-03070-1.

[5] A. N. Tait, T. F. De Lima, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Multi-channel control for microring weight banks", *Optics Express*, vol. 24, no. 8, pp. 8895–8906, 2016.

[6] A. N. Tait, T. F. De Lima, E. Zhou, *et al.*, "Neuromorphic photonic networks using silicon photonic weight banks", *Scientific reports*, vol. 7, no. 1, pp. 1–10, 2017.

[7] X. Meng, D. Kong, K. Kim, *et al.*, "Digital-analog hybrid matrix multiplication processor for optical neural networks", *arXiv preprint arXiv:2401.15061*, 2024.