Evaluation of hardware and model architectures for 3D CNNs on hyperspectral data

Madeline Loui Anderson, Hyewon Je ong, Joanna Kondylis, Ferdi Kossmann

I. Brief overview of project

Hyper-spectral images (HSI) are images with tens to hundreds of channels. Neural network architectures designed for HSIs often make use of 3D convolutions that slide over all three dimensions of the image tensor. In this project, we evaluate the energy-accuracy tradeoff of architectures using 3D convolutions versus ones only using 2D convolutions (i.e. project category 3: New workload of 3D convolutions). We further analyze if off-the-shelf hardware architectures can be modified to perform 3D convolutions using less energy (i.e. project category 5: Sweep design parameters of previously proposed architecture).

II. Motivation

The information captured by each pixel in an RGB image is limited to the primary colors red, green, and blue. Hyperspectral images (HSI) address this limitation by letting each pixel capture information from across the electromagnetic spectrum. By capturing more information than just RGB values, HSIs can allow for more accurate predictions in tasks such as classification, object detection, anomaly detection, and image segmentation. This has driven recent work on developing deep learning models for HSI analysis [1], [4], [13] and has led to wide adoption of HSIs in commercial, industrial, and military applications [8], including land-cover detection, agricultural development, environmental protection and urban planning.

However, HSI analysis comes with its own set of challenges. In contrast to RGB images which consist of three channels, HSIs typically consist of tens to hundreds of channels. This is problematic since HSI models are often deployed in contexts with tight energy constraints, such as solar-powered satellites with onboard compute [3]. In such deployments, the energy consumption of deep neural networks with large convolution kernels and/or high-dimensional fully connected layers is prohibitively high.

To avoid large convolution kernels, HSI models often use 3D convolutions [5], [7], [9], where the convolution kernel not only slides over the x and y dimensions of the image but also over the channel dimension. In order to gain a better understanding on how 3D convolutions can be used to build efficient HSI models, we evaluate them in two aspects: First, we study the accuracy-energy tradeoff of different neural network architectures on off-the-shelf hardware. Our analysis hereby includes architectures that use 3D convolutions and others that only use 2D convolutions. Second, we study how the energy consumption of 3D convolutions changes for different hardware architectures (i.e. when modifying the original off-the-shelf architecture). If we are successful, this work will show how HSI analysis can be made more efficient through building models with 3D convolutions and which hardware design parameters allow for especially efficient execution of 3D convolutions.

III. Technical Contributions

Several works have investigated the computational efficiency and power consumption of HSI learning models on different hardware architectures. The authors of [6] conduct an experimental analysis of performance, cost, latency, and energy consumption for different layers of a 3D CNN using the NVIDIA Jetson Tegra TX2. Another study investigates the energy consumption for the standard vector machine (SVM), multinomial logistic regression (MLR) and random forest (RF) algorithms on the ODROID-XU4, a heterogeneous computing device [12]. [10] proposes a hardware/software implementation for compressive sensing on a system-on-chip (SoC) field-programmable gate array (FPGA) and is shown to operate at competitive speeds with greatly reduced energy consumption. While such studies provide useful insights into various aspects of HSI analysis, most related works focus on a specific piece of hardware and algorithms such as compression and basic machine learning models. Because of the recent growth of deep learning and the need for energy-efficient processing for onboard/remote sensing applications, our approach is to explore the computational performance trade-offs of deep neural networks for HSI data, particularly 2D and 3D CNNs. This work contributes the following: (1) evaluate the energy-accuracy tradeoff of different 2D and 3D CNN workloads for HSI data, (2) explore the energy, latency, and throughput of 2D vs. 3D convolutions using different hardware designs using the Timeloop and Accelergy tools [11]. Together, these analyses aim to provide insight on design choices when implementing CNN models for HSI tasks in the context of hardware constraints.

IV. Evaluation

A. Dataset

We will be using the AVIRIS Indian Pines Data [2]. This dataset was gathered by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in North-western Indiana in June 1992. It contains 220 spectral reflectance bands in the wavelength range 0.4-2.5 μm. This HSI has a spatial dimension of 145x145, a spectral resolution of 10 nm, and a spatial resolution of 20 m by pixel. The ground truth contains 16 land cover classes.

B. Experiments

We explore the energy-accuracy tradeoff and energy efficiency of 3D CNN with the input hyperspectral dataset. We use the accelerator evaluation tools Timeloop and Accelergy [11].

a) Energy-accuracy benefits of 3D CNNs: We explore different model architectures with respect to their energy per inference and their accuracy. We hereby evaluate architectures with 3D convolutions and replace the 3D convolutions through 2D convolutions to have a baseline comparison against 2D CNNs. This allows us to determine whether 3D CNNs can achieve a superior energy-accuracy tradeoff over 2D CNNs on the dataset described in subsection IV-A. We analyze and report the accuracy, energy, cycles, and latency per inference, and the used area of the accelerator.

b) Evaluation of Energy on different hardware architectures: In order to further optimize 3D CNNs, we investigate whether the energy consumption of 3D convolutions can further be reduced through a more specialized hardware architecture, including Eyeriss or NVDLA style architectures. While the experiments in (a) assume fixed hardware properties (e.g. dimensions of PE array, number of PEs, buffer sizes, etc.), we now investigate whether tuning these properties allows for further energy savings. We compare the pattern of energy use and cycles for each component and subcomponents. We then use this profiling information to optimize the architecture for energy efficiency. We use the unmodified Eyeriss and NVDILA style architectures as baselines to our modified designs.

V. Timeline

- **4/10-4/20**: Write helpers to analyze 3D CNNs and 2D CNNs with Timeloop/Accelergy on hardware spec provided by the class. *(Madeline, Hyewon, Joanna, Ferdi)*
- **5/21-4/30**: In parallel:
  1) Evaluate different 3D CNNs and 2D CNNs architectures for accuracy, energy per inference and time per inference on hardware spec provided by the class. *(Joanna, Ferdi)*
  2) Evaluate different hardware architectures for energy per inference, area and time per inference. *(Madeline, Hyewon)*
- **5/1-5/8**: Write project paper and prepare presentation. *(Madeline, Hyewon, Joanna, Ferdi)*

References

[1] N. Audebert, B. Le Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 159–173, June 2019.

[2] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe, "220 band aviris hyperspectral image data set: June 12, 1992 indian pine test site 3," Sep 2015. [Online]. Available: https://purr.purdue.edu/publications/1947/1

[3] S. S. C and G. R, "Onboard target detection in hyperspectral image based on deep learning with fpga implementation," *Microprocessors and Microsystems*, vol. 85, p. 104313, 2021.

[4] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2094–2107, 2014.

[5] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3d-2d convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7570–7588, 2021.

[6] J. M. Haut, S. Bernabé, M. E. Paoletti, R. Fernandez-Beltran, A. Plaza, and J. Plaza, "Low–high-power consumption architectures for deep-learning models applied to hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 776–780, 2018.

[7] D.-H. Jung, J. D. Kim, H.-Y. Kim, T. S. Lee, H. S. Kim, and S. H. Park, "A hyperspectral data 3D convolutional neural network classification model for diagnosis of gray mold disease in strawberry leaves," *Front Plant Sci*, vol. 13, p. 837020, Mar. 2022.

[8] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14 118–14 129, 2018.

[9] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 10, pp. 8693–8703, 2021.

[10] J. M. P. Nascimento, M. P. Véstias, and G. Martín, "Hyperspectral compressive sensing with a system-on-chip fpga," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 3701–3710, 2020.

[11] A. Parashar, P. Raina, Y. S. Shao, Y.-H. Chen, V. A. Ying, A. Mukkara, R. Venkatesan, B. Khailany, S. W. Keckler, and J. Emer, "Timeloop: A systematic approach to dnn accelerator evaluation," in *2019 IEEE international symposium on performance analysis of systems and software (ISPASS)*. IEEE, 2019, pp. 304–315.

[12] N. Pereira, J. Plaza, J. M. Haut, and A. Plaza, "On the evaluation of machine learning algorithms for hyperspectral image classification on a heterogeneous computing device," in *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, 2021, pp. 1–6.

[13] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *Journal of Imaging*, vol. 5, no. 5, 2019. [Online]. Available: https://www.mdpi.com/2313-433X/5/5/52