Low–High-Power Consumption Architectures for Deep-Learning Models Applied to Hyperspectral Image Classification

Juan M. Haut, Student Member, IEEE, Sergio Bernabé, Mercedes E. Paoletti, Student Member, IEEE, Ruben Fernandez-Beltran, Antonio Plaza, Fellow, IEEE, and Javier Plaza, Senior Member, IEEE

Abstract—Convolutional neural networks have emerged as an excellent tool for remotely sensed hyperspectral image (HSI) classification. Nonetheless, the high computational complexity and energy requirements of these models typically limit their application in on-board remote sensing scenarios. In this context, low-power consumption architectures are promising platforms that may provide acceptable on-board computing capabilities to achieve satisfactory classification results with reduced energy demand. For instance, the new NVIDIA Jetson Tegra TX2 device is an efficient solution for on-board processing applications using deep-learning (DL) approaches. So far, very few efforts have been devoted to exploiting this or other similar computing platforms in on-board remote sensing procedures. This letter explores the use of low-power consumption architectures and DL algorithms for HSI classification. The conducted experimental study reveals that the NVIDIA Jetson Tegra TX2 device offers a good choice in terms of performance, cost, and energy consumption for on-board HSI classification tasks.

Index Terms—Deep learning (DL), embedded computing, hyperspectral image (HSI) classification, low-power consumption architectures.

I. INTRODUCTION

The use of miniaturized satellites (SmallSats) is becoming an increasingly popular trend in many of the existing Earth observation programs [1], allowing for a substantial reduction of financial costs and hardware complexity [2]. As a result, this technology has been successfully employed in a wide range of remote sensing applications, such as monitoring of the atmosphere, land-cover categorization or mapping of urban areas, and the Earth surface [3]. Nonetheless, the increasing demand for extended computing capabilities able to deal with new applications has introduced the need to seek for architectures able to not only increase computing capacity but also to reduce energy consumption. These requirements may eventually constrain the use of these small devices under highly demanding scenarios, such as the use of deep-learning (DL) techniques for the classification of hyperspectral image (HSI) data [4], [5].

Broadly speaking, HSI collects hundreds of narrow spectral bands in order to simultaneously provide detailed spectral and spatial information, which makes this data especially useful to accurately identify different materials [6], [7]. Many approaches have been proposed to perform HSI classification; however, the intrinsic complexity of the HSI domain leads to the fact that only the most advanced convolutional neural networks (CNNs) are able to consistently provide satisfactory results on different remote sensing applications [4], [8]. Furthermore, the selection of efficient computing platform is another critical aspect to take into account, especially when dealing with highly demanding methodologies from a computational point of view. Even though some novel methods pursue to reduce the number of training samples in order to obtain robust classifiers [9], these approaches usually result in computationally demanding models with limited practical application in constrained hardware environments.

On the one hand, commodity clusters [10] and graphic processing unit (GPU) platforms [11] have been traditionally used to process HSIs, but those systems are hardly adaptable to on-board processing requirements which generally introduce strong constraints in terms of energy consumption. On the other hand, field-programmable gate array (FPGA) devices [12] offer a good compromise between performance and energy consumption, but they generally require a significant effort from the design and programmability point of view, which may eventually limit their practical application. In this sense, an attractive alternative is the Tegra GPU architecture which, in the last years, has dominated mobile platforms and embedded devices as the Internet of Things. High rated MPixel/s/Watio, less heat, and less space are important keys when facing the on-board processing challenge.

Traditionally, space electronic systems have been highly customized based on the FPGA approach; however, the Tegra architecture is able to provide remarkably higher flexibility...
while becoming more scalable, affordable, and reliable. Even though there are many on-board computing tasks in which Tegra devices may be suitable to process and manage HSI data, it is still necessary to conduct additional research to fully test this hardware applicability. Although there is a handful of jobs that embeds HSI-processing algorithms in efficient platforms [13]–[15], there are very few jobs focused on adapting DL models for remote HSI processing using similar architectures. For instance, Randhe et al. [16] propose to integrate an HSI-CNN model (implemented with Caffe framework) into a Jetson TK1 application, reducing the complexity via principal component analysis (PCA). However, no implementation details are provided. In this sense, it must be highlighted that although some projects developed by Air Force and NASA experts aim at designing radiation-hardened Tegra hardware for on-board purposes, there are very few research works in the literature aimed at testing the actual performance capability to process HSI data using the most recent DL models for on-board exploitation.

This letter deeply explores, for the first time within the remote sensing research community, the use of DL algorithms over the new NVIDIA Jetson Tegra TX2 low-energy consumption architecture by conducting a comparative study of low–high-power consumption hardware applied to HSI classification tasks. The most recent Earth observation programs work for providing high processing level products which require an increasing demand of ground-segment hardware resources [1]. As a result, studying new alternatives to relieve this workload via on-board low-consumption devices is an interesting option to alleviate ground-segment HSI data computations. In this regard, the target of this letter is based on shedding light on the use of the new NVIDIA Jetson Tegra TX2 device for on-board HSI classification when it is compared to other popular hardware alternatives available in regular ground-segment processing units, such as Intel Xeon and NVIDIA GeForce GTX devices. Initially, Sections II and III describe the considered low–high-power consumption architectures as well as the DL-based HSI classification models. Then, Section IV presents the experimental comparison and highlights the most interesting results. Finally, Section V provides interesting conclusions concerning the energy consumption-based viability of moving specific HSI data computations from the ground-segment resources to on-board platforms via the NVIDIA Jetson Tegra TX2 device.

II. HIGH-VERSUS LOW-POWER CONSUMPTION ARCHITECTURES

Leading manufacturers of high-performance computing platforms, such as NVIDIA, launched the Jetson Tegra TX1 device in 2015 as a low-power consumption device. This platform was one of the first supercomputers built on a module carrying a Tegra processor from NVIDIA and incorporating an ARM processor. In 2017, NVIDIA announced the new Jetson Tegra TX2 as a compact card design for low-power scenarios. This device belongs to the NVIDIA Pascal family and is an embedded system. The chip features 256 Compute Unified Device Architecture (CUDA) cores that are based on the same DNA that is featured on the Titan X (Pascal) GPU. The ARM v8 CPU complex comprises two Denver 2 and four A57 cores with a coherent heterogeneous multiprocessor architecture geared for multithreading.

In contrast, high-power consumption architectures represent now the most widely used choice when power restriction is not necessary. Most of these solutions are based on a workstation featuring a professional Intel Xeon processor in conjunction with one or several NVIDIA GPUs from the Pascal family. Among the main features of the latter is the use of unified memory to solve the limited capacity available on the GPU main memory to process large amounts of data. This mechanism creates a pool of managed memory that is shared between the GPU and the CPU, using a single pointer that is accessible to both the CPU and GPU, bridging the CPU–GPU divide. The data can be read or written from code running on either CPUs or GPUs using calls to `cudaMallocManaged()`. An important aspect is that the Pascal GPU architecture is the first one with hardware support for virtual memory page faulting and migration, via its page migration engine.

In this letter, the Jetson Tegra TX2 device (referred to hereinafter as Jetson) is compared against a professional heterogeneous platform (Intel Xeon processor equipped with a GPU NVIDIA GeForce GTX 1080 and referred to hereinafter as Xeon) focused on a detailed comparative study in performance and energy consumption terms. To the best of our knowledge, this kind of analysis has not been previously conducted in the HSI processing literature using DL models for on-board exploitation, and in our opinion, it is very important in order to really calibrate the possibility of using low-power consumption platforms for efficient HSI processing in real remote sensing missions.

From a hardware point of view, the main differences between the considered devices are based on the number of CUDA processing cores, memory configuration, and thermal design power (TDP). Specifically, the Xeon environment offers over 10 times more CUDA cores and streaming multiprocessors than the Jetson device. Regarding the memory configuration, we can find some major differences between both the professional (GDDR5X) and Jetson (LPDDR4) platforms. The Xeon platform exhibits higher bandwidth (over 4×) and lower voltage. The 16-nm fin field-effect transistor technology allows to explore new horizons for discrete memory I/O data rates, from an initial rate between 10 and 12 Gb/s to a potential up to 16 Gb/s. Moreover, it is possible to reduce the latency gap between the local memory and shared internal/external memory through cache prefetching. In the considered professional platform, this technique allows 64B data per memory access to boost execution performance by fetching instructions or data from their original storage in slower memory to a faster local memory before it is actually needed. However, LPDDR4 memory is able to achieve lower memory I/O data rates (between 3.20 and 4.27 Gb/s) allowing cache prefetching to 16B. In this way, the power consumption is reduced by lowering the supply voltage (1.1 V) and maintaining an acceptable bandwidth.

Last but not least, power consumption is another important restriction to be considered in on-board processing. In this case, the Jetson device presents two performance
which is an ARM GPU environment composed by a dual-core NVIDIA Denver2 at 2.00 GHz together with a quad-core ARM Cortex-A57 at 2.00 GHz, 8-GB 128-bit LPDDR4, and integrated 256-core Pascal GPU at 1300 MHz; and 2) the Xeon (multicore heterogeneous system), which is a 2×Intel Xeon E5-2695v3 processors with 14 cores each, running at 2.30 GHz, and 64 GB of DDR3 RAM memory. An NVIDIA GeForce GTX 1080 GPU with 2560 CUDA cores operating at 1772 MHz and dedicated memory of 8 GB.

Regarding the considered software environment, it consists of Debian GNU/Linux 9 and Ubuntu 16.04 as operating systems for both NVIDIA Jetson TX2 and multicore heterogeneous systems, respectively. Tensorflow 1.7 and CUDA 8 for GPU functionality.

B. Results

Table II presents the results of our CNN-based classification experiments, conducted on the IP and UP data sets using the two considered hardware environments. In columns, we show the considered input patch size (i.e., 19, 29, and 39), the percentage of training data (i.e., 5%, 10%, and 15%), and the corresponding overall accuracy (%) as well as the average energy consumption (Wh) and computational time (s) for Xeon and Jetson environments.

According to the reported quantitative results, it is possible to highlight some important observations. Regarding the classification accuracy, the two considered hardware environments exhibit a similar overall performance. Even though the Xeon environment provides a slightly better average overall accuracy than the Jetson one (+0.007%), the differences between both hardware architectures are always under the standard deviation values, which indicates that these small variations are not statistically relevant and, hence, both environments perform similarly in terms of overall accuracy.

Regarding the energy consumption and processing time metrics, experiments reveal several remarkable differences which deserve to be mentioned. Specifically, the Xeon hardware reports an average energy consumption of 0.4553 Wh whereas the Jetson environment only requires, on average, 0.0452 Wh which makes the former technology 10.06× more energy demanding than the latter one. When considering the computational time, the Xeon and Jetson environments obtain an average computational time of 9.14s and 69.79s, respectively. As a result, the Jetson hardware is 7.63× slower than Xeon, nonetheless, it is also 10.06× more energy efficient, which generates a positive balance of 2.43 in the energy/performance ratio when considering the Jetson environment.

When analyzing the results in more detail, some interesting points about the tested configurations can be highlighted. More specifically, the obtained quantitative metrics show that the amount of training data does not have a relevant effect on the differences between both hardware environments. That is, increasing the number of training samples from 5% to 10% or 15% does not have an important impact on the computational time, because both Xeon and Jetson environments take advantage of their GPU-based architectures to process the input data, that is, NVIDIA GeForce GTX 1080 and Pascal GPU, respectively. However, considering a bigger input
patch size affects the two considered hardware configurations in a different way. On the one hand, the Jetson architecture has fewer and slower CPU cores than the Xeon one which logically introduces an unavoidable processing delay as the networks' parameters increase. Note that the number of parameters that the CNN model requires to adjust substantially increases with the input patch size, being 4,16×7,79×4,16 when using 29×29 and 39×39 sizes, respectively. On the other hand, the Jetson hardware shares the memory between ARM CPU and Pascal GPU units which makes this hardware less efficient than the Xeon one when considering very large input spatial sizes, e.g., 39×39, because of the two specific memories present in the Xeon environment. Regarding the considered batch sizes, a similar trend can be observed because Jetson seems to provide a better energy/performance ratio with respect to Xeon when smaller batch sizes are considered.

Fig. 1 displays the runtime and energy differences between Xeon and Jetson environments considering 15% of training data and a spatial size of 19×19 (first column), 29×29 (second column), and 39×39 (third column). Red bars: runtime improvement provided by Xeon with respect to Jetson (in number of times better). Blue bars: energy savings provided by Jetson with respect to Xeon (also in times better). All experiments have been conducted using the AVIRIS IP data set.

remote sensing HSI classification field, the typical input patch and batch sizes are substantially smaller than the maximum values tested in this letter. For instance, a normal patch size value could be 19×19 with 100 batch size (see [4]). As a result, the Jetson hardware environment is shown to be a highly suitable architecture for on-board remote sensing HSI classification tasks, because the energy savings in the acquisition platform are substantially higher than the runtime increase in the ground-segment unit.

Despite the fact that the Xeon environment has shown to obtain a significantly lower computational time than Jetson hardware, it is important to highlight that the latter environment has a much more reduced power consumption while maintaining the classification accuracy which provides an excellent scenario for on-board remote sensing processing tasks. Fig. 2 shows a detailed graphical comparison between the power consumption of both hardware environments over the IP data set in order to better assess the obtained energy results. As we can see, the Jetson energy consumption (displayed in the first row) is substantially lower than the one corresponding to the Xeon configuration (shown in the second row). Besides, the advantage provided by the Jetson architecture becomes especially relevant when considering relatively small batch and patch sizes because of the aforementioned memory limitation of the NVIDIA Jetson Tegra TX2 hardware. With all these considerations in mind, the Jetson environment has shown to provide a competitive advantage

Fig. 1. Runtime and energy differences between Xeon and Jetson environments considering 15% of training data and a spatial size of 19×19 (first column), 29×29 (second column), and 39×39 (third column). Red bars: runtime improvement provided by Xeon with respect to Jetson (in number of times better). Blue bars: energy savings provided by Jetson with respect to Xeon (also in times better). All experiments have been conducted using the AVIRIS IP data set.
in constrained scenarios where power consumption, physical space, and financial costs are important decision factors. Precisely, this is the case of remote sensing platforms where this kind of hardware can be an optimal choice to relieve the ground-segment computations when classifying HSI data using relatively simple CNN-based architectures with a constrained number of parameters (i.e., two CNN layers with max-pooling, over a $19 \times 19$ input patch size and a batch size up to 100 samples). Consequently, the experimental results and the exhaustive power consumption analysis conducted in this letter reveal the viability of integrating the new NVIDIA Jetson TX2 for on-board remote sensing HSI classification.

V. CONCLUSION

This letter studies the possibility of exploiting the new NVIDIA Jetson Tegra TX2 device for on-board HSI classification in order to relieve ground-segment computations when generating high-level remote sensing products. Our experimental results, conducted using two different hardware environments and two reference HSI data sets, indicate that the Jetson device provides satisfactory energy/performance results for on-board HSI classification when considering constrained CNN-based architectures. Future work will be focused on analyzing other HSI processing algorithms on additional low-power consumption hardware platforms.

ACKNOWLEDGMENT

The authors would like to thank Dr. F. D. Igual for his helpful assistance.

REFERENCES


