Deep Learning-oriented GPUs analysis

Uxio Garcia Andrade

Laboratorio de Fundamentos de Computadores I
Grupo 01
uxio.garcia.andrade@rai.usc.es

06/03/2018

Abstract

Comparative analysis of Graphics Processing Units (GPUs) for a machine learning startup, centered on its deep learning suitability. In order to select which GPU best fits the needs of the company, several cutting edge technologies are subjected to analysis, taking into consideration features such as price, memory bandwidth, processing power or VRAM size. Furthermore, GPUs' performance is tested and contrasted by results from diverse benchmarks, as well as other prestigious studies to back up our conclusions.

Keywords: GPU, deep learning, machine learning, AI, NVIDIA, benchmark

I. Introduction

Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed [1]. One of its most revolutionary subfields is Deep Learning, which is based on learning data representations, as opposed to task-specific algorithms. [2] It has countless applications, being healthcare institutions and self-driving cars companies our best clients.

Nevertheless, the implementation of such disruptive technologies often involves expensive hardware updates. Due to this, our engineering team has been lately complaining about the lack of deep learning-oriented equipment, focusing on the paramount urge of acquiring new cost-efficient GPUs.

To avoid possible merging problems, or running into any compatibility issues, we have decided to primarily evaluate NVIDIA GPUs, as AMD's OpenCL toolkit has some problems when working with major DL frameworks. This will be our most restrictive but necessary constraint.

The rest of this paper is organized as follows. Section 2 discusses how to escrutine GPUs' suitability for DL, section 3 shows the results we have obtained, and section 4 offers conclusions.

II. Methodology

Our research has shown that there are 3 main specifications to focus on when it comes to Deep Learning: memory bandwdith, which is probably the most important one, as it determines the amount of data the GPU is able to handle; processing power, which indicates how fast the GPU can manage data; and VRAM size, although its relevance depends on the kind of work we need to do. Moreover, price and cost-efficiency will also be evaluated, as our budget restrictions play a decisive role when selecting GPUs. These criteria have been derived from the studies done by [3] and [4]

In order to compare GPUs, we have gathered information from diverse benchmarks. The first one [5] considers computational power of the GPUs, by performing general purpose computing and taking advantage of the huge parallel computation capacity of

Table 1: *Specifications*

	1070	1070 Ti	1060	P2000
Bandwidth(GB/s)	**256**	256	192	140
VRAM(GB)	**8**	8	6	5
CUDA Cores	**2432**	1920	1280	1024
Price($)	429	469	**309**	479

GPUs. On the other hand, the second one [6] analyses high-end GPUs, showing an overall rating of expensive hardware.

III. Results

Once applied an initial filter to the NVIDIA GPUs, we decided to focus on the following ones, as they are they only high-end ones to meet our price constraints.

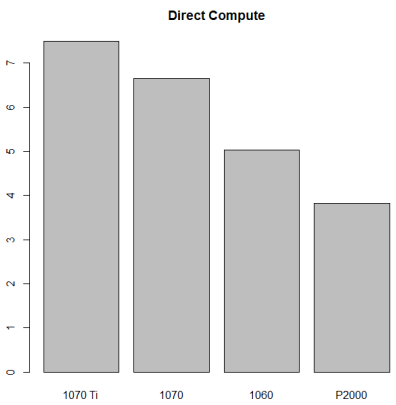
- GeForce GTX 1070
- GeForce GTX 1070 Ti
- GeForce GTX 1060
- Quadro P2000

Now, we will delve into the specifications of each GPU, displayed in Table 1, in which prices have been established according to NVIDIA's official webpage [7]. It can be clearly perceived that the GTX 1070 Ti comes out first in every single aspect, with the single exception of the price, so we will need to persist in our study to determine whether its technical features are worth the price disparity.

A. Benchmarks

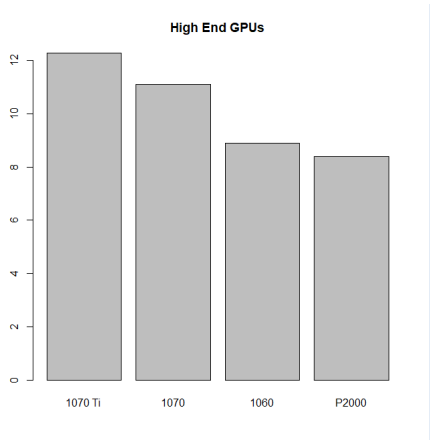
Figure 1 shows a barplot generated with the programming language R, in which the results of the first benchmark can ben appreciated. GTX 1070 Ti is the successful candidate once again. Even if the only considerable difference between the 1070 Ti and the 1070 is the number of CUDA cores, it is a notable and decisive dissimilarity, as it accelerates computations up to a 15%.

Figure 1: *Direct Compute Benchmark*



Moreover, Figure 2's distribution is almost the same as the previous one, although the disparities among the GPUs are smaller. As we are taking into account a wider variety of characteristics in this one, such as price, the resulting plot will be more balanced. Nevertheless, it also shows that even when contemplating price, GTX 1070 Ti is still preferable than the rest of its competitors.

Figure 2: *High-end GPUs benchmark*



IV. Conclusion

Deep learning is one of the most impactful and promising technologies right now, which will exceed its current limitations, being able to reach general-purpose programs as well as the general public. In addition to this, models are expected to require less involvement from human engineers and more demanding hardware.

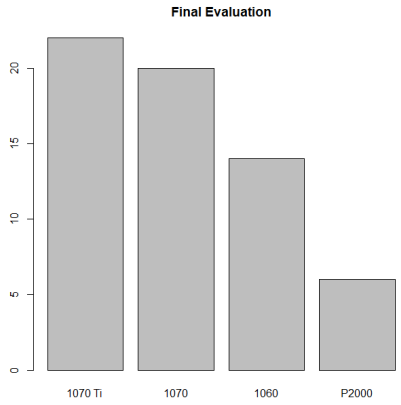
Table 2: *Final ratings*

	1070 Ti	1070	1060	P2000
Price	2	3	**4**	1
Bandwidth	**4**	4	2	1
CUDA	**4**	3	2	1
VRAM	**4**	4	2	1
Bench1	**4**	3	2	1
Bench2	**4**	3	2	1
Total	**22**	20	14	6

After considering plenty of features, GeForce GTX 1070 and GeForce GTX 1070 Ti have rised as the optimal candidates. In order to make a final decision, we have designed Table 2, which assigns to each specification of each GPU a score from 1 to 4 depending on the results shown in the previous parts, whose results have been plotted in Figure 3.

From Table 2, we can conclude the GTX 1070 Ti is our unequaled choice, beating its contenders in almost every single aspect. However, there is a chance that it is out of budget due to price fluctuations. In that case, we would opt for the GTX 1070, whose specifications are quite similar.

Figure 3: *Final results*



References

[1] Definition used by Andrew Ng in his Machine Learning online course https://www.coursera.org/learn/machine-learning/supplement/aAgxl/what-is-machine-learning, [online] last viewed 8 March 2018.

[2] Wikipedia's definition for Deep Learning https://en.wikipedia.org/wiki/Deep$_l$*earning*, [online] last viewed 8 March 2018.

[3] Slav Ivanov's blog entry on how to pick a GPU for deep learning https://blog.slavv.com/picking-a-gpu-for-deep-learning-3d4795c273b9, [online] last viewed 8 March 2018.

[4] Tim Dettmers' blog entry on using GPUs for deep learning http://timdettmers.com/2017/04/09/which-gpu-for-deep-learning/, [online] last viewd 8 March 2018.

[5] Direct compute benchmark. https://www.videocardbenchmark.net/directCompute.html, [online] last viewed 8 March 2018.

[6] High end GPUs benchmark https://www.videocardbenchmark.net/gpu$_v$*alue.html*, [online] last viewed 8 March 2018.

[7] NVIDIA http://www.nvidia.co.uk/page/home.html, [online] last viewed 8 March 2018.