NVIDIA-BTF Türkiye Education Grant Proposal
Heterogeneous Parallel Programming Course with Hands-on Experience

Işıl Öz

August 22, 2023

1 Project Goals and Objectives

In this project, we aim to develop a hands-on project experience for our undergraduate Heterogeneous Parallel Programming course in the Computer Engineering Department at Izmir Institute of Technology [3]. The main objectives of the project are as follows:

- We propose a practical term project for the undergraduate students registered for the Heterogeneous Parallel Programming course.
- The project will be based on YOLO model [9] variants (YOLOv3, YOLOv4, YOLOv5, and Tiny versions), an algorithm that uses neural networks to provide real-time object detection to identify and recognize objects in images. The project will potentially target object detection in scenarios for rescue operations in serious earthquake disasters [8, 10]. Our aim is to utilize NVIDIA Jetson Nano Developer Kits as the development and execution environment, such that it can be integrated into UAVs as accelerators. While full-system development, i.e., an embedded system with UAV integration, is out of the scope of the project, the students will be able to experience a potential real-life application.
- The students, who learn theoretical concepts from the course material prepared by utilizing NVIDIA Accelerated Computing Teaching Kit [4] modules and practice programming examples by attending NVIDIA Workshop, Fundamentals of Accelerated Computing with CUDA C/C++ [2], will gain experience on a practical real-life application. Eventually, we will have a complete teaching workflow based on NVIDIA GPU technologies.

2 Project Summary

In the computer engineering department at Izmir Institute of Technology in Turkey, the Heterogeneous Parallel Programming course has been offered based on the NVIDIA Accelerated Computing teaching kit. I am also an NVIDIA Deep Learning Institute (DLI) Instructor and University Ambassador and I encourage the registered students to attend the Accelerated Computing workshops organized in the department. The semester-long technical elective course covers GPU hardware, CUDA basics, advanced CUDA features, and parallel application development topics. While the content is updated each year, the main concepts and the corresponding teaching kit modules are presented in Table 1.

Table 1: Weekly Course Topics and Accelerated Computing Teaching Kit Modules.

Course Topic	**Teaching Kit Module**
Parallelism	Module 17 - Computational Thinking For Parallel Programming
Introduction to CUDA	Module 2 - Introduction to CUDA C
CUDA Threads	Module 3 - CUDA Parallelism Model
CUDA Memory	Module 4 - Memory and Data Locality
Tiling	Module 4 - Memory and Data Locality
Convolution	Module 8 - Parallel Computation Patterns (Stencil)
Parallel Patterns	Module 9-10 - Parallel Computation Patterns (Reduction - Scan)
CUDA Performance	Module 6 - Memory Access Performance
Dynamic Parallelism	Module 23 - Dynamic Parallelism
CUDA Libraries	Module 25 - Using CUDA Libraries
CUDA CNN	–

While the slides from the teaching kit are utilized in the specific modules, lab exercises and quiz questions are not used since there is no lab session or quiz in the course. Instead, self-developed programming assignments and midterm/final questions are designed for the course assessment and evaluation. Additionally, a final term project is assigned to the students, where *Project Guidelines* document of the Teaching Kit is utilized for defining the purpose, outline, and grading rubric of the project. The students are expected to propose and implement a complete CUDA application, conduct an experimental study, and perform a comparative analysis by comparing different CUDA implementations with other programming models, like OpenACC or other libraries. However, the project topics consist of more theoretical parts and only include CUDA-based simple programs. While the state-of-the-art cuDNN libraries are introduced in the course, the students are not able to find opportunities to practice these high-level implementations. With this project, we aim to fill the gap between low-level CUDA programming and high-level deep learning application, i.e., DL-based YOLO models for computer vision. Specifically, our main motivation is to develop a real-time object detection for a disaster relief situation, where it is critical to know about the survivors' location and damage's criticality. The project will require to develop an application running on the embedded GPU by utilizing Pytorch YOLO libraries [5],

which are based on NVIDIA cuDNN library functions. Jetson Nano Developer Kits enable us to have an embedded system with fast execution of YOLO object-detection algorithms to potentially help rescue operations.

3 Student Involvement

The students, who learn theoretical and practical CUDA programming and CuDNN library basics, will develop a high-level real-life application. Knowing the underlying execution details impacting the performance of the programs, they can evaluate the practical aspects of the GPU-accelerated programs. Not only do they develop simple core CUDA programs, but they also practice a complete real-life application targeting an earthquake relief scenario. They will be encouraged to study GPU-based deep learning concepts and be prepared to develop larger DNN models. A few graduate students have also registered for the course in previous years. We can expect them to continue their master's thesis on a related topic and conduct research based on the experience gathered in the project.

Starting from the 2018-2019 academic year, each year 10-20 students are registered in the Heterogeneous Parallel Programming course, and in average 60-80% of them can get a passing grade. Table 2 presents the statistics about the course in the last five years. It presents the number of students in terms of enrolled in the course, failed (got F) from the course, and received the highest letter grade, AA. Additionally, *Course Evaluation* column demonstrates the average score of the evaluation survey (out of 5), where the number in parenthesis represents the score for the question about the demonstration of the course content based on the quality of the course material and effective examples.

Table 2: Heterogeneous Parallel Programming course statistics.

Term	**#Students Taken**	**#Students Failed**	**#Students w/ AA**	**Course Evaluation**
2022-2023	10	4	x	3.74 (4.00)
2021-2022	11	2	1	4.23 (4.27)
2020-2021	12	4	5	2.84 (2.89)
2019-2020	20	8	3	3.94 (3.94)
2018-2019	16	6	1	3.79 (3.60)

For our project, which will be introduced as a final project in the course evaluation, we request 20 Jetson Nano Developer Kits. Each student is able to work individually and is required to complete the project. Moreover, they need to have peripherals like power adapters and SD memory cards, as detailed in Section 5.

4 Project Impact Measurement

We receive course evaluation reports at the end of each semester at Izmir Institute of Technology. Since the students grade different criteria and comment about the course, we can assess the project's impact based on the students' perceptions.

We already have a conference paper, based on my teaching experience in the course at SuperComputing (SC) Conference, Workshop on Best Practices for HPC Training and Education [6], and it has been published in the Journal of Computational Science Education [7]. However, there is no practical real-life experience in the course. With the Jetson Nano Developer kits, we aim to have a practical course project, and it will be a good teaching experience, therefore we are planning to prepare a conference paper as well. I am also a program committee member of the SC Workshop on Education for High-Performance Computing (EduHPC'23) [1], which can be a good fit for our contribution next year. We request financial support for our registration and travel expenses as given in the budget details.

5 Project Budget

Since less than 20 students have been registered in the course in the last five years 2, we request 20 Jetson Nano 2GB Developer Kits. Table 3 presents our other budget requests for the project. To utilize the developer kits in our classroom, we need peripherals including a power adapter, micro SD memory card, cooling fan, case, camera, and Wi-Fi network adapter. We included a sample amount (3.595,20 TL per kit) taken from a Turkish supplier in Figure 1. Additionally, we request financial support for one person's registration and travel expenses for the international conference or workshop attendance to present the paper.

Table 3: Project Budget Details.

	Unit	Quantity	Total
Jetson Nano Peripherals (power adaptor, SD card, cooling fan, case, camera, wi-fi adaptor)	160 USD	20	3200 USD
Conference/workshop registration and travel expenses		1	3000 USD
TOTAL			**6200 USD**

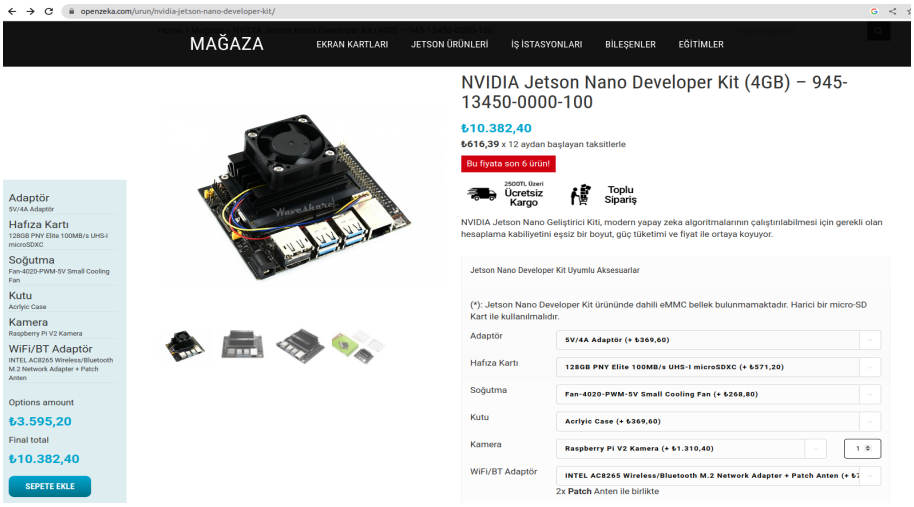

Figure 1: NVIDIA Jetson Nano Developer Kits (Options Amount Requested).

References

[1] Eduhpc-23: Workshop on education for high-performance computing. `https://tcpp.cs.gsu.edu/curriculum/?q=eduhpc23`. Accessed: 2023-08-22.

[2] Fundamentals of accelerated computing with cuda c/c++. `https://courses.nvidia.com/courses/course-v1:DLI+C-AC-01+V1/`. Accessed: 2023-08-22.

[3] Heterogenous parallel programming, izmir institute of technology. `https://ceng.iyte.edu.tr/courses/ceng-443/`. Accessed: 2023-08-22.

[4] Nvidia dli accelerated computing teaching kit. `https://www.nvidia.com/en-us/training/teaching-kits/`. Accessed: 2023-08-22.

[5] Pytorch yolov5. `https://pytorch.org/hub/ultralytics_yolov5/`. Accessed: 2023-08-22.

[6] Teaching accelerated computing and deep learning at a large-scale with the nvidia deep learning institute. `https://sc22.supercomputing.org/proceedings/workshops/workshop_pages/ws_bphpcte110.html`. Accessed: 2023-08-22.

[7] Bálint Gyires-Tóth, Işıl Öz, and Joe Bungo. Teaching accelerated computing and deep learning at a large-scale with the nvidia deep learning institute. *The Journal of Computational Science Education*, 14:23–30, July 2023.

[8] Yafei Jing, Yuhuan Ren, Yalan Liu, Dacheng Wang, and Linjun Yu. Automatic extraction of damaged houses by earthquake based on improved yolov5: A case study in yangbi. *Remote Sensing*, 14(2), 2022.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[10] Junhao Wang. An improved yolo algorithm for object detection in all day scenarios. In Han Qiu, Cheng Zhang, Zongming Fei, Meikang Qiu, and Sun-Yuan Kung, editors, *Knowledge Science, Engineering and Management*, pages 475–486, Cham, 2021. Springer International Publishing.