AI-Powered Video Monitoring: Assessing the NVIDIA Jetson Orin Devices for Edge Computing Applications

6. Smart mobility

Note: Do not include author names or affiliation anywhere in the digest – the review process is double blind

Abstract

The NVIDIA Jetson devices have become a standard in AI and edge computing applications. This paper presents the performance evaluation of the Jetson Orin family, NVIDIA's latest devices released in 2023, in the context of surveillance applications considering CVEDIA-RT software. The NVIDIA Jetson Orin AGX developer kit is used as a means to emulate the Orin NX and Orin Nano devices. A testing procedure based on augmented scripts is described, where performance metrics of the devices, such as RAM, GPU and CPU usage, are assessed. Considering the surveillance use case, footage from a parking lot is used as a benchmark. Results demonstrate that all devices are able to achieve at least an average 10 FPS, indicating the devices as potential candidates for the aforementioned application. Moreover, it is shown that superior devices present superior FPS performance when compared to the lower end Nano counterparts. Lastly, for the final version of the paper, the authors plan to evaluate the YOLOv7 algorithm on an actual Orin device.

Index Terms

Edge computing, machine vision. object detection, surveillance

I. Introduction

Remote monitoring and management systems are essential to modern intelligent surveillance systems. They are commonly deployed at construction sites, parking lots and remote locations and perform tasks such as detecting cars and people. With the latest generation of edge devices, which are capable of internet connection, a larger range of applications has been made available. In particular, these devices enable real-time surveillance with smart communication and intruder detection, for example [1], [2].

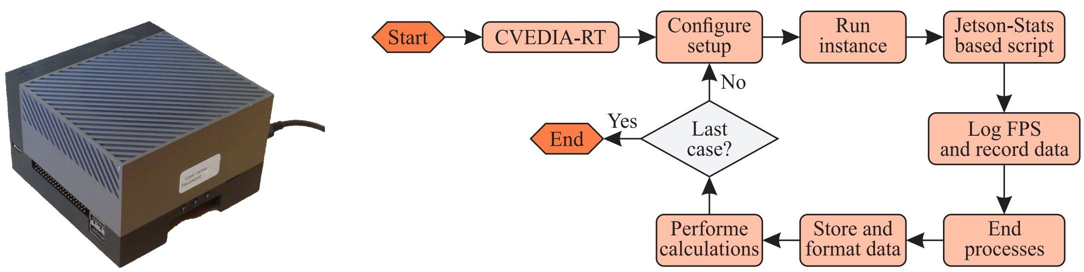
Existing studies have achieved success in many camera-based intelligent surveillance systems applications, typically using an AI computation device powering a vision neural network [3]. Such systems, typically restrained by the power supply, need to consider AI computation performance and efficiency at the same time. The NVIDIA Jetson family has been an industry standard of AI computation devices. On the software side, one-stage object detection models, such as You-Only-Look-Once (YOLO) [4] and Single Shot Detector (SSD) [5], are the primary choices running on these devices. These models have different variations that suit the need for various computation requirements, and their detection performance is positively related to computation. For example, at an 8 FPS target, an NVIDIA Jetson Nano can run a YOLO model but only at a low resolution, compromising its detection accuracy

[6]. The more powerful Jetson TX2 provides more computation overhead but comes with a larger size and power consumption [7]. New devices with better performance-to-power ratio are much desired.

In this context, this paper presents the performance evaluation of NVIDIA's latest devices released in 2023, the Jetson Orin family. Tests are carried out considering surveillance applications using CVEDIA-RT software. As a means to easily test all the NX and Nano series devices, the NVIDIA Jetson Orin AGX developer kit is used to seamlessly emulate them. A testing procedure is developed based on augmenting pre-existing edge device analytics tools, enabling performance metrics of the devices, such as RAM, GPU and CPU usage, to be assessed. Video data from a parking lot pre-built into CVEDIA-RT is used as a benchmark to analyze the performance of the surveillance application scenario. Results for average and maximum FPS, as well as average utilization metrics, are provided for the devices of the family, allowing direct comparison. Future improvements regarding the final version of the paper are listed in the conclusion of this digest.

II. METHODOLOGY

Understanding the capabilities of the newer NVIDIA processors is essential to determine their capability for IoT edge computing tasks, such as the ones present in intelligent surveillance systems, for example. In this section, the Jetson Orin AGX device and the emulation process of the remaining Orin devices are described. A straightforward and systematic performance testing procedure is also developed.



(a) (b)

Fig. 1. Methodology. (a) nVidia Jetson AGX Orin Developer Kit. (b) Flowchart of the proposed performance testing procedure.

A. AGX Device and Emulation

The Jetson Orin family of devices is composed of six different modules: AGX Orin 64Gb, AGX Orin 32Gb, Orin NX 16Gb, Orin NX 8Gb, Orin Nano 8Gb and Orin Nano 4Gb. Given the large number of devices, developing and testing each alternative can be cumbersome. However, the Jetson AGX Orin Developer Kit, presented in Fig. 1(a), enables the development in a single platform along with native emulation of the aforementioned devices. This device is based on the AGX Orin 64Gb and is capable of upwards of 275 trillion operations per second (TOPS) [8]. The emulation process is achieved by flashing the developer kit with the desired image, following predetermined commands. Hence, the transition between different modules can be obtained by a simple reflashing of the device.

B. *Performance testing procedure*

CVEDIA-RT, a cross-platform and modular AI inference engine, is used in order to evaluate the Jetson Orin devices. It features low and high-level interfaces, as well as a number of use cases pre-built into the software, such as intelligent traffic systems, surveillance and crowd estimation. The flowchart of the proposed testing procedure is presented in Fig. 1(b). Initially, the AGX Orin Developer Kit is flashed according to the desired module. Then, after the setup is complete, CVEDIA-RT software is executed and the necessary configurations are applied to the setup. Next, an instance of the surveillance algorithm is executed, using the benchmark video from a parking lot, provided along with CVEDIA-RT. During execution, a Jetson-Stats based script is used as a means to extract and record data for instantaneous RAM, GPU and CPU utilization during a 10-minute time window. In addition, FPS values are also extracted from the CVEDIA-RT software. Once the instance is finished, the processes are ended and the data is formatted and stored in .csv files. Lastly, calculations are performed in order to determine the average values for the desired metrics, as well as determining the maximum achieved FPS value. Following this, the board is reflashed for a different Orin module and the procedure is repeated until every module has been evaluated.

III. Experimental results

This section presents the performance metrics for the different performance tests carried out for the Orin modules. Note that for this paper, only the Orin NX 16Gb, Orin NX 8Gb, Orin Nano 8Gb and Orin Nano 4Gb have been considered. The AGX modules have not been evaluated due to their higher price points and excessive hardware features, which are not justified or necessary for applications discussed in this paper.

In order to evaluate the surveillance use case, a video from an a parking lot, provided along with CVEDIA-RT, is used as a benchmark. During execution, the algorithm is responsible for detecting individuals within a predetermined area. A frame of the video is presented in Fig. 2.

Fig. 2. Frame from the parking lot video provided with CVEDIA-RT, with the monitored area highlighted in blue.

The average and maximum FPS values for the different Orin modules are presented in Fig. 3. It can be seen that the four modules are able to achieve a minimum 10 FPS average value, and peak values of upwards of 33 FPS. Moreover, it is observed that the NX module present a 15 FPS minimum average value, a higher performance when compared to the Nano modules. Note that this is expected given the superior hardware of the NX devices.

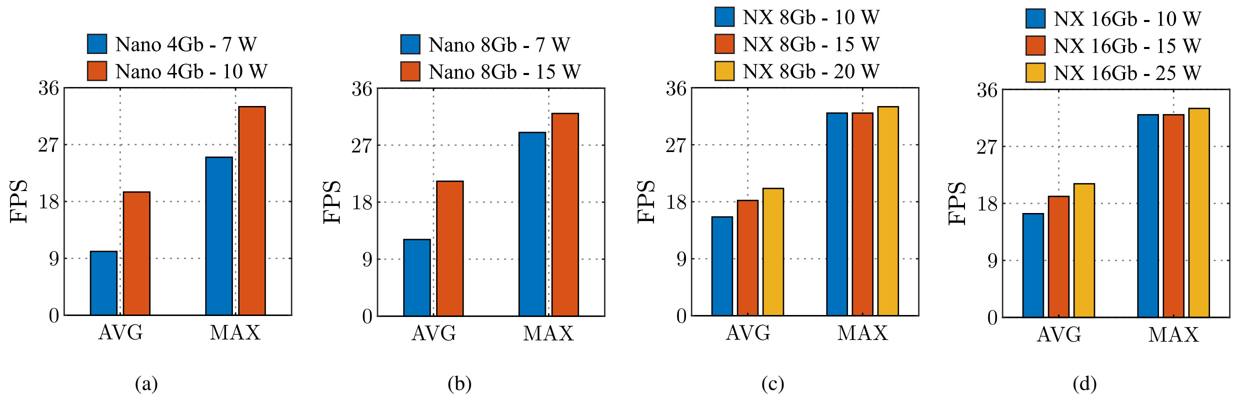


Fig. 3. Average and maximum FPS values for the different Orin devices. (a) Orin Nano 4Gb. (b) Orin Nano 8Gb. (c) Orin NX 8Gb. (d) Orin NX 16Gb.

Fig. 4 presents the average utilization of RAM, GPU and CPU for the four different Orin devices. Initially, it can be observed that all of the modules are capable of running the computer vision algorithm without using the entirety of the available resources, even on the lower-end module. Moreover, it can also be seen that the different devices present relatively similar utilization, with no obvious trend being observed. This is justified by the varying FPS performance, shown in Fig. 3. Lastly, it can be seen that most devices show a reduction in CPU usage with the increase of power configuration.

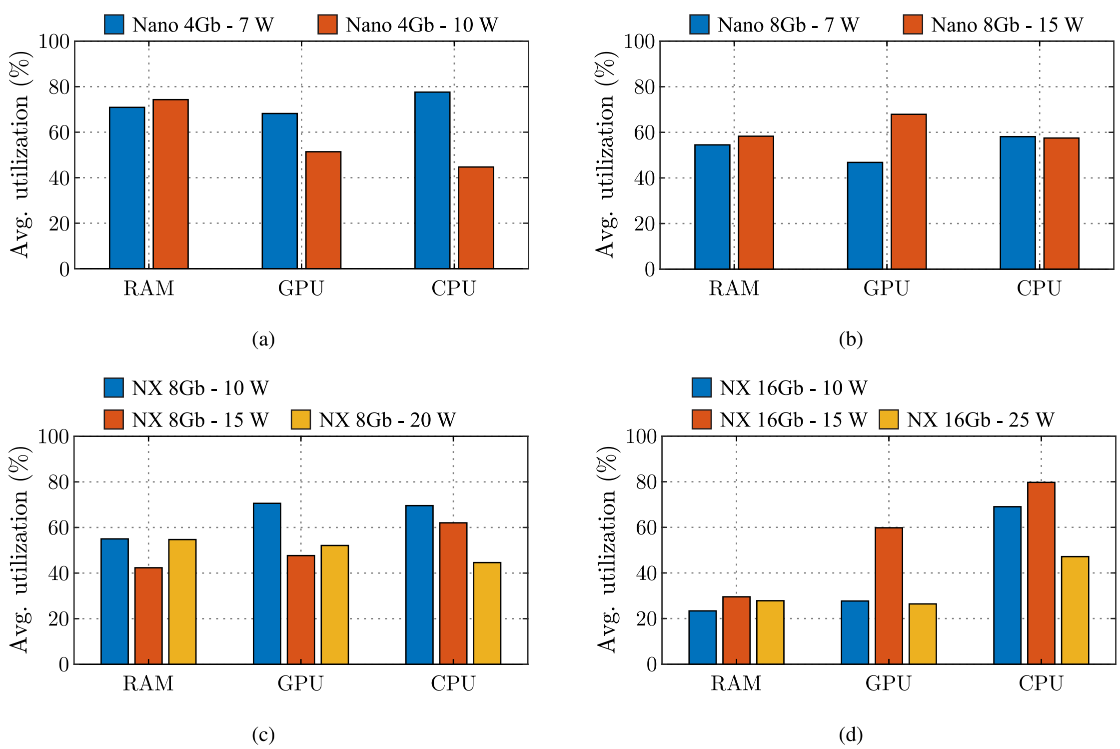


Fig. 4. Average utilization of RAM, GPU and CPU for the different Orin devices. (a) Orin Nano 4Gb. (b) Orin Nano 8Gb. (c) Orin NX 8Gb. (d) Orin NX 16Gb.

IV. Conclusion

This paper presents the performance analysis of the NVIDIA Jetson Orin family of devices for machine vision applications considering CVEDIA-RT software. It can be seen that both the Orin NX and Orin Nano devices are able to achieve an average of 10 FPS in traffic management benchmark, regardless of the power setting and memory configuration. Moreover, it is observed that the higher end devices present superior FPS performance when compared to the lower end Nano counterparts. Due to the limited digest length, significant improvements will be included in the final paper, them being: a detailed description of the AGX device, an illustration of the emulation process and an in-depth discussion of the testing procedure. In addition, a comparison with other machine vision tasks will be presented. Lastly, the authors plan to evaluate the YOLOv7 algorithm on an actual Orin device.

References

[1] I. Lashkov, R. Yuan, and G. Zhang, "Edge-computing-empowered vehicle tracking and speed estimation against strong image vibrations using surveillance monocular camera," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 11, pp. 13 486–13 502, 2023.

[2] M. Adl, M. Alizadeh, S. Habibi, C. Vidal, and A. Emadi, "Traffic enforcement at intersections monitored by a single fisheye camera containing noisy detection and tracking data," in *IECON 2022 – 48th Annual Conference of the IEEE Industrial Electronics Society*, 2022, pp. 1–6.

[3] V. Mandal and Y. Adu-Gyamfi, "Object detection and tracking algorithms for vehicle counting: a comparative analysis," *Journal of big data analytics in transportation*, vol. 2, no. 3, pp. 251–261, 2020.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[6] Y. Huangfu, M. Ahrabi, R. Tahal, J. Huang, A. Mohammad-Alikhani, S. Reymann, B. Nahid-Mobarakeh, S. Shirani, and S. Habibi, "Efficient edge computing device for traffic monitoring using deep learning detectors," *IEEE Canadian Journal of Electrical and Computer Engineering*, vol. in publication process, no. 0, 2023.

[7] C. Chen, B. Liu, S. Wan, P. Qiao, and Q. Pei, "An edge traffic flow detection scheme based on deep learning in an intelligent transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1840–1852, 2021.

[8] "Develop for all six nvidia jetson orin modules with the power of one developer kit — nvidia technical blog," https://developer.nvidia.com/blog/develop-for-all-six-nvidia-jetson-orin-modules-with-the-power-of-one-developer-kit/, (Accessed on 11/01/2023).