ISSN: 2278-4632
Vol-13, Issue-07, July 2023

VLSI Implementation of Deep Learning Architectures for Advanced Image Recognition on Embedded Systems

Dr. Ravva Gurunadha
Associate Professor, Department of Electronics and Communications Engineering
JNTU-GV College of Engineering Vizianagaram, A.P, India
gururavva@gmail.com

Abstract

The advent of deep learning has revolutionized image recognition tasks, yet deploying such architectures on embedded systems presents challenges due to their limited computational resources. This paper investigates VLSI (Very-Large-Scale Integration) implementation strategies for deep learning models tailored for advanced image recognition on embedded systems. We present two case studies: (1) the implementation of a Convolutional Neural Network (CNN) for real-time object detection on an FPGA (Field-Programmable Gate Array), and (2) the deployment of a lightweight Neural Architecture Search (NAS)-optimized model on an ASIC (Application-Specific Integrated Circuit) for high-efficiency image classification. By comparing these case studies, we highlight trade-offs in performance, resource utilization, and deployment efficiency.

Keywords: Deep Learning, VLSI Implementation, Convolutional Neural Networks (CNN), Neural Architecture Search (NAS), FPGA (Field-Programmable Gate Array), ASIC (Application-Specific Integrated Circuit), Embedded Systems, Image Recognition

Introduction

Deep learning has revolutionized image recognition, but deploying these models on constrained embedded systems presents significant challenges. VLSI technologies like FPGA and ASIC offer potential solutions for efficient deployment. This paper explores VLSI implementations of deep learning architectures through two case studies, evaluating their effectiveness in real-world applications [1-8]. The rapid advancement of deep learning has led to remarkable achievements in image recognition, but deploying complex neural networks on resource-constrained systems remains a challenge due to limited computational power, memory, and energy budgets. The paper is mostly about making efficient VLSI implementations of deep learning architectures that work on embedded platforms. It also talks about hardware acceleration techniques and how new memory technologies might be able to get around the problems that come with traditional von Neumann architectures. Deploying these models on limited embedded systems presents significant challenges, despite the fact that deep learning has improved picture recognition. Large-scale integrated circuit (VLSI)

ISSN: 2278-4632
Vol-13, Issue-07, July 2023

technologies like FPGA and ASIC provide potential options for effective deployment. This article uses two case studies to study the implementation of deep learning architectures on very large-scale integrated circuits (VLSI) and assess their effectiveness in real-world applications. However, deploying massive neural networks on resource-constrained devices remains a challenge due to limited computing capability, memory, and energy restrictions. Despite its rapid rise, deep learning has resulted in astonishing achievements in picture identification. When it comes to deep learning architectures, the research mostly focuses on creating efficient VLSI implementations that are compatible with embedded systems. It also covers alternative hardware acceleration solutions and how new memory technologies may be able to overcome the constraints associated with old von Neumann architectures [9-12].

2. Methodology

Case Study 1: CNN for Real-Time Object Detection on FPGA

2.1.1. Overview

The ability of Convolutional neural networks (CNNs) to learn hierarchical features from images makes them widely used for object detection tasks. Implementing a CNN on the FPGA provides a balance between performance and resource efficiency, leveraging the FPGA's parallel processing capabilities.

2.1.2. Methodology

- **Architecture Design**: Create a custom CNN architecture optimized for the FPGA, focusing on efficient data flow and reduced precision arithmetic**.** Key components include convolutional layers, pooling layers, and fully connected layers.
- **Implementation**: Use hardware description languages (HDL) such as VHDL or Verilog to implement the CNN. Utilize high-level synthesis (HLS) tools to streamline the process and optimize the hardware for real-time processing.
- **Performance Metrics**: Evaluate performance based on throughput, latency, and resource utilization (e.g., logic elements, memory usage). Compare the FPGA implementation's performance against software-based CNN models on CPUs and GPUs.

ISSN: 2278-4632
Vol-13, Issue-07, July 2023

2.1.3. Results

- **Advantages**: High parallelism leads to improved real-time processing speeds. The architecture can be easily reconfigured to accommodate various image sizes and complexities.
- **Limitations**: FPGA implementations may face constraints in on-chip memory and power consumption. Development complexity and design time are also considerations.

3. Comparison of Case Studies

3.1. Performance

- **CNN on FPGA**: Provides high-speed processing with the flexibility to adapt to different image recognition tasks. FPGA resource constraints and design complexity influence performance.
- The NAS-Optimized Model on ASIC provides optimal performance and power efficiency for the specific image classification task**.** ASIC's custom design ensures high throughput and low latency.

3.2. Resource utilization

- **CNN on FPGA**: Efficient use of FPGA resources through parallel processing and pipelining. However, FPGA designs may face limitations in memory and power consumption.
- **NAS-Optimized Model on ASIC**: Highly efficient in terms of power and area due to ASIC's fixed design. Specific tasks optimize the ASIC implementation, but it lacks flexibility.

3.3. Cost and Development Time

- **CNN on FPGA offers a lower initial development cost and a faster time-to-market due to its reconfigurable nature.** Because of custom hardware design, development complexity and time are significant.
- The NAS-Optimized Model on ASIC has higher initial design and manufacturing costs, but it offers long-term cost benefits and better performance. Development time is longer due to the complexity of ASIC design and fabrication.

4. Conclusion

ISSN: 2278-4632
Vol-13, Issue-07, July 2023

The specific requirements of the application determine whether to use FPGA or ASIC when developing deep learning architectures on embedded systems. CNNs that are based on FPGAs provide more flexibility and high-speed processing for image recognition tasks, while ASIC-based NAS-optimized versions offer improved performance and efficiency for certain applications. Future studies should explore hybrid methodologies and VLSI technology developments to improve the deployment of deep learning models on embedded devices. The purpose of this work is to discuss the design and implementation of effective deep learning architectures for performing complex image recognition tasks on embedded devices. The results of this study demonstrate significant reductions in power consumption, latency, and accuracy. The described techniques have the potential to run a wide variety of intelligent applications on devices with limited resource accessibility. Future research will concentrate on developing more efficient designs and new hardware technologies.

References

1. Udendhran, R., M. Balamurugan, Annamalai Suresh, and R. Varatharajan. "Enhancing image processing architecture using deep learning for embedded vision systems." Microprocessors and Microsystems 76 (2020): 103094.
2. Pérez, Ignacio, and Miguel Figueroa. "A heterogeneous hardware accelerator for image classification in embedded systems." Sensors 21, no. 8 (2021): 2637.
3. Chen, Yanjiao, Baolin Zheng, Zihan Zhang, Qian Wang, Chao Shen, and Qian Zhang. "Deep learning on mobile and embedded devices: State-of-the-art, challenges, and future directions." ACM Computing Surveys (CSUR) 53, no. 4 (2020): 1-37.
4. Eldafrawy, Mohamed, Andrew Boutros, Sadegh Yazdanshenas, and Vaughn Betz. "FPGA logic block architectures for efficient deep learning inference." ACM Transactions on Reconfigurable Technology and Systems (TRETS) 13, no. 3 (2020): 1-34.
5. Lopez-Montiel, Miguel, Ulises Orozco-Rosas, Moisés Sánchez-Adame, Kenia Picos, and Oscar Humberto Montiel Ross. "Evaluation method of deep learning-based embedded systems for traffic sign detection." IEEE Access 9 (2021): 101217-101238.
6. Akkad, Ghattas, Ali Mansour, and Elie Inaty. "Embedded deep learning accelerators: A survey on recent advances." IEEE Transactions on Artificial Intelligence (2023).
7. Zaman, Kh Shahriya, Mamun Bin Ibne Reaz, Sawal Hamid Md Ali, Ahmad Ashrif A. Bakar, and Muhammad Enamul Hoque Chowdhury. "Custom hardware architectures for deep learning on portable devices: a review." IEEE Transactions on Neural Networks and Learning Systems 33, no. 11 (2021): 6068-6088.

ISSN: 2278-4632
Vol-13, Issue-07, July 2023

8. Manikandababu, C. S., M. Jagadeeswari, and H. Mohammed Irfan. "Low-Power VLSI Design for Image Analysis in Embedded Vision Systems." In 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), pp. 838-844. IEEE, 2023.
9. Giri, Davide, Kuan-Lin Chiu, Giuseppe Di Guglielmo, Paolo Mantovani, and Luca P. Carloni. "ESP4ML: Platform-based design of systems-on-chip for embedded machine learning." In 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1049-1054. IEEE, 2020.
10. An, Hyochan, Sam Schiferl, Siddharth Venkatesan, Tim Wesley, Qirui Zhang, Jingcheng Wang, Kyojin D. Choo et al. "An ultra-low-power image signal processor for hierarchical image recognition with deep neural networks." IEEE Journal of Solid-State Circuits 56, no. 4 (2020): 1071-1081.
11. Messaoud, Seifeddine, Soulef Bouaafia, Amna Maraoui, Ahmed Chiheb Ammari, Lazhar Khriji, and Mohsen Machhout. "Deep convolutional neural networks-based Hardware–Software on-chip system for computer vision application." Computers & Electrical Engineering 98 (2022): 107671.
12. Amuru, Deepthi, Andleeb Zahra, Harsha V. Vudumula, Pavan K. Cherupally, Sushanth R. Gurram, Amir Ahmad, and Zia Abbas. "AI/ML algorithms and applications in VLSI design and technology." Integration 93 (2023): 102048.