Fpga Implementations Of Neural Networks

Machine Learning and Its Hardware Implementation
FPGA Implementations of Neural Networks
Artificial Neural Networks - ICANN 96
Fpga Implementation of Hopfield Neural Network
Deep Learning and Convolutional Neural Networks for Medical Image Computing
In-situ Implementation and Training of Convolutional Neural Network on FPGAs
Third International Symposium on Neural Networks, ISNN 2006, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part III
Proceedings of ESAI 2019, Fez, Morocco
NengoFPGA
Reconfigurable Computing: Architectures, Tools and Applications
Second International Workshop, ARC 2006, Delft, The Netherlands, March 1-3, 2006 Revised Selected Papers
FPGA Implementation of Reduced Precision Convolutional Neural Networks
Advances in Neural Networks-isnn 2006
Applied Reconfigurable Computing. Architectures, Tools, and Applications
Field-Programmable Logic and Applications
Neural Information Processing
The NEURON Book
Advances on P2P, Parallel, Grid, Cloud and Internet Computing
International Work-Conference on Artificial and Natural Neural Networks, IWANN'99, Alicante, Spain, June 2-4, 1999, Proceedings
Design of a Neural Network for FPGA Implementation
Selected Papers from the XIX International Conference on Neuroinformatics, October 2-6, 2017, Moscow, Russia
2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS)
22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11-14, 2012, Proceedings, Part I
6th International Conference, Bochum, Germany, July 16 - 19, 1996. Proceedings
Advances in Neural Computation, Machine Learning, and Cognitive Research
Applications to Electrical, Electronics and Computer Science and Engineering
Artificial Neural Networks and Machine Learning -- ICANN 2012
Implementation of an FPGA-based Artificial Neural Network
Efficient Processing of Deep Neural Networks
Thinking Machines
Embedded Systems and Artificial Intelligence
Application of FPGA to Real-Time Machine Learning
Precision Medicine, High Performance and Large-Scale Datasets
FPGA Implementations of Neural Networks
Artificial Intelligence and Applied Mathematics in Engineering Problems
Hardware Reservoir Computers and Software Image Processing
14th International Conference , FPL 2004, Leuven, Belgium, August 30-September 1, 2004, Proceedings
FPGA Implementation of a PC-AT Computer to Neural Network Interface
FPGA Implementation of PSO Algorithm and Neural Networks

DENNIS ROMAN

Machine Learning and Its Hardware Implementation Springer
In recent years deep learning algorithms have shown extremely high performance on machine learning tasks such as image classification and speech recognition. In support of such applications, various FPGA accelerator architectures have been proposed for convolutional neural networks (CNNs) that enable high performance for classification tasks at lower power than CPU and GPU processors. However, to date, there has been little research on the use of FPGA implementations of deconvolutional neural networks (DCNNs). DCNNs, also known as generative CNNs, encode high-dimensional probability distributions and have been widely used for computer vision applications such as scene completion, scene segmentation, image creation, image denoising, and super-resolution imaging. We propose an FPGA architecture for deconvolutional networks built around an accelerator which effectively handles the complex memory access patterns needed to perform strided deconvolutions, and that supports convolution as well. We also develop a three-step design optimization method that systematically exploits statistical analysis, design space exploration and VLSI optimization. To verify our FPGA deconvolutional accelerator design methodology we train DCNNs offline on two representative datasets using the generative adversarial network method (GAN) run on Tensorflow, and then map these DCNNs to an FPGA DCNN-plus-accelerator implementation to perform generative inference on a Xilinx Zynq-7000 FPGA. Our DCNN implementation achieves a peak performance density of 0.012 GOPs/DSP.
FPGA Implementations of Neural Networks Springer
The two-volume set LNCS 7552 + 7553 constitutes the proceedings of the 22nd International Conference on Artificial Neural Networks, ICANN 2012, held in Lausanne, Switzerland, in September 2012. The 162 papers included in the proceedings were carefully reviewed and selected from 247 submissions. They are organized in topical sections named: theoretical neural computation; information and optimization; from neurons to neuromorphism; spiking dynamics; from single neurons to networks; complex firing patterns; movement and motion; from sensation to perception; object and face recognition; reinforcement learning; bayesian and echo state networks; recurrent neural networks and reservoir computing; coding architectures; interacting with the brain; swarm intelligence and decision-making; mulitlayer perceptrons and kernel networks; training and learning; inference and recognition; support vector machines; self-organizing maps and clustering; clustering, mining and exploratory analysis; bioinformatics; and time weries and forecasting.
Artificial Neural Networks - ICANN 96 Springer Science & Business Media
During the 1980s and early 1990s there was signi?cant work in the design and implementation of hardware neurocomputers. Nevertheless, most of these efforts may be judged to have been unsuccessful: at no time have have ha- ware neurocomputers been in wide use. This lack of success may be largely attributed to the fact that earlier work was almost entirely aimed at developing custom neurocomputers, based on ASIC technology, but for such niche - eas this technology was never suf?ciently developed or competitive enough to justify large-scale adoption. On the other hand, gate-arrays of the period m- tioned were never large enough nor fast enough for serious arti?cial-neur- network (ANN) applications. But technology has now improved: the capacity and performance of current FPGAs are such that they present a much more realistic alternative. Consequently neurocomputers based on FPGAs are now a much more practical

proposition than they have been in the past. This book summarizes some work towards this goal and consists of 12 papers that were selected, after review, from a number of submissions. The book is nominally divided into three parts: Chapters 1 through 4 deal with foundational issues; Chapters 5 through 11 deal with a variety of implementations; and Chapter 12 looks at the lessons learned from a large-scale project and also reconsiders design issues in light of current and future technology.

Fpga Implementation of Hopfield Neural Network Springer

This is Volume III of a three volume set constituting the refereed proceedings of the Third International Symposium on Neural Networks, ISNN 2006. 616 revised papers are organized in topical sections on neurobiological analysis, theoretical analysis, neurodynamic optimization, learning algorithms, model design, kernel methods, data preprocessing, pattern classification, computer vision, image and signal processing, system modeling, robotic systems, transportation systems, communication networks, information security, fault detection, financial analysis, bioinformatics, biomedical and industrial applications, and more.

Deep Learning and Convolutional Neural Networks for Medical Image Computing Springer

The development of neural networks has now reached the stage where they are employed in a large variety of practical contexts. However, to date the majority of such implementations have been in software. While it is generally recognised that hardware implementations could, through performance advantages, greatly increase the use of neural networks, to date the relatively high cost of developing Application-Specific Integrated Circuits (ASICs) has meant that only a small number of hardware neurocomputers has gone beyond the research-prototype stage. The situation has now changed dramatically: with the appearance of large, dense, highly parallel FPGA circuits it has now become possible to envisage putting large-scale neural networks in hardware, to get high performance at low costs. This in turn makes it practical to develop hardware neural-computing devices for a wide range of applications, ranging from embedded devices in high-volume/low-cost consumer electronics to large-scale stand-alone neurocomputers. Not surprisingly, therefore, research in the area has recently rapidly increased, and even sharper growth can be expected in the next decade or so. Nevertheless, the many opportunities offered by FPGAs also come with many challenges, since most of the existing body of knowledge is based on ASICs (which are not as constrained as FPGAs). These challenges range from the choice of data representation, to the implementation of specialized functions, through to the realization of massively parallel neural networks; and accompanying these are important secondary issues, such as development tools and technology transfer. All these issues are currently being investigated by a large number of researchers, who start from different bases and proceed by different methods, in such a way that there is no systematic core knowledge to start from, evaluate alternatives, validate claims, and so forth. FPGA Implementations of Neural Networks aims to be a timely one that fill this gap in three ways: First, it will contain appropriate foundational material and therefore be appropriate for advanced students or researchers new to the field. Second, it will capture the state of the art, in both depth and breadth and therefore be useful researchers currently active in the field. Third, it will cover directions for future research, i.e. embryonic areas as well as more speculative ones.

In-situ Implementation and Training of Convolutional Neural Network on FPGAs Springer

This book presents the latest research findings, innovative research results, methods and development techniques related to P2P, grid, cloud and Internet computing from both theoretical and practical perspectives. It also reveals the synergies among such large-scale computing paradigms. P2P, grid, cloud and Internet computing technologies have rapidly become established as breakthrough paradigms for solving complex problems by enabling aggregation and sharing of an increasing variety of distributed computational resources at large scale. Grid computing originated as a paradigm for high-performance computing, as an alternative to expensive supercomputers through different forms of large-scale distributed computing. P2P computing emerged as a new paradigm after client–server and web-based computing and has proved useful in the development of social networking, B2B (business to business), B2C (business to consumer), B2G (business to government), and B2E (business to employee). Cloud computing has been defined as a “computing paradigm where the boundaries of computing are determined by economic rationale rather than technical limits,” and it has fast become a computing paradigm with applicability and adoption in all application domains and which provides utility computing at a large scale. Lastly, Internet computing is the basis of any large-scale distributed computing paradigms; it has developed into a vast area of flourishing fields with enormous impact on today’s information societies, and serving as a universal platform comprising a large variety of computing forms such as grid, P2P, cloud and mobile computing.

Third International Symposium on Neural Networks, ISNN 2006, Chengdu, China, May 28 - June 1, 2006, Proceedings, Part III Springer

This book features research presented at the 1st International Conference on Artificial Intelligence and Applied Mathematics in Engineering, held on 20–22 April 2019 at Antalya, Manavgat (Turkey). In today’s world, various engineering areas are essential components of technological innovations and effective real-world solutions for a better future. In this context, the book focuses on problems in engineering and discusses research using artificial intelligence and applied mathematics. Intended for scientists, experts, M.Sc. and Ph.D. students, postdocs and anyone interested in the subjects covered, the book can also be used as a reference resource for courses related to artificial intelligence and applied mathematics.

Proceedings of ESAI 2019, Fez, Morocco Springer

This book constitutes the refereed proceedings of the Third International Workshop on Applied Reconfigurable Computing, ARC 2007, held in Mangaratiba, Brazil, in March 2007. The 27 full papers and 10 short papers presented together with a late-comer contribution from ARC 2006 are organized in topical sections on architectures, mapping techniques and tools, arithmetic, and applications.

NengoFPGA Springer Science & Business Media

This book gathers selected research papers presented at the First International Conference on Embedded Systems and Artificial Intelligence (ESAI 2019), held at Sidi Mohamed Ben Abdellah University, Fez, Morocco, on 2–3 May 2019. Highlighting the latest innovations in Computer Science, Artificial Intelligence, Information Technologies, and Embedded Systems, the respective papers will encourage and inspire researchers, industry professionals, and policymakers to put these methods into practice.

Reconfigurable Computing: Architectures, Tools and Applications Springer Science & Business Media

The authoritative reference on NEURON, the simulation environment for modeling biological neurons and neural networks that enjoys wide use in the experimental and computational neuroscience communities. This book shows how to use NEURON to construct and apply empirically based models. Written primarily for neuroscience investigators, teachers, and students, it assumes no previous knowledge of computer programming or numerical methods. Readers with a background in the physical sciences or mathematics, who have some knowledge about brain cells and circuits and are interested in computational modeling, will also find it helpful. The NEURON Book covers material that ranges from the inner workings of this program, to practical considerations involved in specifying the anatomical and biophysical properties that are to be represented in models. It uses a problem-solving approach, with many working examples that readers can try for themselves.

Second International Workshop, ARC 2006, Delft, The Netherlands, March 1-3, 2006 Revised Selected Papers LAP Lambert Academic Publishing

This book contains the papers presented at the 14th International Conference on Field Programmable Logic and Applications (FPL) held during August 30th- September 1st 2004. The conference was hosted by the Interuniversity Micro- Electronics Center (IMEC) in Leuven, Belgium. The FPL series of conferences was founded in 1991 at Oxford University (UK), and has been held annually since: in Oxford (3 times), Vienna, Prague, Darmstadt, London, Tallinn, Glasgow, Villach, Belfast, Montpellier and Lisbon. It is the largest and oldest conference in reconfigurable computing and brings together academic researchers, industry experts, users and newcomers in an informal, welcoming atmosphere that encourages productive exchange of ideas and knowledge between the delegates. The fast and exciting advances in field programmable logic are increasing steadily with more and more application potential and need. New ground has been broken in architectures, design techniques, (partial) run-time reconfiguration and applications of field programmable devices in several different areas. Many of these recent innovations are reported in this volume. The size of the FPL conferences has grown significantly over the years. FPL in 2003 saw 216 papers submitted. The interest and support for FPL in the programmable logic community continued this year with 285 scientific papers submitted, demonstrating a 32% increase when compared to the year before. The technical program was assembled from 78 selected regular papers, 45 additional short papers and 29 posters, resulting in this volume of proceedings. The program also included three invited plenary keynote presentations from Xilinx, Gilder Technology Report and Altera, and three embedded tutorials from Xilinx, the Universit¨ at Karlsruhe (TH) and the University of Oslo.

FPGA Implementation of Reduced Precision Convolutional Neural Networks Springer

This book constitutes the thoroughly refereed post-proceedings of the Second International Workshop on Reconfigurable Computing, ARC 2006, held in Delft, The Netherlands, in March 2006. The 22 revised full papers and 35 revised short papers presented were thoroughly reviewed and selected from 95 submissions. The papers are organized in topical sections on applications, power, image processing, organization and architecture, networks and communication, security, and tools.

Advances in Neural Networks-isnn 2006 Springer Science & Business Media

Low-power, high-speed neural networks are critical for providing deployable embedded AI applications at the edge. We describe a Xilinx FPGA implementation of Neural Engineering Framework (NEF) networks with online learning that outperforms mobile Nvidia GPU implementations by an order of magnitude or more. Specifically, we provide an embedded Python-capable PYNQ FPGA implementation supported with a Xilinx Vivado High-Level Synthesis (HLS) workflow that allows sub-millisecond implementation of adaptive neural networks with low-latency, direct I/O access to the physical world. The outcome of this work is NengoFPGA, a seamless and user-friendly extension to the neural compiler Python package Nengo. To reduce memory requirements and improve performance we tune the precision of the different intermediate variables in the code to achieve competitive absolute accuracy against slower and larger floating-point reference designs. The online learning component of the neural network exploits immediate feedback to adjust the network weights to best support a given arithmetic precision. As the space of possible design configurations of such quantized networks is vast and is subject to a target accuracy constraint, we use the Hyperopt hyper-parameter tuning tool instead of manual search to find Pareto optimal designs. Specifically, we are able to generate the optimized designs in under 500 short iterations of Vivado HLS C synthesis before running the complete Vivado place-and-route phase on that subset, a much longer process not conducive to rapid exploration. For neural network populations of 64-4096 neurons and 1-8 representational dimensions our optimized FPGA implementation generated by Hyperopt has a speedup of 10-484× over a competing cuBLAS implementation on the Jetson TX1 GPU while using 2.4-9.5× less power. Our speedups are a result of HLS-specific reformulation (15× improvement), precision adaptation (3× improvement), and low-latency direct I/O access (1000× improvement).

Applied Reconfigurable Computing. Architectures, Tools, and Applications Morgan & Claypool Publishers

With the improvement in processing systems, machine learning applications are finding widespread use in almost all sectors of technology. Image recognition is one application of machine learning which has become widely popular with various architectures and systems aimed at improving recognition performance. With classification accuracy now approaching saturation point, many researchers are now focusing on resource and energy efficiency. With the increased demand for learning applications in embedded devices, it is of paramount importance to optimize power and energy consumption to increase utility in these low power embedded systems. In recent months, reduced precision neural networks have caught the attention of some researchers. Reduced data width deep nets offer the potential of saving valuable resources on hardware platforms. In turn, these hardware platforms such as Field Programmable Gate Arrays (FPGAs) offer the potential of a low power system with massive parallelism increasing throughput and performance. In this research, we explore the implementations of a deep learning architecture on FPGA in the presence of resource and energy constraints. We study reduced precision neural networks and implement one such architecture as a proof of concept. We focus on binarized convolutional neural network and its implementation on FPGAs. Binarized convolutional nets have displayed a classification accuracy of up to 88% with some smaller image sets such as CIFAR-10. This number is on the rise with some of the new architectures. We study the tradeoff between architecture depth and its impact on accuracy to get a better understanding of the convolutional layers and their impact on the overall performance. This is done from a hardware perspective giving us better insight enabling better resource allocation on FPGA fabric. Zynq ZCU-102 has been used for accelerator implementation. High level synthesis tool (Vivado HLS) from Xilinx is used for CNN definition on FPGA fabric.

Field-Programmable Logic and Applications Academic Press

This book lies at the interface of machine learning – a subfield of computer science that develops algorithms for challenging tasks such as shape or image recognition, where traditional algorithms fail – and photonics – the physical science of light, which underlies many of the optical communications technologies used in our information society. It provides a thorough introduction

to reservoir computing and field-programmable gate arrays (FPGAs). Recently, photonic implementations of reservoir computing (a machine learning algorithm based on artificial neural networks) have made a breakthrough in optical computing possible. In this book, the author pushes the performance of these systems significantly beyond what was achieved before. By interfacing a photonic reservoir computer with a high-speed electronic device (an FPGA), the author successfully interacts with the reservoir computer in real time, allowing him to considerably expand its capabilities and range of possible applications. Furthermore, the author draws on his expertise in machine learning and FPGA programming to make progress on a very different problem, namely the real-time image analysis of optical coherence tomography for atherosclerotic arteries.

Neural Information Processing Springer Science & Business Media

Thinking Machines: Machine Learning and Its Hardware Implementation covers the theory and application of machine learning, neuromorphic computing and neural networks. This is the first book that focuses on machine learning accelerators and hardware development for machine learning. It presents not only a summary of the latest trends and examples of machine learning hardware and basic knowledge of machine learning in general, but also the main issues involved in its implementation. Readers will learn what is required for the design of machine learning hardware for neuromorphic computing and/or neural networks. This is a recommended book for those who have basic knowledge of machine learning or those who want to learn more about the current trends of machine learning. Presents a clear understanding of various available machine learning hardware accelerator solutions that can be applied to selected machine learning algorithms Offers key insights into the development of hardware, from algorithms, software, logic circuits, to hardware accelerators Introduces the baseline characteristics of deep neural network models that should be treated by hardware as well Presents readers with a thorough review of past research and products, explaining how to design through ASIC and FPGA approaches for target machine learning models Surveys current trends and models in neuromorphic computing and neural network hardware architectures Outlines the strategy for advanced hardware development through the example of deep learning accelerators

The NEURON Book Springer Nature

Back-Propagation (BP) Algorithm is one of the efficient learning algorithms for the training of Artificial Neural Networks (ANN). The efficient hardware implementation of the BP Algorithm can find its application in the broad field of applications. The common computing platforms to build the BP algorithm based ANN Systems are Application Specific Integrated Circuits (ASICs) and General-Purpose Processors (GPP) based computers. However, due to a high demand of maintaining a trade-off between performance and flexibility, such computing machines become a bottleneck for further advanced improvements. In the last few decades, there has been significant progress in the field of Field Programmable Gate Arrays (FPGAs), which are based on the reconfigurable hardware platform. One of the main advantages of FPGAs are its flexibility, it is possible to reprogram the same hardware and achieve good performance by allowing parallel computation at the same time. The focus of this thesis is to implement the BP algorithm based ANN system on reconfigurable platform(FPGA). The proposed designs are coded on the software platform, MATLAB and in Verilog Hardware Description Language (Verilog HDL) on FPGA and synthesized on artix-7 FPGA evaluation kit. The validation of the design is verified on two benchmarks and comparisons are observed and discussed between two platforms.

Advances on P2P, Parallel, Grid, Cloud and Internet Computing Springer Science & Business Media

This thesis proposes to implement a new parallel convolutional binarized neural network (i.e. PC-BNN) on FPGA with accurate inference. The embedded PC-BNN is designed for image classification on CIFAR-10 dataset and explores the hardware architecture and optimization of customized CNN topology.

International Work-Conference on Artificial and Natural Neural Networks, IWANN'99, Alicante, Spain, June 2-4, 1999, Proceedings Springer Science & Business Media

This book provides a structured treatment of the key principles and techniques for enabling efficient processing of deep neural networks (DNNs). DNNs are currently widely used for many artificial intelligence (AI) applications, including computer vision, speech recognition, and robotics. While DNNs deliver state-of-the-art accuracy on many AI tasks, it comes at the cost of high computational complexity. Therefore, techniques that enable efficient processing of deep neural networks to improve metrics—such as energy-efficiency, throughput, and latency—without sacrificing accuracy or increasing hardware costs are critical to enabling the wide deployment of DNNs in AI systems. The book includes background on DNN processing; a description and taxonomy of hardware architectural approaches for designing DNN accelerators; key metrics for evaluating and comparing different designs; features of the DNN processing that are amenable to hardware/algorithm co-design to improve energy efficiency and throughput; and opportunities for applying new technologies. Readers will find a structured introduction to the field as well as a formalization and organization of key concepts from contemporary works that provides insights that may spark new ideas.

Design of a Neural Network for FPGA Implementation Springer Nature

This work was to establish whether it was possible to achieve a reasonable speedup by implementing FPGA based Hopfield neural networks for some simple constraint satisfaction problems. The results are significant - our initial implementation using standard Xilinx FPGAs yielded 2-3 orders of magnitude speedup over the Sun Blade 2000 workstation comes with 1.2-GHz version of the 64-bit UltraSPARC III Cu processor. The main problem with the work to date is that the problems are both unrealistically small and simplistic. That is the constraints on the N-Queen problem are simpler than those found in many real world scheduling applications. Thus, it is not clear whether we will be able to optimize the neuron structure for more complex problems since the weights matrix may not contain as many zero elements. Thus a new method for speed improvement of Hopfield neural networks for solving constraint satisfaction problems using Field Programmable Gate Arrays (FPGAs) was proposed and implemented.