論文 / 著書情報
Article / Book Information

題目(和文)	RISC-VをベースとしたAI画像認識システムのための特定用途向け命令セットプロセッサ
Title(English)	Application-Specific Instruction-set Processor for AI Visual System based on RISC-V
著者(和文)	汪漢森
Author(English)	Hansen Wang
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第12861号, 授与年月日:2024年9月20日, 学位の種別:課程博士, 審査員:一色 剛,髙橋 篤司,本村 真人,佐々木 広,藤木 大地
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第12861号, Conferred date:2024/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

（博士課程）
Doctoral Program

論 文 要 旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報通信 系 コース	申請学位（専攻分野）： Academic Degree Requested	博士 (工学) Doctor of
学生氏名： Student's Name	WANG HANSEN	審査員主査： Chief Examiner	一色 剛

要旨（英文 800 語程度）
Thesis Summary (approx.800 English Words)

Computer vision (CV) and artificial intelligence (AI) have become pivotal technologies, finding applications across numerous fields. In healthcare, AI-driven CV systems assist in diagnostics by analyzing medical images, detecting abnormalities, and guiding surgical procedures. In agriculture, they enable precision farming by monitoring crop health, detecting pests, and optimizing resource usage. Surveillance systems leverage these technologies for real-time monitoring and threat detection, enhancing security in public and private spaces. Moreover, AI-powered CV is transforming industries such as automotive with advanced driver-assistance systems (ADAS) and autonomous vehicles, retail with automated checkouts and inventory management, and manufacturing with quality control and predictive maintenance.

The conventional approach to AI model implementation involves using GPUs with APIs like Compute Unified Device Architecture (CUDA). However, GPUs are power-hungry compared to FPGAs and Application-Specific Integrated Circuits (ASICs), making them unsuitable for low-power applications like edge and IoT devices. FPGAs, configurable to execute various digital functions including CNN inference tasks, offer power efficiency by instantiating only the necessary computational elements. However, FPGA programming requires specialized skills and significant time for performance tuning, limiting their flexibility. ASICs, designed for specific tasks, provide higher efficiency and lower power consumption than GPUs and FPGAs, making them ideal for performance-critical applications. Nevertheless, ASICs lack flexibility and require substantial time and resources for design and fabrication.

To address these challenges, we introduce an energy-efficient implementation of the computer vision AI model application-specific instruction-set processor based on RISC-V. The RISC-V architecture, chosen for its open and modular nature, enables scalable implementations of custom instructions. Compared to other AXI-connected architectures, our deeply integrated overall design utilizes custom instructions to directly control the CE, enabling on-chip YOLO execution, including post-processing. Furthermore, while many other ASIC or FPGA architectures tailored for CNNs only support object detection, our design adeptly handles YOLOv8 Object Detection, Instance Segmentation, and Pose Detection. Hence, we name it as Artificial Intelligence Visual System on Chip (AIV-SoC).

The AIV-SoC, grounded in the RISC-V architecture and utilizing the LLVM-C2RTL toolkit, incorporates several optimization techniques such as neural network layer scheme optimization, piecewise linear function approximation, multi-cycle path, 8-bit fixed-point data per-group quantization, and 4-pattern transposed convolution. These optimizations enhance both the versatility and power efficiency of the SoC, demonstrating superior performance compared to GPU and FPGA platforms. It can deal with YOLOv8s with a 352x352 input image resolution and achieve impressive frames per second (FPS) rates of 67.1, 55.2, and 64.9 for object detection, instance segmentation, and pose detection, respectively. Hence, it is suitable for deployment in resource-constrained environments like healthcare, agriculture, and surveillance.

The AIV-SoC designed in this work also supports various RESNET and has been tested using the ImageNet2012 (ILSVRC 2012) dataset in simulations. The RESNET is a widely recognized classic deep neural network (DNN) architecture for image classification, comprising convolution layers (CL), max-pooling layers, shortcut operations, fully connected (FC) layers, and average-pooling layers. These layers all involve significant amounts of loops, and the underlying principle of accelerator acceleration is essentially the unrolling of these loops. We have also summarized a set of theories

on the loop unrolling of various AI layers, which can be used to describe the working principle of accelerators more simply. This is helpful for facilitating communication among chip designers, model deployment engineers, and model trainers, and for optimizing the overall system macroscopically and systematically.

The YOLOv8 model employed in this implementation is a real-time object detection algorithm that divides the input image into a grid and predicts bounding boxes and class probabilities directly within each grid cell. The YOLOv8 version used in this study, developed by Ultralytics, is notable for its balance between accuracy and real-time inference speed. In this work, the YOLOv8 model is extended to support instance segmentation and pose detection. Instance segmentation, which provides pixel-level delineation of object boundaries, is achieved using the YOLOv8 instance segmentation model (yolov8s-seg.pt). For pose detection, which involves estimating the spatial arrangement of key body joints, the YOLOv8 pose model (yolov8s-pose.pt) is used.

In the object detection simulations, we used the COCO dataset (with a mAP loss of 0.8%) and the global wheat 2000 dataset (with a mAP loss of 1.9%). For instance segmentation simulations, we used the COCO dataset (with a mAP loss of 0.8%) and the ROBOFLOW universe crack dataset (with a mAP loss of 0%). In the pose detection simulations, we used the COCO dataset (with a mAP loss of 5.4%).

Additionally, this work explores the implementation of activation functions, such as SiLU and softmax, in hardware using piecewise linear function approximation. This approach offers speed and area efficiency advantages over traditional lookup table or polynomial approximation methods. It effectively addresses the increasing demand for diverse activation functions, providing valuable insights for the design of similar embedded devices.

In conclusion, this thesis presents a comprehensive approach to designing a flexible, energy-efficient, and highly usable SoC for parallelism-tolerant compute-intensive tasks. The combination of RISC-V architecture, YOLOv8 model optimizations, and advanced hardware design techniques provides a versatile platform suitable for various applications, from autonomous vehicles to medical image analysis. Our design also exhibits high scalability, paving the way for future advancements in AI-driven visual systems. It demonstrates the potential of integrating cutting-edge machine learning models with innovative hardware solutions.