Geng Yang

Email: gengyang@stu.xidian.edu.cn

Address: No.2 Taibai South Road, Xi'an, 710071, Shaanxi, China

EDUCATION

- **Ph.D. Candidate in Information and Communication Engineering** Sep. 2019– Dec. 2024 (Exp.)
  State Key Lab of Integrated Service Networks, Xidian University, Xi'an, Shaanxi, China
  Advisor: Prof. Yunsong Li and Prof. Jie Lei
  (**direct admission recommendation**)
  GPA:90.14/100
- **B.S. in Communication Engineering** Aug. 2015 – Jun. 2019
  School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi, China
  GPA: 3.7 (4.0)

RESEARCH INTERESTS

- FPGA-based Edge/Large Model Inference Acceleration
- Algorithm-Hardware Co-design and Automation

SKILLS

- HLS C, Python, Verilog, Matlab, Latex

RESEARCH SUMMARY

- **Two National-level Contests:**
  - Achieved one second prize and one third prize
- **Four conference Papers (top tier in FPGA filed):**
  - First author on two CCF-C conference papers
  - First author on one CCF-B conference poster and one CCF-C conference poster
- **Six Publication Papers:**
  - First author on one SCI-Q1 journal **(top tier in remote sensing filed)**
  - one SCI-Q4 journal **(top tier in remote sensing filed)**
  - First student author on one SCI-Q2 journal
  - Co-author on three SCI-Q1 journals
- **Three National Invention Patents**
  - First student author on one patent
  - Third student author on two patents
- Five Significant Research Projects
  - Led one international collaborative project as the student leader
  - Led two national projects as the student leader
  - Led one project funded by the National Natural Science Foundation of China as the student leader
  - Directed two projects funded by the Xidian University Innovation Fund

RESEARCH OVERVIEW

My PhD research focuses on developing deep learning-based intelligent processing methods for high-resolution Earth observation and other significant aerospace remote sensing applications. The main goal is to tackle the challenges associated with high-performance and automated deployment of DNN models to FPGA-based hardware architectures under stringent resource limitations:

- **In 2019**, based on **Project f**, I proposed a fast hyperspectral anomaly detection algorithm and an effective deeply pipelined hardware architecture (**Paper 7 and Patent 8**). Some of the computational operators perform better than the official AMD-Xilinx HLS operators.
- **In 2021**, I pivoted my research to deep learning hardware acceleration and led two award-winning projects in national competitions (**Contest 1 and 2**). My primary responsibilities included designing compact models, optimizing hardware-friendly pruning and quantization, and implementing high-performance hardware architecture. I played a crucial role in overall coordination and the core system design. Our team successfully deployed high-accuracy CNNs on domestic cloud and embedded FPGA platforms. we launched **Projects d and e** during the same year to transform our competition achievements into practical applications.
- **In 2022**, I proposed OSCAR-RT, the first end-to-end co-design framework for on-satellites SAR ship detection, integrating SAR-aware CNN model adapting, hardware-guided progressive pruning, mixed-precision quantization and fully pipelined interlayer streaming accelerator (**Paper 6**). OSCAR-RT enabled efficient

deployment of CNN models like MobileNetV1, MobileNetV2, and SqueezeNet on the AMD-Xilinx VC709 FPGA. Additionally, I initiated **Project c** and pursued research leading to **Patents 9 and 10** that year.

- **In 2022**, I further explored hardware deployment of binary neural networks with extreme 1-bit weight and activation (**Paper 5 and Project b**). This work is the first to quantify and mitigate the hardware inefficiency in SOTA BNNs, which are mainly caused by various AFP components and increased model size that were proposed for accuracy gains. We ultimately proposed HyBNN, the first hybrid BNN and 4-Bit-Net design that directly binarizes (and quantizes) the original depthwise separable convolution (DSC) blocks to keep compact model size and high model accuracy.
- **Since 2023**, beyond my previous work on all-on-chip fully pipelined hardware architectures, I have further developed E4SA: low-bit systolic array-based universal architecture (**Paper 4)**. Leveraging the exceptional performance of E4SA, I have collaborated on international **Project a** with Professor Zhenman Fang's team at Simon Fraser University in Canada and Professor Yanzhi Wang's team at Northeastern University in the USA. This project focuses on accelerating Stable Diffusion on edge FPGA platforms. Our latest papers on this project are accept to the FPL 2024 conference (the first exploration in this filed: **Paper 1 and 2)**.

PUBLICATIONS

1 **Geng Yang**, Yanxue Xie, Zhongjia Xue, Sung-En Chang, Yanyu Li, Peiyan Dong, Jie Lei, Weiying Xie, Yanzhi Wang, Xue Lin, Zhenman Fang, SDA: Low-Bit Stable Diffusion Acceleration on Edge FPGAs.The 34nd International Conference on Field-Programmable Logic and Applications (***FPL2024***)

2 **Geng Yang**, Jie Lei, Zhenman Fang, Jiaqing Zhang, Junrong Zhang, Weiying Xie, Yunsong Li, SA4: A Comprehensive Analysis and Optimization of Systolic Array Architecture for 4-bit Convolutions.The 34nd International Conference on Field-Programmable Logic and Applications (***FPL2024***)

3 Xingyu Tian, **Geng Yang**, Zhenman Fang, FLUD: A Scalable and Configurable Systolic Array Design for LU Decomposition on FPGAs.(under review)

4 **Geng Yang**, Jie Lei, Zhenman Fang, Jiaqing Zhang, Junrong Zhang, Weiying Xie, Yunsong Li, E4SA: An Ultra-Efficient Systolic Array Architecture for 4-Bit Convolution Neutral Networks.The 32nd ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (***FPGA2024*** poster)

5 **Geng Yang**, Jie Lei, Zhenman Fang, Yunsong Li, Jiaqing Zhang, Weiying Xie, HyBNN: Quantifying and Optimizing Hardware Efficiency of Binary Neural Networks. ACM Transactions on Reconfigurable Technology and Systems (***TRETS 2023***, ***FPT 2023*** Journal Track, ***FCCM 2023*** Poster)

6 **Geng Yang,** Jie Lei, Weiying Xie, Zhenman Fang, Yunsong Li,Jiaxuan Wang, XinZhang, Algorithm/Hardware Co-Design for Real-Time On-Satellite CNN based Ship Detection in SAR Imagery. IEEE Transactions on Geoscience and Remote Sensing, 2022 (***IEEE TGRS***)

7 Jie Lei, **Geng Yang**, Weiying Xie, Yunsong Li, Xiuping Jia, A Low-Complexity Hyperspectral Anomaly Detection Algorithm and its FPGA Implementation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 14:907-921 (***IEEE JSTARS***)

8 Jie Lei, **Geng Yang**, Mengbo Zhang, Weiying Xie, Yunsong Li, Tao Jiang, Kai Liu, Long Gao, FPGA-based Hyperspectral Anomaly Detection System, ***Patent***: CN202110484719.7 on April 30, 2021

9 Jie Lei, Jiaxvan Wang, **Geng Yang**,Weiying Xie, Yunsong Li, Object Detection Joint Pruning and Quantization for SAR Imagery, ***Patent***, CN202111488427.7 on December 8, 2021

10 Jie Lei, Yi Guo, **Geng Yang**,Weiying Xie, Yunsong Li, Object Detection Based on Feature Fusion with Masked Networks for SAR Imagery, ***Patent***: CN202211567684.4 on October 7, 2022

11 Jiaqing Zhang, Jie Lei, Weiying Xie, **GengYang**, Daixun Li, Yunsong Li. Multimodal Informative ViT: Information Aggregation and Distribution for Hyperspectral and LiDAR Classification, IEEE Transactions on Circuits and Systems for Video Technology, 2024 (***IEEE TCSVT***)

12 Xin Zhang, Weiying Xie, Yunsong Li, Jie Lei, Qian Du, **Geng Yang**, Rank-Aware Generative Adversarial Network for Hyperspectral Band Selection. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60:1-12 (***IEEE TGRS***)

13 Jiaqing Zhang, Jie Lei, Weiying Xie, Yunsong Li, Xiuping Jia, **Geng Yang**. Guided Hybrid Quantization for Object Detection in Multimodal Remote Sensing Imagery via One-to-one Self-teaching. IEEE Transactions on Geoscience and Remote Sensing, 2022 (***IEEE TGRS***)

CONTESTS

① National contest for DNN model inference on an cloud FPGA system supervised by Prof. Jie Lie, the **second prize** and **¥3,690,000** project funding, **Student Leader**, Jun. 2021

② National contest for DNN model inference on an embedded FPGA system supervised by Prof. Weiying Xie, the **third prize** and **¥950,000** project funding, **Student Leader,** Jun. 2021

PROJECTS

a. International Collaborative Project: **Stable Diffusion Acceleration on Edge FPGA,** hardware architecture design lead**,** Oct. 2023 – May. 2024

b. Xidian Univerisity Innovative Fund Project: **Algorithm Design and Hardware Acceleration of Binary Neutral Network for Real-time On-satellite Applications** (¥15,000), lead, May. 2023 – May. 2024

c. Xidian Univerisity Innovative Fund Project: R**esearch on DNN Algorithm and Hardware Architecture Design Methods for Remote Sensing Applications** (¥15,000), lead, May. 2022 – May. 2023

d. National Project: **Research on DNN Model Compression Algorithms for XXX** (¥950,000), student lead, Dec. 2020 – Dec. 2022

e. National Project: **Deep Neural Network Design and FPGA Acceleration Technology for XXX,** (¥3,650,000), student lead, Dec. 2020 – Dec. 2022

f. National Natural Science Foundation of China, **Research on a Novel Hyperspectral Target Detection Algorithm to Be Implemented on Satellite for On-board Real-time Data Processing**(¥630,000), hardware architecture design lead, Jan. 2021 – Dec. 2024

AWARDS

- GuoRui Scholarship of the 14th Research Institute of China Electronics Technology Group Corporation in 2021,2022
- Second Scholarship of Xidian University in 2020,2021,2022,2023,2024
- Collaboration-Innovation Scholarship of China Electronics Technology Group Corporation and Xidian University in 2020
- Outstanding Graduate Student of Xidian University in 2019

OTHER HIGHLIGHTS

- Conference secondary reviewer for FPL2024, FPGA 2024, DAC 2024, DAC 2023, DATE2024
- Oral presentation titled *HyBNN: Quantifying and Optimizing Hardware Efficiency of Binary Neural Networks* at International Conference on Field-Programmable Technolohy (***FPT 2023***, Yokohama, Janpan) in Dec. 2023.
- Oral presentation titled *Hyperspectral Anomaly Detection Algorithm and Its Hardware Implementation for on-satellite Real-time Processing* **(Excellent Report)** at the 6th National Imaging Spectroscopic Earth Observation Symposium (Ningbo, China) in Oct. 2021.
- Awarded with **High Distinction** in Academic English and Research Communication Skills of The University of Adelaide (online) in 2021