**Formula 02**

|  |  |
| --- | --- |
| VIETNAM NATIONAL UNIVERSITY, HANOI  **INFORMATION TECHNOLOGY INSTITUTE** | **VIETNAM SOCIALIST REPUBLIC**  **Independence – Freedom – Happiness** |

**RECRUITMENT INTERN  
FPGA & Embedded system design engineer**

**Title: Real-Time Vietnamese and Japanese ASR on FPGA Using QuartzNet12x1 and Vitis AI**

* **Job Description**

In today’s world, Large Language Models (LLMs) play a vital role in many domains. For Automatic Speech Recognition (ASR), QuartzNet12x1 is a state-of-the-art convolutional neural network (CNN)-based model designed to convert speech into text. It is typically trained on paired audio–text datasets (with ground truth transcripts), enabling accurate recognition across different languages. QuartzNet12x1 is lightweight yet powerful, making it suitable for deployment on edge devices such as the Xilinx Kria KV260, while still maintaining strong accuracy.

Traditionally, ASR models are deployed on GPU-powered servers, where audio input is sent over the network for inference. This approach, while effective, introduces latency and requires significant bandwidth. In contrast, edge computing enables ASR to run locally, reducing latency, bandwidth usage, and dependence on network connectivity. This is especially valuable in real-world applications such as smart cars or autonomous robots, where real-time processing and reliability are critical.

This project aims to develop Vietnamese and Japanese ASR systems on FPGA, addressing challenges in efficiency and latency. The ASR model will be optimized and implemented on the Xilinx Kria SOM KV260 development kit to achieve real-time operation. A hardware/software co-design approach will be employed to accelerate ASR at the edge, ensuring faster, more energy-efficient, and reliable performance.

**Objective:**  
The project's primary goal is to accelerate Quartznet12x1 by Vitis-AI framework when deployed into Xilinx SOM Kria-260, including the followings:

* Quantize QuartzNet12x1 using Vietnamese and Japanese speech datasets with ground truth transcripts.
* Optimize and transform the model into a hardware-friendly version suitable for FPGA deployment.
* Ensure the system meets requirements for latency, memory efficiency, energy consumption, flexibility, and reliability in real-time ASR applications.

**Keywords**: Auto Speech Recognition; Transformers; Quartznet12x1; Convolutional neural networks; Edge-AI; Internet-of-Things; Field Programmable Gate Array.

* **Project Supervision**

The project is supervised by Prof. Tran Xuan Tu and Dr. Bui Duy Hieu. Prof. Xuan-Tu Tran and Dr. Duy-Hieu Bui have been working on hardware design and accelerations for FPGA since 2010. They have my publications and patents specifically on hardware design for image processing and security. This work is of interest in designing HW accelerators for edge devices.

* **Candidate Profile**

**Education:**

* Student in related topics with GPA >= 3.4/4.0, Highly motivated to expand expertise in FPGA and AI acceleration.

**Research Experience**:

* Having experiences with FPGA development flow (including Vivado/Vitis/Vitis-HLS)
* Having experiences working with Xilinx KV260 devboard and Vitis AI is a plus.
* Having experiences working with embeded system programmin and realtime embebed system design.

**Technical Skills**:

* Programming: Python, C++ and hardware description languages such as VHDL, Verilog or SystemVerilog.
* Basic knowledge on neural networks, FPGA fundamentals, Linux systems.
* Experience with ARM programming, electronic circuit design is a plus.
* Strong teamwork, independent research, and self-learning ability.
* **Bibliography**

1. Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, Michael Auli, “Quartznet12x1: A Framework for Self-Supervised Learning of Speech Representations,” arXiv archive, 2020, <https://doi.org/10.48550/arXiv.2006.11477>.
2. Santosh Gondi, “Wav2Vec2.0 on the Edge: Performance Evaluation,” arXiv archive, 2022, <https://doi.org/10.48550/arXiv.2202.05993>.
3. Y. Fukuda, K. Yoshida and T. Fujino, "Evaluation of Model Quantization Method on Vitis-AI for Mitigating Adversarial Examples," in IEEE Access, vol. 11, pp. 87200-87209, 2023, doi: 10.1109/ACCESS.2023.3305264.
4. Zhengdong Li, Frederick Ziyang Hong, C. Patrick Yue, “FPGA-based Acceleration of Neural Network for Image Classification using Vitis AI,” Arxiv archive 2024, <https://doi.org/10.48550/arXiv.2412.20974>.
5. Tianqi Chen, Thierry Moreau, Ziheng Jiang, Lianmin Zheng, Eddie Yan, Meghan Cowan, Haichen Shen, Leyuan Wang, Yuwei Hu, Luis Ceze, Carlos Guestrin, and Arvind Krishnamurthy. 2018. TVM: an automated end-to-end optimizing compiler for deep learning. In Proceedings of the 13th USENIX conference on Operating Systems Design and Implementation (OSDI'18). USENIX Association, USA, 579–594.

* **Internship location:** Building E3, VNU Information Technology Institute, 144 Xuan Thuy, Cau Giay, Hanoi.
* **How to apply:** Interested candidates, please email your CV and academic record to hieubd [at] vnu.edu.vn. Email subject: [INTERN] - [Position name] - [Fullname]