

Nghiên cứu và thiết kế phần cứng cho Convolutional neural network

Võ Phan Hoàng Kha

Trường Đại học Công nghệ thông tin TP Hồ Chí Minh

Trần Quốc Lượng

Trường Đại học Công nghệ thông tin TP Hồ Chí Minh

What ?

Chuyển đổi thành công một mô hình Convolutional neural network bằng ngôn ngữ lập trình cấp cao (Python) sang ngôn ngữ phần cứng như Verilog, để tạo IP (Intellectual Property) phần cứng để chạy trên FPGA (Field-Programmable Gate Array).

Why ?

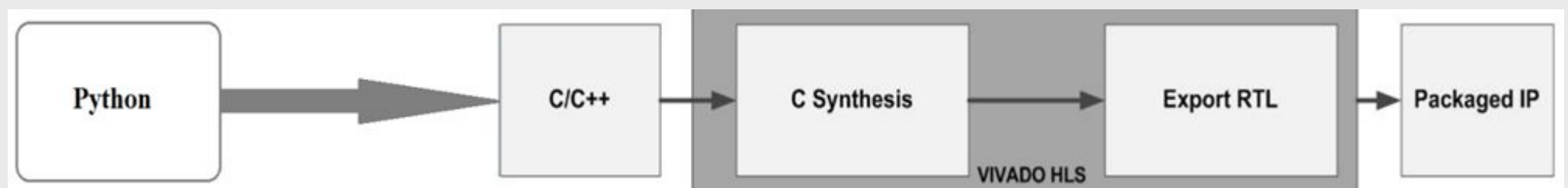
Các mô hình học sâu, đặc biệt là Convolutional neural network, cần một lượng lớn tài nguyên. Trong khi đó, FPGA (Field-Programmable Gate Array) cho phép triển khai mô hình học sâu ở diện tích nhỏ mà vẫn đảm bảo tốc độ và độ chính xác. Do đó, đề tài sẽ giúp tối ưu thời gian xử lý, năng lượng tiêu thụ, tài nguyên tính toán so với các mô hình học sâu khác.

Overview

Chuyển đổi từ Python sang C/C++

Công cụ Vivado HLS

IP phần cứng



Hình 1: Quy trình chuyển đổi từ Python sang ngôn ngữ phần cứng dưới dạng IP

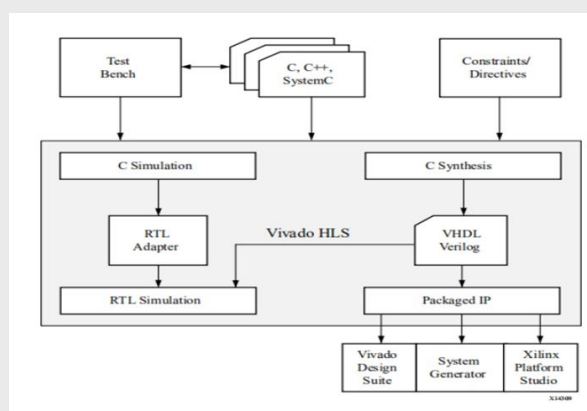
Description

1. Chuyển đổi từ Python sang C/C++

- Nghiên cứu và mô phỏng lại hoạt động của mô hình bằng C/C++ với các trọng số và hiệu số được lấy từ Python.

2. Công cụ tổng hợp cấp cao Vivado HLS

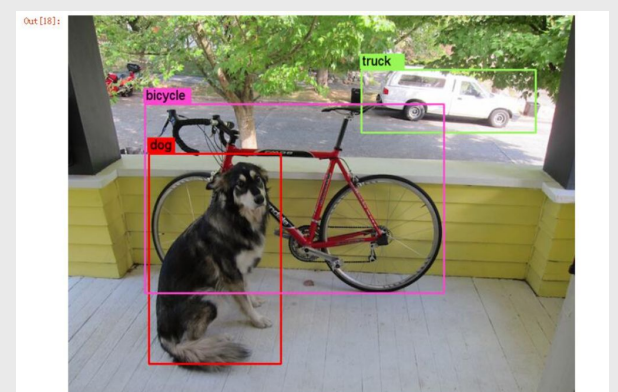
- Là một phần mềm của Xilinx, cho phép người sử dụng tạo ra các IP (Intellectual Property) dưới dạng ngôn ngữ mô tả phần cứng (HDL - Hardware Description Language)
- Sau khi thành công xây dựng mô hình bằng C/C++ từ các tham số từ Python. Mô hình sẽ được đưa vào công cụ Vivado HLS để được chuyển thành IP phần cứng.



Hình 2: Quy trình thiết kế, đầu ra, đầu vào của Vivado HLS

3. IP Phần cứng

- Sau khi đã có IP phần cứng, việc mô phỏng để kiểm tra sẽ được tiến hành để đảm bảo tính đúng đắn cho IP.
- Tiếp đến, IP này sẽ được mô phỏng trực tiếp trên FPGA, kiểm tra thời gian thực hiện, tính toán hao tổn tài nguyên.



Hình 3: Minh họa kết quả phân loại vật thể của IP phần cứng