

Hardware Accelerated Human Pose Tracking

Acceleration of deep learning algorithms with FPGAs for high performance and edge applications

Student



Michael Schmid

Introduction: With the advent of Artificial Intelligence (AI) and the ever improving neural networks, many new and unimaginable challenges can be overcome. In this paper, we focus on the task of Human Pose Estimation (HPE). The latest deep learning models combined with state-of-the-art hardware can perform HPE in real time.

HPE refers to the estimation of a kinematic human body model from images.

Swiss startup VRMotion has developed a pose tracking system to localize the 3D position of the human pose for their virtual-reality flight simulator.

A multi-camera system with a world-class and computationally expensive deep learning model runs on multiple graphics processing units (GPUs) with sufficient performance. To introduce HPE directly into other possible applications, a small and low-cost system is desirable. Therefore, the pose tracking system should run on edge devices.

The Interdisciplinary Center for Artificial Intelligence (ICAI) at the OST, together with VRM Switzerland, has developed a new, computationally effective neural network for HPE called ICAIPose. In this work ICAIPose is to be implemented on FPGA edge devices.

Approach: With the newly introduced Field Programmable Gate Arrays (FPGAs) for AI applications from Xilinx, the market leader in FPGAs, new opportunities arise.

The main goal of this is to show how to implement ICAIPose on the high-performance Adaptive Compute Acceleration Platform FPGA VCK190 and the KV260 edge device. This requires a camera interface for the two hardware platforms. In order to test this system thoroughly, a given deep learning model was run with the cameras. Xilinx's Vitis AI deep learning framework allowed to implement ICAIPose on the newly built FPGA platform with the cameras. The ICAIPose network could be compiled for the Deep-Learning Processor Unit (DPU) on the FPGA with minor adjustments to the network. Thanks to the included Vitis AI Runtime Engine with its easy-to-use Python API, communication with the DPU is done via an embedded Linux on the FPGAs microprocessor.

Conclusion: First of all, ICAIPose runs on both FPGA platforms with the proposed camera interfaces. ICAIPose runs at 27 frames per second (fps) on the VCK190 and 8 fps on the KV260 in its original configuration.

The rather modest performance of the KV260 is due to the large ICAIPose neural network. On a GPU based edge device, the NVIDIA Jetson Xavier NX the performance is also 8 fps.

The Vitis AI framework from Xilinx has been extensively tested and shows its strengths, but also some teething troubles. For running deep neural networks on FPGAs, Vitis AI is by far the best

developed framework and should be considered before implementing custom hardware-accelerated algorithms for deep learning.

L: Input Image with the drawn pose, detected by ICAIPose. R: Confidence map of the 15 keypoints: (head, left knee etc.)
Own presentation



Pose estimation with an example from a used dataset. Different networks implemented on GPU and FPGA
Own presentation



Examiner

Prof. Dr. Paul Zbinden

Subject Area

Electrical Engineering

Project Partner

VRM-Switzerland,
Dübendorf, Zürich