



Contribution ID: 18

Type: **Lightning Talk**

AI Upscaling with Super Resolution CNNs on FPGAs and ASICs

Wednesday 27 September 2023 17:25 (5 minutes)

Deep learning techniques have demonstrated remarkable performance in super resolution (SR) tasks for enhancing image resolution and granularity. These architectures extract image features with a convolutional block and add the extracted features to the upsampled input image transported through a skip connection, which is then converted from a depth to higher resolution space. However, SR can be computationally expensive due to large three-dimensional inputs with outputs many times larger while demanding low latency, making large scale implementation in commercial video streaming applications challenging. To address this issue, we explore the viability of SR deployment on-chip to FPGA and ASIC devices for low latency and low power inference. We train and optimize our model using a range of techniques, including quantization-aware training, batch normalization, heterogeneous quantization, and FIFO depth optimization to achieve an implementation which fits within our resource and accuracy constraints. Using the DIV2K diverse image dataset and supplying input images which are downsampled by a factor of three, we achieve >30 PSNR at $>2b$ quantization. We use this initial FPGA implementation as a proof of concept for future ASIC implementations.

Authors: HAWKS, Ben (Fermi National Accelerator Lab); DI GUGLIELMO, Giuseppe (Fermilab); CAMPOS, Javier; MITREVSKI, Jovan (Fermi National Accelerator Lab. (US)); MUHIZI, Jules; TRAN, Nhan (Fermi National Accelerator Lab. (US)); FORELLI, Ryan (Lehigh University)

Presenter: FORELLI, Ryan (Lehigh University)

Session Classification: Contributed Talks

Track Classification: Contributed Talks