Minutes of the CRG meeting #1

Date 29/11/2022 **Location Bld. 2, 2290**

Present

Attendees: Prof. Ahmed Eltawil, Prof. Suhaib Fahmy, Prof.Fadi Kurdahi, Prof. Khaled Salama, Mojtaba Alshams, Olga Krestinskaya, Hao Liu, Deigo Silva, Kamilya Smagulova and Li Zhang

Online: Dr. Mohammed Fouda

Agenda

Overall progress to date and student presentations

1. Brief introduction for annual meeting and milestones by Prof. Eltawil.
2. Overall progress to date and different students roles and progress by Dr. Fouda.
3. Presentation by students:
   - Diego presented a state-of-the-art frameworks on Autonomous driving as well as the progress on object detection using Yolov5 model on event-based data
   - Mojtaba presented five different Distributed Learning (DL) topologies and four synchronization techniques. .
   - Hao presented task offloading with compute-bound.
   - Olga presented Hardware-Aware Network Architecture Search framework adapted for In-Memory computing designs.
   - Li demonstrated results on the kernel-wise mixed precision for LeNet and VGG-small networks.

After PIs discussion of each presentation, the following suggestions and considerations were made:

These action items and suggestions will be followed on over the next two month with a second meeting targeted for end of January. Each item will be discussed in the meeting with expected follow-up and suggestions.

1. Diego:
   a. To estimate the number of FLOPS (and power consumption) and other hardware-related parameters during the inference phase.
   b. To evaluate how much data comes from each sensor.
   c. To compare the performance of a low level signal processing function (filters) and neural network in the Head.
   d. To consider implementation of Multi-task Transformer NN instead of CNN, since consistency of matrix size in all layers can allow efficient utilization of hardware resources (with less number of steps compared to non-uniform CNN layers).
   e. To get familiarized with UCI Chauffeur Service (in particular, a pixel to steering) and ZOOX L5.

2. Mojtaba:
    a. Advantages and disadvantages of each DL topology should be summarized based on Autonomous Driving application features.
    b. The proposed distributed learning topology should be resilient to the scenario when a car is out of a network.
    c. The DL should consider both – (a) a full training from scratch and (b) a scenario for updating pretrained models with personalization and fine-tuning on the edge (car configurations, location-based training, incremental training, adaptive training).
    d. To consider that the transfer learning will depend on the geography.
    e. To analyze Checkpoints vs. Communication cost
    f. The CRG project assumes heterogeneous system. To contact Adilya for information on the customizable code designed for heterogenous devices and links.
    g. To consider topology for an alternative application in healthcare.

3. Hao
    a. Memory-bound was not considered. The system should has no delays to memory access.
    b. To cooperate with Olga (NAS) in order to identify how many split points to include and what are the best places.
    c. To consider precision dimension
    d. To consider the design of longer but splitable network.
    e. To assume the scenario when AP saturated by other users or links are down.
    f. Simple and long tail – entropy for complex tail

4. Olga
    a. Predictor is the key element in the algorithm and it is important to ensure its reliability
    b. Small case study to validate the algorithm. To take sampled model from the space and to use hardware simulator to estimate accuracy degradation, to list applied modifications.
5. Li:
    a. The presented results should be improved.
    b. To perform frame analysis.
    c. To consider implementation of Mixed Precision within neural network (e.g. lowest possible at head, tail) and to correlate it with data.
    d. Dynamically change quantization to improve energy efficiency (e.g. like DVFS but with precision).
    e. To cooperate with Olga (NAS) and Hao (split computing) to analyze precision at split cuts.

Minutes of the CRG meeting #2

Date 30/11/2022 **Location Bld. 2, 2290**

Present

Attendees: Prof. Ahmed Eltawil, Prof. Suhaib Fahmy, Prof.Fadi Kurdahi, Prof. Khaled Salama, Mojtaba Alshams, Olga Krestinskaya, Hao Liu, Deigo Silva, Kamilya Smagulova and Li Zhang

Online: Rachid Karami, Mariam Rakka

Agenda

Student presentations and PI discussion

1. Mariam presented the programming module -compiler.
2. Rachid presented the AP and high-level control flow.
3. Kamilya presented the summary on Distributed Collaborative Machine Learning (DCML) such as federated learning, split learning, splitfed learning.
4. PI discussion

During discussion the following suggestions were made by PIs:

1. Mariam:
   a. To check if Neurosim already supports Mixed Precision
   b. Mariam and Olga should coordinate and have discussion, in particular on Neurosim functions
   c. Similarly to SIAM, to consider different interconnects between tiles
2. Rachid:
   a. To consider local memory within AP (without global bus).
   b. Latency in the system (technique to take advantage.
   c. To check the packet-processing in RISC-V ISA for acceleration.
   d. Diego and Rachid: discussion should include structure of the data; efficient memory organization for application, transformers architecture.
   e. To check the ARM and ARCH (??) case – to check the reason it failed.
   f. Pipeline of work: global, localized and transfer learning.
3. Kamilya:
   a. To align features of Split Learning with the CRG project requirements.
   b. To identify the best split cut considering both model and hardware resources (and number of clients/cars): cooperation with Hao (split computing), Olga (NAS) and Diego (model and dataset)

To start writing the review papers and to start to work on identified gaps and with each other in overlapping areas of the project. The next follow-up meeting is planned to be held in January.