Contribution ID: **47**   Type: **Poster**

The performance profiling of ptycho-W1Net AI algorithm on DCU and HUAWEI NPU 910

Thursday, 14 March 2024 16:10 (30 minutes)

APS at USA completed the high-precision training and inference in Nvidia GPU clusters taking the ptychoNN algorithm combined with ePIE Conjugate Gradient method. By the reference of that idea, we came up with a new model called W1-Net whose training speed was faster with higher precision of inference. After this development, we implemented the model onto DCU cluster. However, the performance was only 1/6 of Nvidia GPU A100. Profiling action was done to the training process and the low speed was caused by the atom operation during the training function. After tuning the code, the training time was reduced by half then the previous model. Apart from DCU, we also trained on HUAWEI NPU card. This paper will show the profiling result of HUAWEI NPU 910*8 cluster.

Significance

The training process is on the heterogeneous computing card on HUAWEI Ascend 910 which is different from Nvidia and training speed can be comparable to Nvidia GPU A100.

References

Title: 'W1-Net:A fast training and highly scalable ptychography convolutional neural network'. This paper is underreview by 'The European Physical Journal Plus'.

Experiment context, if any

The data comes from this website: https://github.com/mcherukara/PtychoNN

Primary authors: Dr WANG, Lei (Institute of High Energy Physics); Dr MU, Yangyang (Institute of High Energy Physics)

Co-authors: Dr XING, Chengye (Institute of High Energy Physics); Dr CHANG, Guangcai (Institute of High Energy Physics); Dr SHI, Jingyan (Institute of High Energy Physics); Dr HU, Jiarui (Institute of High Energy Physics); Dr HU, Yu (Institute of High Energy Physics); Dr LIU, Jianli (Institute of High Energy Physics); Dr SUN, Haokai (Institute of High Energy Physics); Dr LIU, Rui (Institute of High Energy Physics); Dr FU, Shiyuan (Institute of High Energy Physics); Dr QI, Fazhi (Institute of High Energy Physics); Dr CHENG, Yaodong (Institute of High Energy Physics)

Presenter: Dr WANG, Lei (Institute of High Energy Physics)

Session Classification: Poster session with coffee break

Track Classification: Track 1: Computing Technology for Physics Research