

Parareal neural networks emulating a parallel-in-time algorithm

Youngkyu Lee¹, Jongho Park¹ and Chang-Ock Lee¹

1) *Department of Mathematical Sciences, KAIST, Daejeon 34141, KOREA*

Corresponding Author : Chang-Ock Lee, colee@kaist.edu

ABSTRACT

As deep neural networks (DNNs) [2] become deeper, the training time increases. In this perspective, multi-GPU parallel computing has become a key tool in accelerating the training of DNNs. In this paper, we introduce a novel methodology to construct a parallel neural network that can utilize multiple GPUs simultaneously from a given DNN. We observe that layers of the DNN can be interpreted as time steps of a time-dependent problem and can be parallelized by emulating a parallel-in-time algorithm called parareal [3]. The parareal algorithm consists of fine structures which can be implemented in parallel and a coarse structure which gives suitable approximations to the fine structures. By emulating it, the layers of the DNN are torn to form a parallel structure, which is connected using a suitable coarse network. We report accelerated and accuracy-preserved results of the proposed methodology applied to ResNet-1001 [1] on various datasets.

REFERENCES

1. He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, "Identity mappings in deep residual networks", *European Conference on Computer Vision*, pp 630–645, Springer, Cham, 2016.
2. LeCun, Yann and Bengio, Yoshua and Hinton, Geoffrey, "Deep learning", *Nature*, Vol. 521, 2015, pp. 436-444.
3. Lions, Jacques-Louis and Maday, Yvon and Turinici, Gabriel, "Résolution d'EDP par un schéma en temps «pararéel»", *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, Vol. 332, 2001, pp. 661-668.