Reading list

Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y. and He, K., 2017. Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv preprint arXiv:1706.02677.

Vancouver

https://arxiv.org/pdf/1706.02677.pdf

Hoffer, E., Hubara, I. and Soudry, D., 2017. Train longer, generalize better: closing the generalization gap in large batch training of neural networks. Advances in neural information processing systems, 30.

https://arxiv.org/pdf/1705.08741.pdf

You, Y., Zhang, Z., Hsieh, C.J., Demmel, J. and Keutzer, K., 2018, August. Imagenet training in minutes. In Proceedings of the 47th International Conference on Parallel Processing (pp. 1-10).

https://www2.eecs.berkeley.edu/Pubs/TechRpts/2020/EECS-2020-18.pdf

Keskar, N.S., Mudigere, D., Nocedal, J., Smelyanskiy, M. and Tang, P.T.P., 2016. On large-batch training for deep learning: Generalization gap and sharp minima. ICLR 2017

https://arxiv.org/pdf/1609.04836.pdf,

Yin, D., Pananjady, A., Lam, M., Papailiopoulos, D., Ramchandran, K. and Bartlett, P., 2018, March. Gradient diversity: a key ingredient for scalable distributed learning. In International Conference on Artificial Intelligence and Statistics (pp. 1998-2007). PMLR.

Vancouver

http://proceedings.mlr.press/v84/yin18a/yin18a.pdf