REFERENCES

[1] Fahim Chowdhury, Yue Zhu, Todd Heer, Saul Paredes, Adam Moody, Robin Goldstone, Kathryn Mohror, and Weikuan Yu. 2019. I/O characterization and performance evaluation of BeeGFS for deep learning. In *8th International Conference on Parallel Processing*. Kyoto, Japan, 80:1–80:10.

[2] Ching-Hsiang Chu, Pouya Kousha, Ammar Ahmad Awan, Kawthar Shafie Khorassani, Hari Subramoni, and Dhabaleswar K. Panda. 2020. NV-group: link-efficient reduction for distributed deep learning on modern dense GPU systems. In *34th ACM International Conference on Supercomputing*. Virtual, 1–12.

[3] Jeffrey Dean et al. 2012. Large scale distributed deep networks. In *25th International Conference on Neural Information Processing Systems*. Lake Tahoe, USA, 1223–1231.

[4] Nikoli Dryden, Roman Böhringer, Tal Ben-Nun, and Torsten Hoefler. 2021. Clairvoyant prefetching for distributed machine learning i/o. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, 1–15.

[5] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning (Adaptive Computation and Machine Learning series)*. MIT Press.

[6] K. He, X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 770–778.

[7] Andrei Ivanov, Nikoli Dryden, Tal Ben-Nun, Shigang Li, and Torsten Hoefler. 2021. Data movement is all you need: a case study on optimizing transformers. In *4th Conference on Machine Learning and Systems*. Virtual.

[8] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. 2015. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

[9] X. Liu, C. Wu, M. Menta, L. Herranz, B. Raducanu, A. D. Bagdanov, S. Jui, and J. van de Weijer. 2020. Generative feature replay for class-incremental learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 915–924.

[10] David Muñoz, Camilo Narváez, Carlos Cobos, Martha Mendoza, and Francisco Herrera. 2020. Incremental learning model inspired in rehearsal for deep convolutional networks. *Knowledge-Based Systems*, 208, 106460.

[11] Derek G. Murray, Jiri Simsa, Ana Klimovic, and Ihor Indyk. 2021. Tf.data: a machine learning data processing framework. (2021). arXiv: 2101.12127 [cs.LG].

[12] Deepak Narayanan, Aaron Harlap, Amar Phanishayee, Vivek Seshadri, Nikhil R. Devanur, Gregory R. Ganger, Phillip B. Gibbons, and Matei Zaharia. 2019. PipeDream: generalized pipeline parallelism for DNN training. In *27th ACM Symposium on Operating Systems Principles*. Huntsville, Canada, 1–15.

[13] Bogdan Nicolae. 2020. DataStates: Towards lightweight data models for deep learning. In *SMC'20: The 2020 Smoky Mountains Computational Sciences and Engineering Conference*. Nashville, United States, 117–129. DOI: 10.1007/978-3-030-63393-6_8.

[14] Bogdan Nicolae, Adam Moody, Elsa Gonsiorowski, Kathryn Mohror, and Franck Cappello. 2019. VeloC: towards high performance adaptive asynchronous checkpointing at large scale. In *IEEE International Parallel and Distributed Processing Symposium*. Rio de Janeiro, Brazil, 911–920.

[15] [n. d.] NVIDIA Data Loading Library. https://developer.nvidia.com/DALI. ().

[16] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. 2019. Continual lifelong learning with neural networks: a review. *Neural Networks*, 113, 54–71.

[17] Sarunya Pumma, Min Si, Wu-chun Feng, and Pavan Balaji. 2019. Scalable deep learning via I/O analysis and optimization. *ACM Transactions on Parallel Computing*, 6, 2, 1–34.

[18] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. 2017. ICaRL: incremental classifier and representation learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, 5533–5542.

[19] Daniel A Reed and Jack Dongarra. 2015. Exascale computing and big data. *Communications of the ACM*, 58, 7, 56–68.

[20] A. Robins. 1993. Catastrophic forgetting in neural networks: the role of rehearsal mechanisms. In *The First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*, 65–68.

[21] Andrei A. Rusu, Neil C. Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell. 2016. Progressive neural networks. arXiv 1606.04671. (2016).

[22] Alex Sergeev and Mike Del Balso. [n. d.] Meet Horovod: uber's open source distributed deep learning framework for tensorflow. https://eng.uber.com/horovod. ().

[23] Shane Snyder, Philip Carns, Kevin Harms, Robert Ross, Glenn K Lockwood, and Nicholas J Wright. 2016. Modular HPC I/O characterization with Darshan. In *5th Workshop on Extreme-scale Programming Tools*. IEEE, 9–17.

[24] Chih-Chieh Yang and Guojing Cong. 2019. Accelerating data loading in deep neural network training. In *IEEE 26th International Conference on High Performance Computing, Data, and Analytics*, 235–245.

[25] Yang You, Zhao Zhang, Cho-Jui Hsieh, and James Demmel Kurt Keutzer. 2018. ImageNet training in minutes. In *47th International Conference on Parallel Processing*. Eugene, USA, 1–10.

[26] Z. Zhang, L. Huang, J. Pauloski, and I. T. Foster. 2020. Efficient I/O for neural network training with compressed data. In *IEEE International Parallel and Distributed Processing Symposium*, 409–418.