

REFERENCES

- Bengio, Y., Léonard, N., & Courville, A. (2013). *Estimating or Propagating Gradients Through Stochastic Neurons for Conditional Computation*. 1–12. <http://arxiv.org/abs/1308.3432>
- Cun, Y. Le, Denker, J. S., & Solla, S. A. (1990). Optimal Brain Damage. *Advances in Neural Information Processing Systems 2*, 2(June 2014), 598–605. <http://dl.acm.org/citation.cfm?id=109230.109298>
- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503, 92–108. <https://doi.org/10.1016/j.neucom.2022.06.111>
- Fang, G., Ma, X., Song, M., Bi Mi, M., & Wang, X. (2023). DepGraph: Towards Any Structural Pruning. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2023-June*, 16091–16101. <https://doi.org/10.1109/CVPR52729.2023.01544>
- Gao, Z., Yao, Y., Wei, X., Yan, T., Zeng, S., Ge, G., Wang, Y., Ullah, A., & Reviriego, P. (2022). Reliability evaluation of FPGA based pruned neural networks. *Microelectronics Reliability*, 130(December 2021), 114498. <https://doi.org/10.1016/j.microrel.2022.114498>
- Han, S., Liu, X., Mao, H., Pu, J., Pedram, A., Horowitz, M. A., & Dally, W. J. (2016). EIE: Efficient Inference Engine on Compressed Deep Neural Network. *Proceedings - 2016 43rd International Symposium on Computer Architecture, ISCA 2016*, 16, 243–254. <https://doi.org/10.1109/ISCA.2016.30>
- Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*, 1–14.
- Han, S., Pool, J., Tran, J., & Dally, W. J. (2015). Learning both weights and connections for efficient neural networks. *Advances in Neural Information Processing Systems, 2015-Janua*, 1135–1143.

- He, K. (n.d.). *Deep Residual Learning for Image Recognition*.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE International Conference on Computer Vision, 2015 Inter*, 1026–1034. <https://doi.org/10.1109/ICCV.2015.123>
- He, Y., Balaprakash, P., & Li, Y. (2020). Fidelity: Efficient resilience analysis framework for deep learning accelerators. *Proceedings of the Annual International Symposium on Microarchitecture, MICRO, 2020-Octob*, 270–281. <https://doi.org/10.1109/MICRO50266.2020.00033>
- He, Y., Kang, G., Dong, X., Fu, Y., & Yang, Y. (2018). Soft filter pruning for accelerating deep convolutional neural networks. *IJCAI International Joint Conference on Artificial Intelligence, 2018-July*, 2234–2240. <https://doi.org/10.24963/ijcai.2018/309>
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *32nd International Conference on Machine Learning, ICML 2015, 1*, 448–456.
- Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H., & Kalenichenko, D. (2017). *Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference*.
- Kwasniewska, A., Szankin, M., Ozga, M., Wolfe, J., Das, A., Zajac, A., Ruminski, J., & Rad, P. (2019). Deep Learning Optimization for Edge Devices: Analysis of Training Quantization Parameters. *IECON Proceedings (Industrial Electronics Conference), 2019-Octob*, 96–101. <https://doi.org/10.1109/IECON.2019.8927153>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2323. <https://doi.org/10.1109/5.726791>
- Leveugle, R., Calvez, A., Maistri, P., & Vanhauwaert, P. (2009). Statistical fault injection: Quantified error and confidence. *Proceedings -Design, Automation and Test in Europe, DATE, May*, 502–506.

- <https://doi.org/10.1109/date.2009.5090716>
- Li, G., Hari, S. K. S., Sullivan, M., Tsai, T., Pattabiraman, K., Emer, J., & Keckler, S. W. (2017). Understanding Error Propagation in Deep Learning Neural Network (DNN) Accelerators and Applications. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC, 2017-Novem*. <https://doi.org/10.1145/3126908.3126964>
- May, T. C., & Woods, M. H. (1979). Alpha-Particle-Induced Soft Errors in Dynamic Memories. *IEEE Transactions on Electron Devices*, 26(1), 2–9. <https://doi.org/10.1109/T-ED.1979.19370>
- Min, J. T. C., & Motani, M. (2020). DropNet: Reducing neural network complexity via iterative pruning. *37th International Conference on Machine Learning, ICML 2020, Part F16814*, 9298–9308. <https://doi.org/10.5555/3524938.3525805>
- Nvidia. (2018). *NVDLA Open Source Project*. <http://nvdla.org/>
- Predić, B., Vukić, U., Saračević, M., Karabašević, D., & Stanujkić, D. (2022). The Possibility of Combining and Implementing Deep Neural Network Compression Methods. *Axioms*, 11(5). <https://doi.org/10.3390/axioms11050229>
- Qian, C., Zhang, M., Nie, Y., Lu, S., & Cao, H. (2023). A Survey of Bit-Flip Attacks on Deep Neural Network and Corresponding Defense Methods. *Electronics (Switzerland)*, 12(4). <https://doi.org/10.3390/electronics12040853>
- Rifqie, D. M., Surianto, D. F., Abdal, N. M., Hidayat M, W., & Ramli, H. (2022). Post Training Quantization in Lenet-5 Algorithm for Efficient Inference. *Journal of Embedded Systems, Security and Intelligent Systems*, 3(1), 60. <https://doi.org/10.26858/jessi.v3i1.34106>
- Ruospo, A., Gavarini, G., Bragaglia, I., Traiola, M., Bosio, A., Sanchez, E., Ruospo, A., Gavarini, G., Bragaglia, I., Traiola, M., Bosio, A., Ruospo, A., Gavarini, G., Bragaglia, I., Traiola, M., Bosio, A., & Sanchez, E. (2023). *Networks To cite this version : HAL Id : hal-03888005 Selective Hardening of Critical Neurons in Deep Neural Networks*.

- Ruospo, A., & Sanchez, E. (2021). On the reliability assessment of artificial neural networks running on ai-oriented mpsoes. *Applied Sciences (Switzerland)*, 11(14). <https://doi.org/10.3390/app11146455>
- Saxena, A. (2022). An Introduction to Convolutional Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*, 10(12), 943–947. <https://doi.org/10.22214/ijraset.2022.47789>
- Sidiropoulos, N. D., De Lathauwer, L., Fu, X., Huang, K., Papalexakis, E. E., & Faloutsos, C. (2017). Tensor Decomposition for Signal Processing and Machine Learning. *IEEE Transactions on Signal Processing*, 65(13), 3551–3582. <https://doi.org/10.1109/TSP.2017.2690524>
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171–177. <https://doi.org/10.1016/j.nlm.2004.06.005>
- Wang, N., Chen, C., & Gopalakrishnan, K. (n.d.). *Ultra-Low Precision 4-bit Training of Deep Neural Networks*. *NeurIPS 2020*.
- Yadav, P. O., Sharma, A., Philip, D., & Alexander, B. (2021). A Self Driving Car using Machine Learning and IOT. *International Journal Of Trendy Research In Engineering And Technology*, 05(04), 07–11. <https://doi.org/10.54473/ijtret.2021.5401>
- Zhao, K., Jain, A., & Zhao, M. (2020). *Iterative Activation-based Structured Pruning*.