













- [3] L. Jiao and J. Zhao, "A Survey on the New Generation of Deep Learning in Image Processing," *IEEE Access*, vol. 7, pp. 172231-172263, 2019. 10.1109/ACCESS.2019.2956508.
- [4] N. Justesen, P. Bontrager, J. Togelius and S. Risi, "Deep Learning for Video Game Playing," *IEEE Transactions on Games*, vol. 12, no. 1, pp. 1-20, March 2020. 10.1109/TG.2019.2896986.
- [5] T. Liang, J. Glossner, L. Wang, S. Shi, and X. Zhang, "Pruning and Quantization for Deep Neural Network Acceleration: A Survey," *Neurocomputing*, vol. 461, pp. 370-403, Oct. 2021, 10.1016/j.neucom.2021.07.045.
- [6] V. Lebedev, V. Lempitsky, "Speeding-up convolutional neural networks: A survey," *Bulletin of the Polish Academy of Sciences. Technical Science*, vol. 66, no. 6, pp. 799-811, 2018, 10.24425/bpas.2018.125927.
- [7] M. D. Zeiler, and R. Fergus. "Stochastic pooling for regularization of deep convolutional neural networks," arXiv preprint arXiv:1301.3557, 2013, 10.48550/arXiv.1301.3557.
- [8] J. Y. Wu, C. Yu, S. W. Fu, C. T. Liu, S. Y. Chien and Y. Tsao, "Increasing Compactness of Deep Learning Based Speech Enhancement Models With Parameter Pruning and Quantization Techniques," *IEEE Signal Processing Letters*, vol. 26, no. 12, pp. 1887-1891, Dec. 2019. 10.1109/LSP.2019.2951950.
- [9] J. Guo, W. Liu, W. Wang, J. Han, R. Li, Y. Lu, S. Hu, "Accelerating Distributed Deep Learning By Adaptive Gradient Quantization," in *Proc. ICASSP*, Barcelona, Spain, 2020, pp. 1603-1607, doi: 10.1109/ICASSP40776.2020.9054164.
- [10] C. Wang, L. Gong, X. Li, and X. Zhou, "A Ubiquitous Machine Learning Accelerator With Automatic Parallelization on FPGA," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 10, pp. 2346-2359, 1 Oct. 2020. 10.1109/TPDS.2020.2990924.
- [11] Y. Toyama, K. Yoshioka, K. Ban, S. Maya, A. Sai and K. Onizuka, "An 8 Bit 12.4 TOPS/W Phase-Domain MAC Circuit for Energy-Constrained Deep Learning Accelerators," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2730-2742, Oct. 2019. 10.1109/JSSC.2019.2926649.
- [12] Y. Wang, Y. Wang, H. Li and X. Li, "An Efficient Deep Learning Accelerator Architecture for Compressed Video Analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1-1, 2021. 10.1109/TCAD.2021.3120076.
- [13] J. Chen and X. Ran, "Deep Learning With Edge Computing: A Review," *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1655-1674, Aug. 2019. 10.1109/JPROC.2019.2921977.
- [14] K. Cao, Y. Liu, G. Meng and Q. Sun, "An Overview on Edge Computing Research," *IEEE Access*, vol. 8, pp. 85714-85728, 2020. 10.1109/ACCESS.2020.2991734.
- [15] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Generation Computer Systems*, vol. 97, pp. 219-235, 2019. 10.1016/j.future.2019.02.050.
- [16] S. W. Yang, "Efficient Deep Learning on Limited System Resources in FPGAs Performance Comparison on Floating Points," M.S. thesis, Dept. Computer & Information Technology, Korea Univ., Seoul, Korea, 2019.
- [17] A. Iwata, Y. Yoshida, S. Matsuda, Y. Sato, and N. Suzumura, "An artificial neural network accelerator using general purpose 24 bits floating point digital signal processors," in *IJCNN*, Washington, DC, USA, vol. 2, 1989, pp.171-175, doi: 10.1109/IJCNN.1989.118695.
- [18] J. Civit-Masot, F. Luna-Perejón, S. Vicente-Díaz, J. M. Rodríguez Corral and A. Civit, "TPU Cloud-Based Generalized U-Net for Eye Fundus Image Segmentation," *IEEE Access*, vol. 7, pp. 142379-142387, 2019. 10.1109/ACCESS.2019.2944692.
- [19] R. Murillo, A. A. D. Barrio, and G. Botella, "Deep PeNSieve: A deep learning framework based on the posit number system," *Digital Signal Processing*, vol. 102, 102762, 2020. 10.1016/j.dsp.2020.102762.
- [20] Q. H. Vo, N. Linh Le, F. Asim, L. W. Kim and C. S. Hong, "A Deep Learning Accelerator Based on a Streaming Architecture for Binary Neural Networks," *IEEE Access*, vol. 10, pp. 21141-21159, 2022. 10.1109/ACCESS.2022.3151916.
- [21] Y. Toyama, K. Yoshioka, K. Ban, S. Maya, A. Sai and K. Onizuka, "An 8 Bit 12.4 TOPS/W Phase-Domain MAC Circuit for Energy-Constrained Deep Learning Accelerators," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 10, pp. 2730-2742, Oct. 2019. 10.1109/JSSC.2019.2926649.
- [22] H. F. Langroudi, Z. Carmichael, and D. Kudithipudi, "Deep Learning Training on the Edge with Low-Precision Posits," arXiv preprint arXiv.1907.13216, 2019. 10.48550/arXiv.1907.13216.
- [23] H. W. Son, D. Y. Lee, and H. W. Kim, "Compact CNN Accelerator Chip Design with Optimized MAC And Pooling Layers," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 25, no. 9, pp. 1158-1165, Sept. 2021, 10.6109/JKIICE.2021.25.9.1158.
- [24] S. F. Hsiao, K. C. Chen, C. C. Lin, H. J. Chang, and B. C. Tsai, "Design of a Sparsity-Aware Reconfigurable Deep Learning Accelerator Supporting Various Types of Operations," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no. 3, pp. 376-387, Sept. 2020. 10.1109/JETCAS.2020.3015238.
- [25] L. Kang, H. Li, X. Li, and H. Zheng, "Design of Convolution Operation Accelerator based on FPGA," in *Proc. Int. Conf. MLBDBI*, Taiyuan, China, 2020, pp. 80-84, doi:10.1109/MLBDBI51377.2020.00021.
- [26] Y. LeCun, C. Cortes, and C. J. C. Burges, "The MNIST Database of handwritten digits," [Online]. Available: <http://yann.lecun.com/exdb/mnist/>.
- [27] L. Wan, M. Zeiler, S. Zhang, Y. L. Cun, and R. Fergus, "Regularization of Neural Networks using DropConnect," in *Proc. the International Conference on Machine Learning*, PMLR, Atlanta, Georgia, USA, 2013, pp. 1058-1066.
- [28] Siham Tabik, Ricardo F. Alvear-Sandoval, María M. Ruiz, José-Luis Sancho-Gómez, Aníbal R. Figueiras-Vidal, Francisco Herrera, "MNIST-NET10: A heterogeneous deep networks fusion based on the degree of certainty to reach 0.1% error rate. Ensembles overview and proposal," *Information Fusion*, vol. 62, pp. 73-80, Oct. 2020. 10.1016/j.inffus.2020.04.002.
- [29] Matuzas77, "MNIST-0.17," [Online]. Available: <https://github.com/Matuzas77/MNIST-0.17>.
- [30] S. S. Kadam, A. C. Adamthe, and A. B. Patil, "CNN Model for Image Classification on MNIST and Fashion-MNIST Dataset," *Journal of Scientific Research*, vol. 64, no. 2, pp. 374-384, 2020. 10.37398/JSR.2020.640251.
- [31] R. F. Alvear-Sandoval, J. L. Sancho-Gómez, and A. R. Figueiras-Vidal, "On improving CNNs performance: The case of MNIST," *Information Fusion*, vol. 52, pp. 106-109, Dec. 2019. 10.1016/j.inffus.2018.12.005.
- [32] A. Baldominos, Y. Saez, and P. Isasi, "A Survey of Handwritten Character Recognition with MNIST and EMNIST," *Appl. Sci.*, vol. 9, no. 15, 3169, Aug. 2019, 10.3390/app9153169.
- [33] A. Velichko A, "Neural Network for Low-Memory IoT Devices and MNIST Image Recognition Using Kernels Based on Logistic Map," *Electronics*, vol. 9, no. 9, 1432, 2020. 10.3390/electronics9091432.
- [34] S. S. Mor, S. Solanki, S. Gupta, S. Dhingra, M. Jain, and R. Saxena, "Handwritten Text Recognition: with Deep Learning and Android," *IJEAT*, vol. 8, no. 3S, pp. 819-825, Feb. 2019.
- [35] S. Himanshu, "Activation Functions : Sigmoid, tanh, ReLU, Leaky ReLU, PReLU, ELU, Threshold ReLU and Softmax basics for Neural Networks and Deep Learning," Jan. 19, 2019. [Online]. Available: <https://himanshud.medium.com/activation-functions-sigmoid-reLU-leaky-reLU-and-softmax-basics-for-neural-networks-and-deep-8d9c70eed91e>.