The Need for Energy-Efficient Machine Learning Algorithms

Student ID

1 Introduction

2 Literature Review

2.1 The Issue of Energy Consumption

Samuel (1959) (Fan et al. 2020) Jumper et al. (2021) (Schwartz et al. 2020) Devlin et al. (2019) (Bender et al. 2021) Fedus et al. (2021) (Amodei & Hernandez 2021) Silver et al. (2016) (Silver et al. 2017) Mahajan et al. (2018) (Lacoste et al. 2019) Strubell et al. (2019) (Cook et al. 2017) Google Sustainability (2021) (Stagg & Ward 2021) Microsoft News Center (2017) (Bizo et al. 2021) (Jain et al. 2019) Al-Jarrah et al. (2015) (United Nations Department of Economic and Social Affairs 2016) Buchanan & Tachet des Combes (2021) (Dodge et al. 2019) Walsh et al. (2021) (So et al. 2019) Daniel et al. (2021) (Acharyya et al. 2019) Marwah et al. (2011) (Xu et al. 2021) Veniat & Denoyer (2017) (He et al. 2015) Xie et al. (2016) (Yoo et al. 2011) Cheng et al. (2010) (Klein 2008) Bietti & Vatanparast (2019) (Lottick et al. 2019) Henderson et al. (2020) (Jouppi et al. 2017) Burr et al. (2021) (Kumar et al. 2020) Seznec et al. (2021) (de Leeuw 2021) Barroso & Hölzle (2007) (Wang & Selvan 2021) Joshi et al. (2020) (Liu et al. 2021) Merolla et al. (2014) (Davies et al. 2018) Fick et al. (2017) (Tang et al. 2018) Kendall et al. (2020) (Agrawal et al. 2021) Bose et al. (2020) (Gokmen & Haensch 2019) Abadi et al. (2016) (Paszke et al. 2019) Chirodea et al. (2021) (David et al. 2020) Chen et al. (2015) (Han et al. 2015) Hanson & Pratt (1988) (Iandola et al. 2016) Mehta et al. (2018) (Redmon & Farhadi 2016) Luo et al. (2019) (Dehghani et al. 2018) Plummer et al. (2020) (Teerapittayanon et al. 2017) Wang et al. (2017) (Sennrich et al. 2015) Morin & Bengio (2005) (Akhand et al. 2008) Rusu et al. (2016) (Yang et al. 2020) Belilovsky et al. (2018) (Simonyan & Zisserman 2014) Wang et al. (2020) (Peters et al. 2018) Cer et al. (2018) (Ruder 2018) Conneau et al. (2017)

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y. & Zheng, X. (2016), TensorFlow: A system for large-scale machine learning, *in* '12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)', pp. 265–283.
URL: *https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf*

Acharyya, P., Rosario, S. D., Flor, R., Joshi, R., Li, D., Linares, R. & Zhang, H. (2019), Autopilot of cement plants for reduction of fuel consumption and emissions, *in* 'ICML 2019 Workshop on Climate Change: How Can AI Help?'.
URL: *https://www.climatechange.ai/papers/icml2019/26*

Agrawal, A., Lee, S. K., Silberman, J., Ziegler, M., Kang, M., Venkataramani, S., Cao, N., Fleischer, B., Guillorn, M., Cohen, M., Mueller, S., Oh, J., Lutz, M., Jung, J., Koswatta, S., Zhou, C., Zalani, V., Bonanno, J., Casatuta, R., Chen, C.-Y., Choi, J., Haynie, H., Herbert, A., Jain, R., Kar, M., Kim, K.-H., Li, Y., Ren, Z., Rider, S., Schaal, M., Schelm, K., Scheuermann, M., Sun, X., Tran, H., Wang, N., Wang, W., Zhang, X., Shah, V., Curran, B., Srinivasan, V., Lu, P.-F., Shukla, S., Chang, L. & Gopalakrishnan, K. (2021), 9.1 A 7nm 4-Core AI Chip with

25.6TFLOPS Hybrid FP8 Training, 102.4TOPS INT4 Inference and Workload-Aware Throttling, *in* '2021 IEEE International Solid- State Circuits Conference (ISSCC)', Vol. 64, pp. 144–146.

Akhand, M. A. H., Islam, M. M. & Murase, K. (2008), Training of neural network ensemble through progressive interaction, *in* '2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)', pp. 2120–2126.

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K. & Taha, K. (2015), 'Efficient machine learning for big data: A review', *Big Data Research* **2**(3), 87–93. Big Data, Analytics, and High-Performance Computing.
URL: *https://www.sciencedirect.com/science/article/pii/S2214579615000271*

Amodei, D. & Hernandez, D. (2021), 'Ai and compute'.
URL: *https://openai.com/blog/ai-and-compute/*

Barroso, L. A. & Hölzle, U. (2007), 'The Case for Energy-Proportional Computing', *IEEE Computer* **40**(12), 33–37.

Belilovsky, E., Eickenberg, M. & Oyallon, E. (2018), 'Greedy Layerwise Learning Can Scale to ImageNet', *arXiv e-prints* p. arXiv:1812.11446.

Bender, E. M., Gebru, T., McMillan-Major, A. & Shmitchell, S. (2021), On the dangers of stochastic parrots: Can language models be too big?, *in* 'Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency', FAccT '21, Association for Computing Machinery, New York, NY, USA, p. 610–623.
URL: *https://doi.org/10.1145/3442188.3445922*

Bietti, E. & Vatanparast, R. (2019), 'Data Waste', *Harvard International Law Journal* **61**.
URL: *https://ssrn.com/abstract=3584251*

Bizo, D., Ascierto, R., Lawrence, A. & Davis, J. (2021), Uptime Institute Global Data Center Survey 2021, Technical report, Uptime Institute Intelligence.
URL: *https://uptimeinstitute.com/2021-data-center-industry-survey-results*

Bose, S. K., Acharya, J. & Basu, A. (2020), 'Is my Neural Network Neuromorphic? Taxonomy, Recent Trends and Future Directions in Neuromorphic Engineering', *arXiv e-prints* p. arXiv:2002.11945.

Buchanan, W. & Tachet des Combes, R. (2021), 'Charting the path towards sustainable AI with Azure Machine Learning Resource Metrics', *Microsoft Azure Machine Learning* .
URL: *https://techcommunity.microsoft.com/t5/green-tech-blog/charting-the-path-towards-sustainable-ai-with-azure-machine/ba-p/2866923*

Burr, G. W., Sebastian, A., Ando, T. & Haensch, W. (2021), 'Ohm's Law + Kirchhoff's Current Law = Better AI: Neural-Network Processing Done in Memory with Analog Circuits will Save Energy', *IEEE Spectrum* **58**(12), 44–49.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Sung, Y.-H., Strope, B. & Kurzweil, R. (2018), 'Universal Sentence Encoder', *arXiv e-prints* p. arXiv:1803.11175.

Chen, T., Li, M., Li, Y., Lin, M., Wang, N., Wang, M., Xiao, T., Xu, B., Zhang, C. & Zhang, Z. (2015), 'MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems', *arXiv e-prints* p. arXiv:1512.01274.

Cheng, H., Tan, P.-N. & Jin, R. (2010), 'Efficient algorithm for localized support vector machine', *IEEE Transactions on Knowledge and Data Engineering* **22**(4), 537–549.

Chirodea, M. C., Novac, O. C., Novac, C. M., Bizon, N., Oproescu, M. & Gordan, C. E. (2021), Comparison of tensorflow and pytorch in convolutional neural network - based applications, *in* '2021 13th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)', pp. 1–6.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L. & Bordes, A. (2017), 'Supervised Learning of Universal Sentence Representations from Natural Language Inference Data', *arXiv e-prints* p. arXiv:1705.02364.

Cook, G., Lee, J., Tsai, T., Kongn, A., Deans, J., Johnson, B., Jardim, E. & Johnson, B. (2017), Clicking Clean: Who is winning the race to build a green internet?, Technical report, Greenpeace.

Daniel, C., Shukla, A. K. & Sharma, M. (2021), Applications of machine learning in harnessing of renewable energy, *in* P. V. Baredar, S. Tangellapalli & C. S. Solanki, eds, 'Advances in Clean Energy Technologies', Springer Singapore, Singapore, pp. 177–187.

David, R., Duke, J., Jain, A., Janapa Reddi, V., Jeffries, N., Li, J., Kreeger, N., Nappier, I., Natraj, M., Regev, S., Rhodes, R., Wang, T. & Warden, P. (2020), 'TensorFlow Lite Micro: Embedded Machine Learning on TinyML Systems', *arXiv e-prints* p. arXiv:2010.08678.

Davies, M., Srinivasa, N., Lin, T.-H., Chinya, G., Cao, Y., Choday, S. H., Dimou, G., Joshi, P., Imam, N., Jain, S., Liao, Y., Lin, C.-K., Lines, A., Liu, R., Mathaikutty, D., McCoy, S., Paul, A., Tse, J., Venkataramanan, G., Weng, Y.-H., Wild, A., Yang, Y. & Wang, H. (2018), 'Loihi: A neuromorphic manycore processor with on-chip learning', *IEEE Micro* **38**(1), 82–99.

de Leeuw, O. (2021), Towards More Energy-Efficient Neural Networks, *in* 'Open Data Science Conference West 2021'.
URL: *https://odsc.com/blog/towards-more-energy-efficient-machine-learning-models/*

Dehghani, M., Gouws, S., Vinyals, O., Uszkoreit, J. & Kaiser, Ł. (2018), 'Universal Transformers', *arXiv e-prints* p. arXiv:1807.03819.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *ArXiv* **abs/1810.04805**.

Dodge, J., Gururangan, S., Card, D., Schwartz, R. & Smith, N. A. (2019), 'Show Your Work: Improved Reporting of Experimental Results', *arXiv e-prints* p. arXiv:1909.03004.

Fan, A., Bhosale, S., Schwenk, H., Ma, Z., El-Kishky, A., Goyal, S., Baines, M., Celebi, O., Wenzek, G., Chaudhary, V., Goyal, N., Birch, T., Liptchinsky, V., Edunov, S., Grave, E., Auli, M. & Joulin, A. (2020), 'Beyond english-centric multilingual machine translation'.

Fedus, W., Zoph, B. & Shazeer, N. (2021), 'Switch Transformers: Scaling to Trillion Parameter Models with Simple and Efficient Sparsity', *arXiv e-prints* p. arXiv:2101.03961.

Fick, L., Blaauw, D., Sylvester, D., Skrzyniarz, S., Parikh, M. & Fick, D. (2017), Analog in-memory subthreshold deep neural network accelerator, *in* '2017 IEEE Custom Integrated Circuits Conference (CICC)', pp. 1–4.

Gokmen, T. & Haensch, W. (2019), 'Algorithm for Training Neural Networks on Resistive Device Arrays', *arXiv e-prints* p. arXiv:1909.07908.

Google Sustainability (2021), Google Environmental Report 2021, Technical report, Google LLC.
URL: *https://www.gstatic.com/gumdrop/sustainability/google-2021-environmental-report.pdf*

Han, S., Pool, J., Tran, J. & Dally, W. J. (2015), 'Learning both Weights and Connections for Efficient Neural Networks', *arXiv e-prints* p. arXiv:1506.02626.

Hanson, S. & Pratt, L. (1988), Comparing Biases for Minimal Network Construction with Back-Propagation, *in* D. Touretzky, ed., 'Advances in Neural Information Processing Systems', Vol. 1, Morgan-Kaufmann.
URL: *https://proceedings.neurips.cc/paper/1988/file/1c9ac0159c94d8d0cbedc973445af2da-Paper.pdf*

He, K., Zhang, X., Ren, S. & Sun, J. (2015), 'Deep Residual Learning for Image Recognition', *arXiv e-prints* p. arXiv:1512.03385.

Henderson, P., Hu, J., Romoff, J., Brunskill, E., Jurafsky, D. & Pineau, J. (2020), 'Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning', *arXiv e-prints* p. arXiv:2002.05651.

Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J. & Keutzer, K. (2016), 'SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and ¡0.5MB model size', *arXiv e-prints* p. arXiv:1602.07360.

Jain, P., Mo, X., Jain, A., Tumanov, A., Gonzalez, J. E. & Stoica, I. (2019), 'The OoO VLIW JIT Compiler for GPU Inference', *ArXiv* **abs/1901.10008**.

Joshi, V., Le Gallo, M., Haefeli, S., Boybat, I., Nandakumar, S. R., Piveteau, C., Dazzi, M., Rajendran, B., Sebastian, A. & Eleftheriou, E. (2020), 'Accurate deep neural network inference using computational phase-change memory', *Nature Communications* **11**, 2473.

Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., Bates, S., Bhatia, S., Boden, N., Borchers, A., Boyle, R., Cantin, P.-l., Chao, C., Clark, C., Coriell, J., Daley, M., Dau, M., Dean, J., Gelb, B., Vazir Ghaemmaghami, T., Gottipati, R., Gulland, W., Hagmann, R., Ho, C. R., Hogberg, D., Hu, J., Hundt, R., Hurt, D., Ibarz, J., Jaffey, A., Jaworski, A., Kaplan, A., Khaitan, H., Koch, A., Kumar, N., Lacy, S., Laudon, J., Law, J., Le, D., Leary, C., Liu, Z., Lucke, K., Lundin, A., MacKean, G., Maggiore, A., Mahony, M., Miller, K., Nagarajan, R., Narayanaswami, R., Ni, R., Nix, K., Norrie, T., Omernick, M., Penukonda, N., Phelps, A., Ross, J., Ross, M., Salek, A., Samadiani, E., Severn, C., Sizikov, G., Snelham, M., Souter, J., Steinberg, D., Swing, A., Tan, M., Thorson, G., Tian, B., Toma, H., Tuttle, E., Vasudevan, V., Walter, R., Wang, W., Wilcox, E. & Yoon, D. H. (2017), 'In-Datacenter Performance Analysis of a Tensor Processing Unit', *arXiv e-prints* p. arXiv:1704.04760.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P. & Hassabis, D. (2021), 'Highly accurate protein structure prediction with alphafold', *Nature* **596**(7873), 583–589.
URL: *https://doi.org/10.1038/s41586-021-03819-2*

Kendall, J., Pantone, R., Manickavasagam, K., Bengio, Y. & Scellier, B. (2020), 'Training End-to-End Analog Neural Networks with Equilibrium Propagation', *arXiv e-prints* p. arXiv:2006.01981.

Klein, G. (2008), 'Performing a project premortem', *Engineering Management Review, IEEE* **36**, 103–104.

Kumar, M., Zhang, X., Liu, L., Wang, Y. & Shi, W. (2020), Energy-Efficient Machine Learning on the Edges, *in* '2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)', pp. 912–921.

Lacoste, A., Luccioni, A., Schmidt, V. & Dandres, T. (2019), 'Quantifying the Carbon Emissions of Machine Learning', *arXiv e-prints* p. arXiv:1910.09700.

Liu, S.-C., Strachan, J. P. & Basu, A. (2021), 'Prospects for Analog Circuits in Deep Networks', *arXiv e-prints* p. arXiv:2106.12444.

Lottick, K., Susai, S., Friedler, S. A. & Wilson, J. P. (2019), 'Energy Usage Reports: Environmental awareness as part of algorithmic accountability', *arXiv e-prints* p. arXiv:1911.08354.

Luo, J.-H., Zhang, H., Zhou, H.-Y., Xie, C.-W., Wu, J. & Lin, W. (2019), 'ThiNet: Pruning CNN Filters for a Thinner Net', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **41**(10), 2525–2538.

Mahajan, D., Girshick, R., Ramanathan, V., He, K., Paluri, M., Li, Y., Bharambe, A. & van der Maaten, L. (2018), 'Exploring the Limits of Weakly Supervised Pretraining', *arXiv e-prints* p. arXiv:1805.00932.

Marwah, M., Shah, A., Bash, C., Patel, C. & Ramakrishnan, N. (2011), 'Using data mining to help design sustainable products', *IEEE Computer* **44**, 103–106.

Mehta, S., Rastegari, M., Shapiro, L. & Hajishirzi, H. (2018), 'ESPNetv2: A Light-weight, Power Efficient, and General Purpose Convolutional Neural Network', *arXiv e-prints* p. arXiv:1811.11431.

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., Jackson, B. L., Imam, N., Guo, C., Nakamura, Y., Brezzo, B., Vo, I., Esser, S. K., Appuswamy, R., Taba, B., Amir, A., Flickner, M. D., Risk, W. P., Manohar, R. & Modha, D. S. (2014), 'A million spiking-neuron integrated circuit with a scalable communication network and interface', *Science* **345**(6197), 668–673.
URL: *https://www.science.org/doi/abs/10.1126/science.1254642*

Microsoft News Center (2017), 'Microsoft announces one of the largest wind deals in the Netherlands with Vattenfall', *Microsoft Media Relations, WE Communications for Microsoft, (425) 638-7777* .
URL: *https://news.microsoft.com/2017/11/02/microsoft-announces-one-of-the-largest-wind-deals-in-the-netherlands-with-vattenfall/*

Morin, F. & Bengio, Y. (2005), Hierarchical Probabilistic Neural Network Language Model, *in* R. G. Cowell & Z. Ghahramani, eds, 'Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics', Vol. R5 of *Proceedings of Machine Learning Research*, PMLR, pp. 246–252. Reissued by PMLR on 30 March 2021.
URL: *https://proceedings.mlr.press/r5/morin05a.html*

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala, S. (2019), 'PyTorch: An Imperative Style, High-Performance Deep Learning Library', *arXiv e-prints* p. arXiv:1912.01703.

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. & Zettlemoyer, L. (2018), 'Deep contextualized word representations', *arXiv e-prints* p. arXiv:1802.05365.

Plummer, B. A., Dryden, N., Frost, J., Hoefler, T. & Saenko, K. (2020), 'Neural Parameter Allocation Search', *arXiv e-prints* p. arXiv:2006.10598.

Redmon, J. & Farhadi, A. (2016), 'YOLO9000: Better, Faster, Stronger', *arXiv e-prints* p. arXiv:1612.08242.

Ruder, S. (2018), 'NLP's ImageNet moment has arrived', *The Gradient* .

Rusu, A. A., Rabinowitz, N. C., Desjardins, G., Soyer, H., Kirkpatrick, J., Kavukcuoglu, K., Pascanu, R. & Hadsell, R. (2016), 'Progressive Neural Networks', *arXiv e-prints* p. arXiv:1606.04671.

Samuel, A. L. (1959), 'Some studies in machine learning using the game of checkers', *IBM J. Res. Dev.* **3**(3), 210–229.
URL: *https://doi.org/10.1147/rd.33.0210*

Schwartz, R., Dodge, J., Smith, N. & Etzioni, O. (2020), 'Green ai', *Communications of the ACM* **63**, 54 – 63.

Sennrich, R., Haddow, B. & Birch, A. (2015), 'Neural Machine Translation of Rare Words with Subword Units', *arXiv e-prints* p. arXiv:1508.07909.

Seznec, M., Gac, N., Orieux, F. & Sashala Naik, A. (2021), 'Real-Time Optical Flow Processing on Embedded GPU: anHardware-Aware Algorithm to Implementation Strategy', *Journal of Real-Time Image Processing* .
URL: *https://hal.archives-ouvertes.fr/hal-03457011*

Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. & Hassabis, D. (2016), 'Mastering the game of go with deep neural networks and tree search', *Nature* **529**, 484–489.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T. P., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. & Hassabis, D. (2017), 'Mastering the game of go without human knowledge', *Nature* **550**, 354–359.

Simonyan, K. & Zisserman, A. (2014), 'Very Deep Convolutional Networks for Large-Scale Image Recognition', *arXiv e-prints* p. arXiv:1409.1556.

So, D. R., Liang, C. & Le, Q. V. (2019), 'The Evolved Transformer', *arXiv e-prints* p. arXiv:1901.11117.

Stagg, R. & Ward, S. (2021), Forestry Statistics 2021: A compendium of statistics about woodland, forestry and primary wood processing in the United Kingdom, Technical report, The Research Agency of the Forestry Commission, 231 Corstorphine Road, Edinburgh, EH12 7AT.
URL: *https://www.forestresearch.gov.uk/tools-and-resources/statistics/forestry-statistics/forestry-statistics-2021/*

Strubell, E., Ganesh, A. & McCallum, A. (2019), 'Energy and Policy Considerations for Deep Learning in NLP', *arXiv e-prints* p. arXiv:1906.02243.

Tang, J., Bishop, D., Kim, S., Copel, M., Gokmen, T., Todorov, T., Shin, S., Lee, K.-T., Solomon, P., Chan, K., Haensch, W. & Rozen, J. (2018), ECRAM as Scalable Synaptic Cell for High-Speed, Low-Power Neuromorphic Computing, *in* '2018 IEEE International Electron Devices Meeting (IEDM)', pp. 13.1.1–13.1.4.

Teerapittayanon, S., McDanel, B. & Kung, H. T. (2017), 'BranchyNet: Fast Inference via Early Exiting from Deep Neural Networks', *arXiv e-prints* p. arXiv:1709.01686.

United Nations Department of Economic and Social Affairs (2016), *World Economic and Social Survey 2016: Climate Change Resilience - An Opportunity for Reducing Inequalities*, World Economic and Social Survey (WESS), UN.
URL: *https://books.google.co.uk/books?id=JlvjDwAAQBAJ*

Veniat, T. & Denoyer, L. (2017), 'Learning Time/Memory-Efficient Deep Architectures with Budgeted Super Networks', *arXiv e-prints* p. arXiv:1706.00046.

Walsh, P., Bera, J., Sharma, V. S., Kaulgud, V., Rao, R. M. & Ross, O. (2021), Sustainable ai in the cloud: Exploring machine learning energy use in the cloud, *in* '2021 36th IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW)', pp. 265–266.

Wang, K., Gao, X., Zhao, Y., Li, X., Dou, D. & Xu, C.-Z. (2020), Pay attention to features, transfer learn faster cnns, *in* 'International Conference on Learning Representations'.
URL: *https://openreview.net/forum?id=ryxyCeHtPB*

Wang, T. & Selvan, A. (2021), 'Google wins MLPerf benchmarks with TPU V4', *Google Cloud Blog* .
URL: *https://cloud.google.com/blog/products/ai-machine-learning/google-wins-mlperf-benchmarks-with-tpu-v4*

Wang, X., Yu, F., Dou, Z.-Y., Darrell, T. & Gonzalez, J. E. (2017), 'SkipNet: Learning Dynamic Routing in Convolutional Networks', *arXiv e-prints* p. arXiv:1711.09485.

Xie, S., Girshick, R., Dollár, P., Tu, Z. & He, K. (2016), 'Aggregated Residual Transformations for Deep Neural Networks', *arXiv e-prints* p. arXiv:1611.05431.

Xu, J., Zhou, W., Fu, Z., Zhou, H. & Li, L. (2021), 'A Survey on Green Deep Learning', *ArXiv* **abs/2111.05193**.

Yang, C., Wang, S., Yang, C., Li, Y., He, R. & Zhang, J. (2020), 'Progressively Stacking 2.0: A Multi-stage Layerwise Training Method for BERT Training Speedup', *arXiv e-prints* p. arXiv:2011.13635.

Yoo, P. D., Ng, J. W. & Zomaya, A. Y. (2011), An energy-efficient kernel framework for large-scale data modeling and classification, *in* '2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum', pp. 404–408.