Automated Deep Learning: Neural Architecture Search Is Not the End

Other titles in Foundations and Trends® in Machine Learning

Causal Fairness Analysis: A Causal Toolkit for Fair Machine Learning
Drago Plečko and Elias Bareinboim
ISBN: 978-1-63828-330-0

A Friendly Tutorial on Mean-Field Spin Glass Techniques for Non-Physicists
Andrea Montanari and Subhabrata Sen
ISBN: 978-1-63828-212-9

Reinforcement Learning, Bit by Bit Xiuyuan Lu, Benjamin Van Roy, Vikranth Dwaracherla, Morteza Ibrahimi, Ian Osband and Zheng Wen
ISBN: 978-1-63828-254-9

Introduction to Riemannian Geometry and Geometric Statistics: From Basic Theory to Implementation with Geomstats
Nicolas Guigui, Nina Miolane and Xavier Pennec
ISBN: 978-1-63828-154-2

Graph Neural Networks for Natural Language Processing: A Survey
Lingfei Wu, Yu Chen, Kai Shen, Xiaojie Guo, Hanning Gao, Shucheng Li, Jian Pei and Bo Long
ISBN: 978-1-63828-142-9

Model-based Reinforcement Learning: A Survey
Thomas M. Moerland, Joost Broekens, Aske Plaat and Catholijn M. Jonker
ISBN: 978-1-63828-056-9

Automated Deep Learning: Neural Architecture Search Is Not the End

Xuanyi Dong
University of Technology Sydney
Google Brain
xuanyi.dxy@gmail.com

David Jacob Kedziora
University of Technology Sydney
david.kedziora@uts.edu.au

Katarzyna Musial
University of Technology Sydney
katarzyna.musial-gabrys@uts.edu.au

Bogdan Gabrys
University of Technology Sydney
bogdan.gabrys@uts.edu.au

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

X. Dong *et al.*. *Automated Deep Learning: Neural Architecture Search Is Not the End.* Foundations and Trends® in Machine Learning, vol. 17, no. 5, pp. 767–920, 2024.

ISBN: 978-1-63828-319-5

Foundations and Trends® in Machine Learning
Volume 17, Issue 5, 2024

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles in the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2024, Volume 17, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Contents

Automated Deep Learning: Neural Architecture Search Is Not the End

Xuanyi Dong[1], David Jacob Kedziora[2], Katarzyna Musial[2] and Bogdan Gabrys[2]

[1] *Complex Adaptive Systems Lab, University of Technology Sydney, Australia and Brain Team, Google Research, USA; xuanyi.dxy@gmail.com*

[2] *Complex Adaptive Systems Lab, University of Technology Sydney, Australia; david.kedziora@uts.edu.au, katarzyna.musial-gabrys@uts.edu.au, bogdan.gabrys@uts.edu.au*

ABSTRACT

Deep learning (DL) has proven to be a highly effective approach for developing models in diverse contexts, including visual perception, speech recognition, and machine translation. However, the end-to-end process for applying DL is not trivial. It requires grappling with problem formulation and context understanding, data engineering, model development, deployment, continuous monitoring and maintenance, and so on. Moreover, each of these steps typically relies heavily on humans, in terms of both knowledge and interactions, which impedes the further advancement and democratization of DL. Consequently, in response to these issues, a new field has emerged over the last few years: automated deep learning (AutoDL). This endeavor seeks to minimize the need for human involvement and is best known for its achievements in neural architecture search (NAS), a topic that has been the focus of several surveys. That stated,

Xuanyi Dong, David Jacob Kedziora, Katarzyna Musial and Bogdan Gabrys (2024), "Automated Deep Learning: Neural Architecture Search Is Not the End", Foundations and Trends® in Machine Learning: Vol. 17, No. 5, pp 767–920. DOI: 10.1561/2200000119.

NAS is not the be-all and end-all of AutoDL. Accordingly, this review adopts an overarching perspective, examining research efforts into automation across the entirety of an archetypal DL workflow. In so doing, this work also proposes a comprehensive set of ten criteria by which to assess existing work in both individual publications and broader research areas. These criteria are: novelty, solution quality, efficiency, stability, interpretability, reproducibility, engineering quality, scalability, generalizability, and eco-friendliness. Thus, ultimately, this review provides an evaluative overview of AutoDL in the early 2020s, identifying where future opportunities for progress may exist.

Keywords: Automated deep learning (AutoDL); neural architecture search (NAS); hyperparameter optimization (HPO); automated data engineering; hardware search; automated deployment; life-long learning; persistent learning; adaptation; automated machine learning (AutoML); autonomous machine learning (AutonoML); deep neural networks; deep learning.

1

Introduction

In the quest for artificial intelligence (AI), history may judge the early 2010s as a mental reset, stimulating a new era of research and development with unrivaled intensity. Within those reformative years, the field of machine learning (ML) witnessed a shifting of priorities and approaches. Two threads of aspiration stand out:

- Deep Learning (DL) – The idea that multi-layered artificial-neuron networks are central to pushing the capabilities of ML.
- Automated Machine Learning (AutoML) – The idea that no part of an ML workflow should necessarily depend on human involvement.

It was inevitable that these two ideologies would eventually converge, fusing into the novel subject of automated deep learning (AutoDL).

Admittedly, while AutoDL is a "hot topic" in 2021, the foundations underlying this surge of activity stretch back for decades. The notion of ML itself (Michie *et al.*, 1994) was established in the 1950s, aiming to tune mathematical models of desirable functions via automated data-driven algorithms. In time, by the turn of the 21st century, numerous ML models and algorithms would be in practical use, with support vector machines and other kernel methods proving particularly popular (Hofmann *et al.*, 2008). However, the concept of a neuron, inextricably linked to human intelligence, always seemed an obvious basis for ML. Depicted computationally as early as the 1940s (McCulloch and Pitts, 1943), their representational power in multi-layered arrangements was evident by the late 1960s, exemplified by the proto-DL "group method of data handling" (GMDH) (Ivakhnenko, 1971). Since then, with stutters around AI winters, numerous types of neural layers and architectural variants have been proposed and adopted. These include recurrent structures (Hopfield, 1982; Little, 1974), convolutional and downsample layers (Fukushima and Miyake, 1982), auto-encoder hierarchies (Ballard, 1987), memory mechanisms (Pollack, 1989), and gating structures (Hochreiter and Schmidhuber, 1997).

As a result, the historical successes of artificial neural networks (ANNs) are undeniably many, encompassing handwriting recognition (LeCun *et al.*, 1989), time series prediction (Weigend and Gershenfeld, 1993), video retrieval (Ji *et al.*, 2012; Yang *et al.*, 2009), mitosis detection (Cireşan *et al.*, 2013), and so on. Yet the advantages of deep neural networks (DNNs), including their status as universal approximators (Hornik *et al.*, 1989), are countered by the unwieldy nature of their complexity. For instance, while backpropagation was established as reverse automatic differentiation in the 1970s (Linnainmaa, 1970), this DNN training technique did not become generally feasible until relatively recently. Thus, the rising dominance of DL in the 2010s (LeCun *et al.*, 2015; Schmidhuber, 2015) is as much an outcome of big data infrastructure and hardware acceleration, specifically graphical processing units (GPUs), as it is the result of any one theoretical advance.

In contrast, the evolution of AutoML is harder to pin down, primarily because the scope of automating higher-level ML mechanisms can be made extremely broad. The extended history of this topic is grappled with elsewhere (Kedziora *et al.*, 2020). Nonetheless, the mainstream interpretation of AutoML – and even the abbreviation itself – has been forged on the back of advances in ML model/algorithm selection and the optimization of their user-defined hyperparameters. Accordingly, if the success of a DNN in the 2012 ImageNet competition (Krizhevsky *et al.*, 2012) heralds the modern DL era, then the release of Auto-WEKA in

2013 (Thornton *et al.*, 2013) marks the start of the modern AutoML era. Within several years, by late 2016, these threads would start to entwine within the sub-field of neural architecture search (NAS) (Baker *et al.*, 2017; Zoph and Le, 2017). This was not the first time that AutoML techniques had been applied to neural networks, but it was the moment that the broader data science community took notice. It was also opportune; the website Papers-With-Code (Facebook AI Research, n.d.) highlights that, while the number of DL publications has skyrocketed since 2012, year-by-year performance improvements on many benchmark datasets have diminished, i.e., those related to vision, text, audio, and speech. There is a sense that, as state-of-the-art (SoTA) DL models have become highly sophisticated, a reliance on human design is locking out broader engagement behind steep learning curves, while also hindering further metric progress. Automation through NAS is a vital step in enabling a broader community to push these technical limits.

Importantly, while NAS launched the modern AutoDL story, it does not encompass it. Model selection, i.e., the design of a neural network, is but one stage of a DL workflow. As illustrated in Figure 1.1, there are many other subtasks involved in ML/DL, such as defining a problem of interest, collecting and organizing data, generating features, deploying and adapting trained models. This workflow may often be sequential in research and development, but real-world applications are much more agile and will typically reiterate through earlier operations and, in the case of large-scale systems, these processes may even be asynchronous. In effect, AI based on DL cannot reach its full potential

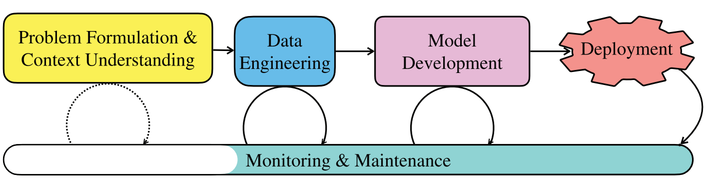

Figure 1.1: Schematic of an end-to-end DL workflow, i.e., the processes involved in applying DL to a problem. Traditionally, human decisions are required for every part of this workflow, such as analyzing a problem context, defining an ML task, designing a model, manually tuning hyperparameters, selecting training strategy.

without considering the entire life cycle of a solution, from its design to the maintenance phase.

We now bring attention to a previous review (Kedziora *et al.*, 2020), which surveyed efforts to automate all aspects of this workflow in the general context of ML, with additional focus on how the resulting mechanisms may be integrated into a single architecture. The review touched on NAS and other elements of DL, but it could not cover the full extent of work in the AutoDL sphere. It did not need to; at a high level, working with DNNs fits smoothly into the conceptual framework of both AutoML and its extension, autonomous machine learning (AutonoML). However, on a practical level, the complexity of DNNs throws up many challenges that have arguably constrained the breadth of developments in AutoDL as compared to standard AutoML. Instead, what is remarkable is the *depth* of research in AutoDL, with numerous innovations brought about by attempts to surmount these obstacles, all with the aim of making the automation of DL feasible. Certainly, it would be remiss to trivialize AutoDL as just a subset of AutoML. Likewise, critically evaluating the limitations of present-day AutoDL is just as worthwhile as highlighting its accomplishments. For instance, the field of DL is sometimes criticized for a tunnel-vision focus on model-performance metrics within a limited set of benchmarks, an attitude which, while valid, risks missing the broader perspective on all that AutoDL may become (Dong *et al.*, 2021a; Gencoglu *et al.*, 2019). In essence, there is a need to consider several questions more thoroughly:

- As we enter the 2020s, what is the current research landscape of DL?
- What makes a "good" DL model?
- How can automated systems best pursue and support this model performance?
- Is the field of AutoDL even advanced enough for such a meta-analysis?

This work is an extension of the broadly scoped AutonoML review (Kedziora *et al.*, 2020) with an in-depth focus on the newly

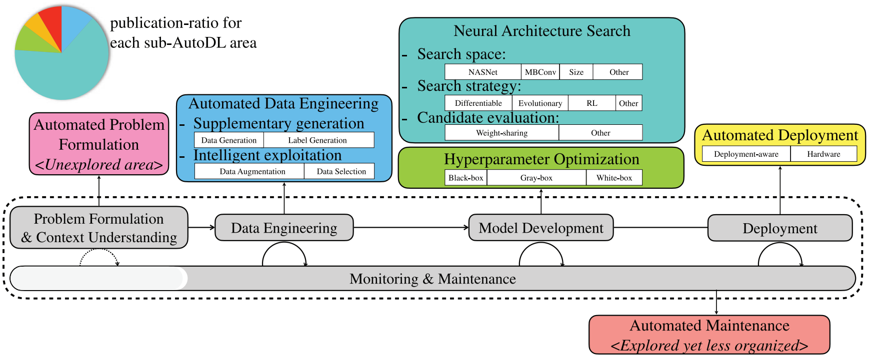

Figure 1.2: (Color online) breakdown schematic for research activity in AutoDL, with surveyed publications attributed to different phases of a DL workflow and then further subcategorized. The pie chart denotes the ratio of publications across all workflow phases, while the white stacked bars denote ratios within each subcategory. All statistics are derived from the Awesome-AutoDL project, at: https://github.com/D-X-Y/Awesome-AutoDL.

popularized topic of AutoDL. While there are many surveys in this sphere[1] (Elsken *et al.*, 2019; Feurer and Hutter, 2019; Ren *et al.*, 2021; Wistuba *et al.*, 2019; Yu and Zhu, 2020), most focus on deep analysis within one or two sub-domains of AutoDL. In contrast, we examine the entire spectrum of the DL workflow, as of 2021, actively identifying areas that have not yet been fully explored or where automation research may still be in its infancy. We first provide an overview of AutoDL in Section 2, introducing several fundamental concepts. Then, partitioning major AutoDL research into sections inspired by a DL workflow, as per Figure 1.2, we explore automation for: task management (Section 3), data preparation (Section 4), neural architecture design (Section 5), hyperparameter selection (Section 6), model deployment (Section 7), and online maintenance (Section 8).

Crucially, a major motivation for this review is a reaction to the sheer quantity of publications in the space of AutoDL; we aim to provide summary assessments of surveyed AutoDL algorithms/research in terms of ten carefully designed criteria, not just accuracy alone. These

[1]One of the authors maintains a publicly accessible curated list of AutoDL resources at: https://github.com/D-X-Y/Awesome-AutoDL.

are introduced in Section 2.4 and form an evaluative framework for overviews within every aforementioned section, as well as, in Section 9, a final critical discussion around the entire field of AutoDL.

References

Abdelfattah, M. S., A. Mehrotra, Ł. Dudziak, and N. D. Lane (2021). "Zero-cost proxies for lightweight NAS". In: *International Conference on Learning Representations (ICLR)*.

Addanki, R., S. B. Venkatakrishnan, S. Gupta, H. Mao, and M. Alizadeh (2019). "Placeto: Learning generalizable device placement algorithms for distributed machine learning". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 3981–3991.

Agresti, A. (2018). *An Introduction to Categorical Data Analysis*. John Wiley & Sons.

Akiba, T., S. Sano, T. Yanase, T. Ohta, and M. Koyama (2019). "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 2623–2631. DOI: 10.1145/3292500.3330701.

Ali, A. R., M. Budka, and B. Gabrys (2019). "A meta-reinforcement learning approach to optimize parameters and hyper-parameters simultaneously". In: *Pricai*. 93–106. DOI: 10.1007/978-3-030-29911-8_8.

Ali, A. R., B. Gabrys, and M. Budka (2018). "Cross-domain meta-learning for time-series forecasting". *Procedia Computer Science*. 126: 9–18. DOI: 10.1016/j.procs.2018.07.204.

Almgren, R. and N. Chriss (2001). “Optimal execution of portfolio transactions”. *Journal of Risk*. 3: 5–40. DOI: 10.21314/JOR.2001.041.

Andrychowicz, M., M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. De Freitas (2016). “Learning to learn by gradient descent by gradient descent”. In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. Vol. 29.

Ankit, A., I. E. Hajj, S. R. Chalamalasetti, G. Ndu, M. Foltin, R. S. Williams, P. Faraboschi, W.-M. W. Hwu, J. P. Strachan, K. Roy, *et al.* (2019). “PUMA: A programmable ultra-efficient memristor-based accelerator for machine learning inference”. In: *Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems*. 715–731. DOI: 10.1145/3297858.3304049.

Astudillo, R. and P. Frazier (2019). “Bayesian optimization of composite functions”. In: *The International Conference on Machine Learning (ICML)*. 354–363.

Baik, S., M. Choi, J. Choi, H. Kim, and K. M. Lee (2020). “Meta-learning with adaptive hyperparameters”. In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Baker, B., O. Gupta, N. Naik, and R. Raskar (2017). “Designing neural network architectures using reinforcement learning”. In: *International Conference on Learning Representations (ICLR)*.

Baker, B., O. Gupta, R. Raskar, and N. Naik (2018). “Accelerating neural architecture search using performance prediction”. In: *International Conference on Learning Representations (ICLR) Workshop*.

Ballard, D. H. (1987). “Modular learning in neural networks”. In: *AAAI Conference on Artificial Intelligence (AAAI)*. 279–284.

Baydin, A. G., R. Cornish, D. M. Rubio, M. Schmidt, and F. Wood (2018). “Online learning rate adaptation with hypergradient descent”. In: *International Conference on Learning Representations (ICLR)*.

Bayer, J., D. Wierstra, J. Togelius, and J. Schmidhuber (2009). “Evolving memory cell structures for sequence learning”. In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. 755–764. DOI: 10.1007/978-3-642-04277-5_76.

Ben-David, S., J. Blitzer, K. Crammer, F. Pereira, *et al.* (2007). "Analysis of representations for domain adaptation". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. DOI: 10.7551/mitpress/7503.003.0022.

Bender, G., P.-J. Kindermans, B. Zoph, V. Vasudevan, and Q. Le (2018). "Understanding and simplifying one-shot architecture search". In: *The International Conference on Machine Learning (ICML)*. 550–559.

Bender, G., H. Liu, B. Chen, G. Chu, S. Cheng, P.-J. Kindermans, and Q. V. Le (2020). "Can weight sharing outperform random architecture search? An investigation with tunas". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 14323–14332. DOI: 10.1109/CVPR42600.2020.01433.

Bengio, Y. (2000). "Gradient-based optimization of hyperparameters". *Neural Computation*. 12(8): 1889–1900. DOI: 10.1162/089976600300015187.

Bengio, Y., S. Bengio, and J. Cloutier (1990). *Learning a Synaptic Learning Rule*. Citeseer.

Bengio, Y., J. Louradour, R. Collobert, and J. Weston (2009). "Curriculum learning". In: *The International Conference on Machine Learning (ICML)*. 41–48. DOI: 10.1145/1553374.1553380.

Bergstra, J., R. Bardenet, Y. Bengio, and B. Kégl (2011). "Algorithms for hyper-parameter optimization". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. Vol. 24.

Bergstra, J. and Y. Bengio (2012). "Random search for hyper-parameter optimization". *Journal of Machine Learning Research (JMLR)*. 13(2).

Biedenkapp, A., J. Marben, M. Lindauer, and F. Hutter (2018). "CAVE: Configuration assessment, visualization and evaluation". In: *Proceedings of the International Conference on Learning and Intelligent Optimization (LION)*. DOI: 10.1007/978-3-030-05348-2_10.

Box, G. E. and D. R. Cox (1964). "An analysis of transformations". *Journal of the Royal Statistical Society: Series B (Methodological)*. 26(2): 211–243. DOI: 10.1111/j.2517-6161.1964.tb00553.x.

Breiman, L. (2001). “Statistical modeling: The two cultures (with comments and a rejoinder by the author)”. *Statistical Science*. 16(3): 199–231. DOI: 10.1214/ss/1009213726.

Brock, A., T. Lim, J. M. Ritchie, and N. Weston (2018). “SMASH: One-shot model architecture search through hypernetworks”. In: *International Conference on Learning Representations (ICLR)*.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.* (2020). “Language models are few-shot learners”. In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. Vol. 33. 1877–1901.

Bulat, A., B. Martinez, and G. Tzimiropoulos (2020). “BATS: Binary architecture search”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. DOI: 10.1007/978-3-030-58592-1_19.

Cai, H., C. Gan, T. Wang, Z. Zhang, and S. Han (2020). “Once-for-all: Train one network and specialize it for efficient deployment”. In: *International Conference on Learning Representations (ICLR)*.

Cai, H., J. Yang, W. Zhang, S. Han, and Y. Yu (2018). “Path-level network transformation for efficient architecture search”. In: *The International Conference on Machine Learning (ICML)*. 678–687. DOI: 10.1609/aaai.v32i1.11709.

Cai, H., L. Zhu, and S. Han (2019). “ProxylessNAS: Direct neural architecture search on target task and hardware”. In: *International Conference on Learning Representations (ICLR)*.

Carpenter, G. A. and S. Grossberg (1987). “A massively parallel architecture for a self-organizing neural pattern recognition machine”. *Computer Vision, Graphics, and Image Processing*. 37(1): 54–115. DOI: 10.1016/S0734-189X(87)80014-2.

Caruana, R., A. Niculescu-Mizil, G. Crew, and A. Ksikes (2004). “Ensemble selection from libraries of models”. In: *The International Conference on Machine Learning (ICML)*. 18. DOI: 10.1145/1015330.1015432.

Celik, B. and J. Vanschoren (2021). “Adaptation strategies for automated machine learning on evolving data”. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. 43(9): 3067–3078. DOI: 10.1109/TPAMI.2021.3062900.

Chandola, V., A. Banerjee, and V. Kumar (2009). "Anomaly detection: A survey". *ACM Computing Surveys (CSUR)*. 41(3): 1–58. DOI: 10.1145/1541880.1541882.

Chandrashekaran, A. and I. R. Lane (2017). "Speeding up hyper-parameter optimization by extrapolation of learning curves using previous builds". In: *The Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*. 477–492. DOI: 10.1007/978-3-319-71249-9_29.

Chen, W., X. Gong, and Z. Wang (2021b). "Neural architecture search on imagenet in four gpu hours: A theoretically inspired perspective". In: *International Conference on Learning Representations (ICLR)*.

Chen, T., I. Goodfellow, and J. Shlens (2016b). "Net2Net: Accelerating learning via knowledge transfer". In: *International Conference on Learning Representations (ICLR)*.

Chen, Y., M. W. Hoffman, S. G. Colmenarejo, M. Denil, T. P. Lillicrap, M. Botvinick, and N. Freitas (2017). "Learning to learn without gradient descent by gradient descent". In: *The International Conference on Machine Learning (ICML)*. 748–756.

Chen, Y.-H., T. Krishna, J. S. Emer, and V. Sze (2016a). "Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks". *IEEE Journal of Solid-State Circuits*. 52(1): 127–138. DOI: 10.1109/JSSC.2016.2616357.

Chen, X., C. Liang, D. Huang, E. Real, K. Wang, Y. Liu, H. Pham, X. Dong, T. Luong, C.-J. Hsieh, *et al.* (2023). "Symbolic discovery of optimization algorithms". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Chen, T., T. Moreau, Z. Jiang, L. Zheng, E. Yan, H. Shen, M. Cowan, L. Wang, Y. Hu, L. Ceze, *et al.* (2018). "TVM: An automated end-to-end optimizing compiler for deep learning". In: *The USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. 578–594.

Chen, M., K. Wu, B. Ni, H. Peng, B. Liu, J. Fu, H. Chao, and H. Ling (2021a). "Searching the search space of vision transformer". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Chen, Y., T. Yang, X. Zhang, G. Meng, X. Xiao, and J. Sun (2019). "DetNAS: Backbone search for object detection". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 6642–6652.

Choi, K., D. Hong, H. Yoon, J. Yu, Y. Kim, and J. Lee (2021). "DANCE: Differentiable accelerator/network co-exploration". In: *The ACM/ESDA/IEEE Design Automation Conference (DAC)*. 337–342. DOI: 10.1109/DAC18074.2021.9586121.

Cireşan, D. C., A. Giusti, L. M. Gambardella, and J. Schmidhuber (2013). "Mitosis detection in breast cancer histology images with deep neural networks". In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. 411–418. DOI: 10.1007/978-3-642-40763-5_51.

Conn, A. R., K. Scheinberg, and L. N. Vicente (2009). *Introduction to Derivative-Free Optimization*. SIAM. DOI: 10.1137/1.9780898718768.

Cubuk, E. D., B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le (2019). "AutoAugment: Learning augmentation strategies from data". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 113–123. DOI: 10.1109/CVPR.2019.00020.

Cubuk, E. D., B. Zoph, J. Shlens, and Q. V. Le (2020). "RandAugment: Practical automated data augmentation with a reduced search space". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR) Workshop*. 702–703. DOI: 10.1109/CVPRW50498.2020.00359.

Dai, X., A. Wan, P. Zhang, B. Wu, Z. He, Z. Wei, K. Chen, Y. Tian, M. Yu, P. Vajda, *et al.* (2021). "FBNetV3: Joint architecture-recipe search using neural acquisition function". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 16276–16285.

Daxberger, E. A., A. Makarova, M. Turchetta, and A. Krause (2020). "Mixed-variable Bayesian optimization". In: *International Joint Conferences on Artificial Intelligence (IJCAI)*. 2633–2639. DOI: 10.24963/ijcai.2020/365.

De Jong, K. (2016). "Evolutionary computation: A unified approach". In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. 185–199. DOI: 10.1145/2908961.2926973.

Dikov, G. and J. Bayer (2019). "Bayesian learning of neural network architectures". In: *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. 730–738.

Domhan, T., J. T. Springenberg, and F. Hutter (2015). "Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves". In: *International Joint Conferences on Artificial Intelligence (IJCAI)*.

Domke, J. (2012). "Generic methods for optimization-based modeling". In: *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. 318–326.

Dong, X., J. Huang, Y. Yang, and S. Yan (2017). "More is less: A more complicated network with less inference complexity". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 5840–5848. DOI: 10.1109/CVPR.2017.205.

Dong, X., L. Liu, K. Musial, and B. Gabrys (2021a). "NATS-Bench: Benchmarking NAS algorithms for architecture topology and size". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. DOI: 10.1109/TPAMI.2021.3054824.

Dong, X., M. Tan, A. W. Yu, D. Peng, B. Gabrys, and Q. V. Le (2021b). "AutoHAS: Efficient hyperparameter and architecture search". In: *International Conference on Learning Representations (ICLR) Workshop*.

Dong, X. and Y. Yang (2019a). "Network pruning via transformable architecture search". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 760–771.

Dong, X. and Y. Yang (2019b). "One-shot neural architecture search via self-evaluated template network". In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 3681–3690. DOI: 10.1109/ICCV.2019.00378.

Dong, X. and Y. Yang (2019c). "Searching for a robust neural architecture in four GPU hours". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 1761–1770. DOI: 10.1109/CVPR.2019.00186.

Dong, X. and Y. Yang (2019d). "Teacher supervises students how to learn from partially labeled images for facial landmark detection". In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 783–792. DOI: 10.1109/ICCV.2019.00087.

Dong, X. and Y. Yang (2020). "NAS-bench-201: Extending the scope of reproducible neural architecture search". In: *International Conference on Learning Representations (ICLR)*.

Du, Z., X. Wang, H. Yang, J. Zhou, and J. Tang (2019). "Sequential scenario-specific meta learner for online recommendation". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 2895–2904.

Duan, Y., J. Schulman, X. Chen, P. L. Bartlett, I. Sutskever, and P. Abbeel (2016). "RL2: Fast reinforcement learning via slow reinforcement learning". *arXiv preprint arXiv:1611.02779*.

Dura-Bernal, S., B. A. Suter, P. Gleeson, M. Cantarelli, A. Quintana, F. Rodriguez, D. J. Kedziora, G. L. Chadderdon, C. C. Kerr, S. A. Neymotin, R. A. McDougal, M. Hines, G. M. Shepherd, and W. W. Lytton (2019). "NetPyNE, a tool for data-driven multiscale modeling of brain circuits". *eLife*. 8(Apr.). DOI: 10.7554/eLife.44494.

Eggensperger, K., F. Hutter, H. H. Hoos, and K. Leyton-Brown (2014). "Surrogate benchmarks for hyperparameter optimization". In: *The European Conference on Artificial Intelligence (ECAI) Workshop*. 24–31.

Eggensperger, K., P. Müller, N. Mallik, M. Feurer, R. Sass, A. Klein, N. Awad, M. Lindauer, and F. Hutter (2021). "HPOBench: A collection of reproducible multi-fidelity benchmark problems for HPO". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Eimer, T., A. Biedenkapp, M. Reimer, S. Adriaensen, F. Hutter, and M. Lindauer (2021). "DACBench: A benchmark library for dynamic algorithm configuration". In: *International Joint Conferences on Artificial Intelligence (IJCAI)*. 1668–1674. DOI: 10.24963/ijcai.2021/230.

Elsken, T., J. H. Metzen, and F. Hutter (2019). "Neural architecture search: A survey". *Journal of Machine Learning Research (JMLR)*. 20(55): 1–21. DOI: 10.1007/978-3-030-05318-5_11.

Epic Games (2019). “Unreal engine”. URL: https://www.unrealengine.com.

Erickson, N., J. Mueller, A. Shirkov, H. Zhang, P. Larroy, M. Li, and A. Smola (2020). “Autogluon-tabular: Robust and accurate automl for structured data”. In: *The International Conference on Machine Learning (ICML) Workshop*.

Facebook AI Research (n.d.). “Papers with code”. URL: https://paperswithcode.com.

Facebook Inc. (2021). “PyTorch V1.10.0”. URL: https://github.com/pytorch/pytorch/releases/tag/v1.10.0.

Falkner, S., A. Klein, and F. Hutter (2018). “BOHB: Robust and efficient hyperparameter optimization at scale”. In: *The International Conference on Machine Learning (ICML)*. 1437–1446.

Fan, Y., F. Tian, T. Qin, X.-Y. Li, and T.-Y. Liu (2018). “Learning to teach”. In: *International Conference on Learning Representations (ICLR)*.

Fang, J., Y. Sun, Q. Zhang, Y. Li, W. Liu, and X. Wang (2020). “Densely connected search space for more flexible neural architecture search”. In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 10628–10637. DOI: 10.1109/CVPR42600.2020.01064.

Feurer, M. and F. Hutter (2019). “Hyperparameter optimization”. In: *Automated Machine Learning*. Springer, Cham. 3–33. DOI: 10.1007/978-3-030-05318-5_1.

Figurnov, M., M. D. Collins, Y. Zhu, L. Zhang, J. Huang, D. Vetrov, and R. Salakhutdinov (2017). “Spatially adaptive computation time for residual networks”. In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. DOI: 10.1109/CVPR.2017.194.

Finn, C., P. Abbeel, and S. Levine (2017). “Model-agnostic meta-learning for fast adaptation of deep networks”. In: *The International Conference on Machine Learning (ICML)*. 1126–1135.

Forrester, A. I., A. Sóbester, and A. J. Keane (2007). “Multi-fidelity optimization via surrogate modelling”. In: *Mathematical, Physical and Engineering Sciences*. Vol. 463. 3251–3269. DOI: 10.1098/rspa.2007.1900.

Frankle, J. and M. Carbin (2018). “The lottery ticket hypothesis: Finding sparse, trainable neural networks”. In: *The International Conference on Machine Learning (ICML)*.

Fukushima, K. (2013). “Training multi-layered neural network neocognitron”. *Neural Networks*. 40: 18–31. DOI: 10.1016/j.neunet.2013.01.001.

Fukushima, K. and S. Miyake (1982). “Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition”. In: *Competition and Cooperation in Neural Nets*. Springer. 267–285. DOI: 10.1007/978-3-642-46466-9_18.

Gama, J., I. Žliobaitė, A. Bifet, M. Pechenizkiy, and A. Bouchachia (2014). “A survey on concept drift adaptation”. *ACM Computing Surveys (CSUR)*. 46(4): 1–37. DOI: 10.1145/2523813.

Ganin, Y. and V. Lempitsky (2015). “Unsupervised domain adaptation by backpropagation”. In: *The International Conference on Machine Learning (ICML)*. 1180–1189.

Gao, J., M. Galley, and L. Li (2018). “Neural approaches to conversational AI”. In: *The International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1371–1374.

Gelernter, D. (1993). *Mirror Worlds: Or the Day Software Puts the Universe in a Shoebox... How It Will Happen and What it Will Mean*. Oxford University Press.

Gemp, I., G. Theocharous, and M. Ghavamzadeh (2017). “Automated data cleansing through meta-learning”. In: *AAAI Conference on Artificial Intelligence (AAAI)*. 4760–4761. DOI: 10.1609/aaai.v31i2.19107.

Gencoglu, O., M. van Gils, E. Guldogan, C. Morikawa, M. Süzen, M. Gruber, J. Leinonen, and H. Huttunen (2019). “HARK side of deep learning–from grad student descent to automated machine learning”. *arXiv preprint arXiv:1904.07633*.

Ghiasi, G., T.-Y. Lin, and Q. V. Le (2019). “NAS-FPN: Learning scalable feature pyramid architecture for object detection”. In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 7036–7045. DOI: 10.1109/CVPR.2019.00720.

Gil, Y., M. Greaves, J. Hendler, and H. Hirsh (2014). "Amplify scientific discovery with artificial intelligence". *Science*. 346(6206): 171–172. DOI: 10.1126/science.1259439.

Goldberg, D. E. and K. Deb (1991). "A comparative analysis of selection schemes used in genetic algorithms". In: *Foundations of Genetic Algorithms*. Vol. 1. 69–93. DOI: 10.1016/B978-0-08-050684-5.50008-2.

Golovin, D., B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley (2017). "Google vizier: A service for black-box optimization". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 1487–1495. DOI: 10.1145/3097983.3098043.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio (2014). "Generative adversarial networks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Google (2017). "Edge TPU". URL: https://cloud.google.com/edge-tpu.

Grossberg, S. (2020). "Toward autonomous adaptive intelligence: Building upon neural models of how brains make minds". *IEEE Transactions on Systems, Man, and Cybernetics: Systems (TSMC)*. 51(1): 51–75. DOI: 10.1109/TSMC.2020.3041476.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi (2019). "A survey of methods for explaining black box models". *ACM Computing Surveys (CSUR)*. 51(5): 1–42. DOI: 10.1145/3236009.

Gunasekar, S., Y. Zhang, J. Aneja, C. C. T. Mendes, A. Del Giorno, S. Gopi, M. Javaheripi, P. Kauffmann, G. de Rosa, O. Saarikivi, *et al.* (2023). "Textbooks are all you need". *arXiv preprint arXiv:2306.11644*.

Guo, Z., X. Zhang, H. Mu, W. Heng, Z. Liu, Y. Wei, and J. Sun (2020). "Single path one-shot neural architecture search with uniform sampling". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer. 544–560. DOI: 10.1007/978-3-030-58517-4_32.

Ha, D., A. Dai, and Q. V. Le (2017). "Hypernetworks". In: *International Conference on Learning Representations (ICLR)*.

Han, Y., G. Huang, S. Song, L. Yang, H. Wang, and Y. Wang (2021). "Dynamic neural networks: A survey". *arXiv preprint arXiv:2102.04906*.

Han, S., X. Liu, H. Mao, J. Pu, A. Pedram, M. A. Horowitz, and W. J. Dally (2016a). "EIE: Efficient inference engine on compressed deep neural network". *ACM SIGARCH Computer Architecture News*. 44(3): 243–254. DOI: 10.1145/3007787.3001163.

Han, S., H. Mao, and W. J. Dally (2016b). "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding". In: *International Conference on Learning Representations (ICLR)*.

Hansen, N. and A. Ostermeier (2001). "Completely derandomized self-adaptation in evolution strategies". *Evolutionary Computation*. 9(2): 159–195. DOI: 10.1162/106365601750190398.

He, K., G. Gkioxari, P. Dollár, and R. Girshick (2017a). "Mask r-cnn". In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 2961–2969.

He, X., L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua (2017b). "Neural collaborative filtering". In: *Proceedings of the International Conference on World Wide Web (WWW)*. 173–182.

He, K., X. Zhang, S. Ren, and J. Sun (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 770–778.

Hendrycks, D. and K. Gimpel (2017). "A baseline for detecting misclassified and out-of-distribution examples in neural networks". In: *International Conference on Learning Representations (ICLR)*.

Hiesinger, P. R. (2021). *The Self-Assembling Brain: How Neural Networks Grow Smarter*. Princeton University Press. DOI: 10.1515/9780691215518.

Hinton, G. E. and D. C. Plaut (1987). "Using fast weights to deblur old memories". In: *Proceedings of the 9th Annual Conference of the Cognitive Science Society*. 177–186.

Hinton, G., O. Vinyals, and J. Dean (2014). "Distilling the knowledge in a neural network". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS) Workshop*.

Hochreiter, S. and J. Schmidhuber (1997). "Long short-term memory". *Neural Computation*. 9(8): 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

Hochreiter, S., A. S. Younger, and P. R. Conwell (2001). "Learning to learn using gradient descent". In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*. 87–94. DOI: 10.1007/3-540-44668-0_13.

Hofmann, T., B. Schölkopf, and A. J. Smola (2008). "Kernel methods in machine learning". *The Annals of Statistics*: 1171–1220. DOI: 10.1214/009053607000000677.

Hopfield, J. J. (1982). "Neural networks and physical systems with emergent collective computational abilities". *Proceedings of the National Academy of Sciences*. 79(8): 2554–2558. DOI: 10.1073/pnas.79.8.2554.

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer feedforward networks are universal approximators". *Neural Networks*. 2(5): 359–366. DOI: 10.1016/0893-6080(89)90020-8.

Houthooft, R., R. Y. Chen, P. Isola, B. C. Stadie, F. Wolski, J. Ho, and P. Abbeel (2018). "Evolved policy gradients". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Hu, H., J. Langford, R. Caruana, E. Horvitz, and D. Dey (2018). "Macro neural architecture search revisited". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS) Workshop*.

Huang, G., Z. Liu, L. Van Der Maaten, and K. Q. Weinberger (2017). "Densely connected convolutional networks". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 4700–4708. DOI: 10.1109/CVPR.2017.243.

Hutter, F., H. H. Hoos, and K. Leyton-Brown (2011). "Sequential model-based optimization for general algorithm configuration". In: *Proceedings of the International Conference on Learning and Intelligent Optimization (LION)*. 507–523. DOI: 10.1007/978-3-642-25566-3_40.

Hutter, F., H. H. Hoos, K. Leyton-Brown, and T. Stützle (2009). "ParamILS: An automatic algorithm configuration framework". *Journal of Artificial Intelligence Research (JAIR)*. 36: 267–306. DOI: 10.1613/jair.2861.

Hwang, C.-L. and A. S. M. Masud (2012). *Multiple Objective Decision Making—Methods and Applications: A State-of-the-Art Survey*. Vol. 164. Springer Science & Business Media.

Ioffe, S. and C. Szegedy (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift". In: *The International Conference on Machine Learning (ICML)*. 448–456.

Ivakhnenko, A. G. (1971). "Polynomial theory of complex systems". *IEEE Transactions on Systems, Man, and Cybernetics (SMC)*. 4: 364–378. DOI: 10.1109/TSMC.1971.4308320.

Jacot, A., F. Gabriel, and C. Hongler (2018). "Neural tangent kernel: Convergence and generalization in neural networks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 8580–8589.

Jaderberg, M., V. Dalibard, S. Osindero, W. M. Czarnecki, J. Donahue, A. Razavi, O. Vinyals, T. Green, I. Dunning, K. Simonyan, *et al.* (2017). "Population based training of neural networks". *arXiv preprint arXiv:1711.09846*.

Jamieson, K. and A. Talwalkar (2016). "Non-stochastic best arm identification and hyperparameter optimization". In: *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. 240–248.

Jang, E., S. Gu, and B. Poole (2017). "Categorical reparameterization with gumbel-softmax". In: *International Conference on Learning Representations (ICLR)*.

Jenatton, R., C. Archambeau, J. González, and M. Seeger (2017). "Bayesian optimization with tree-structured dependencies". In: *The International Conference on Machine Learning (ICML)*. 1655–1664.

Ji, S., W. Xu, M. Yang, and K. Yu (2012). "3D convolutional neural networks for human action recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. 35(1): 221–231. DOI: 10.1109/TPAMI.2012.59.

Jiang, W., L. Yang, E. H.-M. Sha, Q. Zhuge, S. Gu, S. Dasgupta, Y. Shi, and J. Hu (2020). "Hardware/software co-exploration of neural architectures". *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*. 39(12): 4805–4815. DOI: 10.1109/TCAD.2020.2986127.

Jin, H., Q. Song, and X. Hu (2019). "Auto-keras: An efficient neural architecture search system". In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1946–1956.

Jomaa, H. S., J. Grabocka, and L. Schmidt-Thieme (2019). "Hyp-RL: Hyperparameter optimization by reinforcement learning". *arXiv preprint arXiv:1906.11527*.

Jozefowicz, R., W. Zaremba, and I. Sutskever (2015). "An empirical exploration of recurrent network architectures". In: *The International Conference on Machine Learning (ICML)*. 2342–2350.

Kadlec, P. and B. Gabrys (2009). "Architecture for development of adaptive on-line prediction models". *Memetic Computing*. 1(4): 241–269. DOI: 10.1007/s12293-009-0017-8.

Kandasamy, K., G. Dasarathy, J. Oliva, J. Schneider, and B. Póczos (2016). "Gaussian process bandit optimisation with multi-fidelity evaluations". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 1000–1008.

Kandasamy, K., W. Neiswanger, J. Schneider, B. Poczos, and E. P. Xing (2018). "Neural architecture search with Bayesian optimisation and optimal transport". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Karnin, Z., T. Koren, and O. Somekh (2013). "Almost optimal exploration in multi-armed bandits". In: *The International Conference on Machine Learning (ICML)*. 1238–1246.

Kedziora, D. J., K. Musial, and B. Gabrys (2020). "AutonoML: Towards an integrated framework for autonomous machine learning". *arXiv preprint arXiv:2012.12600*.

Kingma, D. P. and J. Ba (2015). "Adam: A method for stochastic optimization". In: *International Conference on Learning Representations (ICLR)*.

Kirkpatrick, J., R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell (2017). "Overcoming catastrophic forgetting in neural networks". *Proceedings of the National Academy of Sciences*. 114(13): 3521–3526. DOI: 10.1073/pnas.1611835114.

Kirsch, L., S. van Steenkiste, and J. Schmidhuber (2020). "Improving generalization in meta reinforcement learning using learned objectives". In: *International Conference on Learning Representations (ICLR)*.

Klein, A. and F. Hutter (2019). "Tabular benchmarks for joint architecture and hyperparameter optimization". *arXiv preprint arXiv:1905.04970*.

Konyushkova, K., S. Raphael, and P. Fua (2017). "Learning active learning from data". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 4228–4238.

Krawczyk, B., L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak (2017). "Ensemble learning for data stream analysis: A survey". *Information Fusion*. 37: 132–156. DOI: 10.1016/j.inffus.2017.02.004.

Krizhevsky, A., G. Hinton, *et al.* (2009). *Learning Multiple Layers of Features from Tiny Images*. Citeseer.

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012). "Imagenet classification with deep convolutional neural networks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 1097–1105.

Larsen, J., L. K. Hansen, C. Svarer, and M. Ohlsson (1996). "Design and regularization of neural networks: The optimal use of a validation set". In: *IEEE Workshop on Neural Networks for Signal Processing (NNSP)*. 62–71.

Larsen, J., C. Svarer, L. N. Andersen, and L. K. Hansen (2002). "Adaptive regularization in neural network modeling". In: *Neural Networks: Tricks of the Trade*. DOI: 10.1007/3-540-49430-8_6.

Lazzaro, J., S. Ryckebusch, M. A. Mahowald, and C. A. Mead (1988). "Winner-take-all networks of $O(n)$ complexity". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. Vol. 1. DOI: 10.21236/ADA451466.

LeCun, Y. (1985). "A learning scheme for asymmetric threshold networks". *Proceedings of Cognitiva*. 85(537): 599–604.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". *Nature*. 521(7553): 436–444. DOI: 10.1038/nature14539.

LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel (1989). "Backpropagation applied to handwritten zip code recognition". *Neural Computation.* 1(4): 541–551. DOI: 10.1162/neco.1989.1.4.541.

Li, Z. and D. Hoiem (2017). "Learning without forgetting". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI).* 40(12): 2935–2947. DOI: 10.1109/TPAMI.2017.2773081.

Li, Y., G. Hu, Y. Wang, T. Hospedales, N. M. Robertson, and Y. Yang (2020). "DADA: Differentiable automatic data augmentation". In: *Proceedings of the European Conference on Computer Vision (ECCV).* Springer. 580–595. DOI: 10.1007/978-3-030-58542-6_35.

Li, L., K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar (2018). "Hyperband: A novel bandit-based approach to hyperparameter optimization". *Journal of Machine Learning Research (JMLR).* 18(1): 6765–6816.

Li, B., X. Jiang, D. Bai, Y. Zhang, N. Zheng, X. Dong, L. Liu, Y. Yang, and D. Li (2021a). "Full-cycle energy consumption benchmark for low-carbon computer vision". *arXiv preprint arXiv:2108.13465.*

Li, K. and J. Malik (2017). "Learning to optimize". In: *International Conference on Learning Representations (ICLR).*

Li, L. and A. Talwalkar (2020). "Random search and reproducibility for neural architecture search". In: *Uncertainty in Artificial Intelligence.* 367–377.

Li, C., Z. Yu, Y. Fu, Y. Zhang, Y. Zhao, H. You, Q. Yu, Y. Wang, and Y. Lin (2021b). "HW-NAS-Bench: Hardware-aware neural architecture search benchmark". In: *International Conference on Learning Representations (ICLR).*

Li, A., O. Spyra, S. Perel, V. Dalibard, M. Jaderberg, C. Gu, D. Budden, T. Harley, and P. Gupta (2019). "A generalized framework for population based training". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD).* 1791–1799. DOI: 10.1145/3292500.3330649.

Lim, S., I. Kim, T. Kim, C. Kim, and S. Kim (2019). "Fast autoaugment". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS).*

Lindauer, M. and F. Hutter (2020). "Best practices for scientific research on neural architecture search". *Journal of Machine Learning Research (JMLR)*. 21(243): 1–18.

Linnainmaa, S. (1970). "The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors". *Master's Thesis (in Finnish), Univ. Helsinki*: 6–7.

Little, W. A. (1974). "The existence of persistent states in the brain". *Mathematical Biosciences*. 19(1–2): 101–120. DOI: 10.1016/0025-5564(74)90031-5.

Liu, H., A. Brock, K. Simonyan, and Q. V. Le (2020). "Evolving normalization-activation layers". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Liu, C., L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei (2019a). "Auto-DeepLab: Hierarchical neural architecture search for semantic image segmentation". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 82–92.

Liu, H., K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu (2018b). "Hierarchical representations for efficient architecture search". In: *International Conference on Learning Representations (ICLR)*.

Liu, H., K. Simonyan, and Y. Yang (2019b). "DARTS: Differentiable Architecture Search". In: *International Conference on Learning Representations (ICLR)*.

Liu, B., M. Wang, H. Foroosh, M. Tappen, and M. Pensky (2015). "Sparse convolutional neural networks". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 806–814. DOI: 10.1109/CVPR.2015.7298681.

Liu, L., T. Zhou, G. Long, J. Jiang, X. Dong, and C. Zhang (2021). "Isometric propagation network for generalized zero-shot learning". In: *International Conference on Learning Representations (ICLR)*.

Liu, L., T. Zhou, G. Long, J. Jiang, and C. Zhang (2019c). "Learning to propagate for graph meta-learning". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Liu, C., B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang, and K. Murphy (2018a). "Progressive neural architecture search". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 19–34. DOI: 10.1007/978-3-030-01246-5_2.

Lorraine, J., P. Vicol, and D. Duvenaud (2020). "Optimizing millions of hyperparameters by implicit differentiation". In: *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. 1540–1552.

Loshchilov, I. and F. Hutter (2016). "CMA-ES for hyperparameter optimization of deep neural networks". In: *International Conference on Learning Representations (ICLR) Workshop*.

Luketina, J., M. Berglund, K. Greff, and T. Raiko (2016). "Scalable gradient-based tuning of continuous regularization hyperparameters". In: *The International Conference on Machine Learning (ICML)*. 2952–2960.

Ma, E. (2019). "NLP Augmentation". URL: https://github.com/makcedward/nlpaug.

MacKay, M., P. Vicol, J. Lorraine, D. Duvenaud, and R. Grosse (2019). "Self-tuning networks: Bilevel optimization of hyperparameters using structured best-response functions". In: *International Conference on Learning Representations (ICLR)*.

Maclaurin, D., D. Duvenaud, and R. Adams (2015). "Gradient-based hyperparameter optimization through reversible learning". In: *The International Conference on Machine Learning (ICML)*. 2113–2122.

Madrid, J. G., H. J. Escalante, E. F. Morales, W.-W. Tu, Y. Yu, L. Sun-Hosoya, I. Guyon, and M. Sebag (2018). "Towards AutoML in the presence of drift: First results". In: *The International Conference on Machine Learning (ICML) Workshop*. DOI: 10.52591/lxai201812039.

Marler, R. T. and J. S. Arora (2004). "Survey of multi-objective optimization methods for engineering". *Structural and Multidisciplinary Optimization*. 26(6): 369–395. DOI: 10.1007/s00158-003-0368-6.

Martel, Y., A. Roßmann, E. Sultanow, O. Weiß, M. Wissel, F. Pelzel, and M. Seßler (2021). "Software architecture best practices for enterprise artificial intelligence". *INFORMATIK*.

McCloskey, M. and N. J. Cohen (1989). "Catastrophic interference in connectionist networks: The sequential learning problem". In: *Psychology of Learning and Motivation*. Vol. 24. Elsevier. 109–165. DOI: 10.1016/S0079-7421(08)60536-8.

McCulloch, W. S. and W. Pitts (1943). "A logical calculus of the ideas immanent in nervous activity". *The Bulletin of Mathematical Biophysics*. 5(4): 115–133. DOI: 10.1007/BF02478259.

Mellor, J., J. Turner, A. Storkey, and E. J. Crowley (2021). "Neural architecture search without training". In: *The International Conference on Machine Learning (ICML)*.

Michie, D., D. J. Spiegelhalter, C. Taylor, *et al.* (1994). "Machine learning". *Neural and Statistical Classification*. 13(1994): 1–298.

Microsoft Corporation (2018). "Neural Network Intelligence". URL: https://github.com/microsoft/nni.

Milutinovic, M., A. G. Baydin, R. Zinkov, W. Harvey, D. Song, F. Wood, and W. Shen (2017). "End-to-end training of differentiable pipelines across machine learning frameworks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS) Workshop*.

Mirhoseini, A., A. Goldie, H. Pham, B. Steiner, Q. V. Le, and J. Dean (2018). "A hierarchical model for device placement". In: *International Conference on Learning Representations (ICLR)*.

Mirhoseini, A., A. Goldie, M. Yazgan, J. W. Jiang, E. Songhori, S. Wang, Y.-J. Lee, E. Johnson, O. Pathak, A. Nazi, *et al.* (2021). "A graph placement methodology for fast chip design". *Nature*. 594(7862): 207–212. DOI: 10.1038/s41586-021-03544-w.

Mirhoseini, A., H. Pham, Q. V. Le, B. Steiner, R. Larsen, Y. Zhou, N. Kumar, M. Norouzi, S. Bengio, and J. Dean (2017). "Device placement optimization with reinforcement learning". In: *The International Conference on Machine Learning (ICML)*. 2430–2439.

Močkus, J. (1975). "On Bayesian methods for seeking the extremum". In: *Optimization Techniques IFIP Technical Conference*. Springer. 400–404. DOI: 10.1007/978-3-662-38527-2_55.

Nardi, L., D. Koeplinger, and K. Olukotun (2019). "Practical design space exploration". In: *IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS)*. 347–358. DOI: 10.1109/MASCOTS.2019.00045.

Nguyen, T.-D., D. J. Kedziora, K. Musial, and B. Gabrys (2021). "Exploring opportunistic meta-knowledge to reduce search spaces for automated machine learning". In: *International Joint Conference on Neural Network (IJCNN)*. DOI: 10.1109/IJCNN52387.2021.9533431.

Niu, T. and M. Bansal (2019). "Automatically learning data augmentation policies for dialogue tasks". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1317–1323. DOI: 10.18653/v1/D19-1132.

Panayotov, V., G. Chen, D. Povey, and S. Khudanpur (2015). "Librispeech: An asr corpus based on public domain audio books". In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 5206–5210. DOI: 10.1109/ICASSP.2015.7178964.

Parisi, G. I., R. Kemker, J. L. Part, C. Kanan, and S. Wermter (2019). "Continual lifelong learning with neural networks: A review". *Neural Networks*. 113: 54–71. DOI: 10.1016/j.neunet.2019.01.012.

Parkhi, O. M., A. Vedaldi, and A. Zisserman (2015). "Deep face recognition". In: *Proceedings of the British Machine Vision Conference (BMVC)*. 41.1–41.12. DOI: 10.5244/C.29.41.

Parsa, M., A. Ankit, A. Ziabari, and K. Roy (2019). "PABO: Pseudo agent-based multi-objective bayesian hyperparameter optimization for efficient neural accelerator design". In: *The IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. 1–8. DOI: 10.1109/ICCAD45719.2019.8942046.

Patterson, D., J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean (2021). "Carbon emissions and large neural network training". *arXiv preprint arXiv:2104.10350*.

Pedregosa, F. (2016). "Hyperparameter optimization with approximate gradient". In: *The International Conference on Machine Learning (ICML)*. 737–746.

Peng, Y., Y. Bao, Y. Chen, C. Wu, and C. Guo (2018). "Optimus: An efficient dynamic resource scheduler for deep learning clusters". In: *Proceedings of the EuroSys Conference (EuroSys)*. 1–14. DOI: 10.1145/3190508.3190517.

Peng, D., X. Dong, E. Real, M. Tan, Y. Lu, G. Bender, H. Liu, A. Kraft, C. Liang, and Q. Le (2020). "PyGlove: Symbolic programming for automated machine learning". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Pham, H., M. Guan, B. Zoph, Q. Le, and J. Dean (2018). "Efficient neural architecture search via parameters sharing". In: *The International Conference on Machine Learning (ICML)*. 4095–4104.

Pineau, J., P. Vincent-Lamarre, K. Sinha, V. Larivière, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and H. Larochelle (2021). "Improving reproducibility in machine learning research: A report from the NeurIPS 2019 reproducibility program". *Journal of Machine Learning Research (JMLR)*. 22.

Pollack, J. B. (1989). "Implications of recursive distributed representations". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 527–536.

Quinlan, J. R. (1986). "Induction of decision trees". *Machine Learning*. 1(1): 81–106. DOI: 10.1007/BF00116251.

Radosavovic, I., R. P. Kosaraju, R. Girshick, K. He, and P. Dollár (2020). "Designing network design spaces". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 10428–10436. DOI: 10.1109/CVPR42600.2020.01044.

Rajpurkar, P., J. Zhang, K. Lopyrev, and P. Liang (2016). "Squad: 100,000+ questions for machine comprehension of text". *arXiv preprint arXiv:1606.05250*. DOI: 10.18653/v1/D16-1264.

Ramachandran, P., B. Zoph, and Q. V. Le (2017). "Searching for activation functions". *arXiv preprint arXiv:1710.05941*.

Ravi, S. and H. Larochelle (2017). "Optimization as a model for few-shot learning". In: *International Conference on Learning Representations (ICLR)*.

Reagen, B., J. M. Hernández-Lobato, R. Adolf, M. Gelbart, P. Whatmough, G.-Y. Wei, and D. Brooks (2017). "A case for efficient accelerator design space exploration via bayesian optimization". In: *IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. 1–6. DOI: 10.1109/ISLPED.2017.8009208.

Real, E., A. Aggarwal, Y. Huang, and Q. V. Le (2019). "Regularized evolution for image classifier architecture search". In: *AAAI Conference on Artificial Intelligence (AAAI)*. 4780–4789. DOI: 10.1609/aaai.v33i01.33014780.

Real, E., C. Liang, D. So, and Q. Le (2020). "AutoML-zero: Evolving machine learning algorithms from scratch". In: *The International Conference on Machine Learning (ICML)*. 8007–8019.

Real, E., S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le, and A. Kurakin (2017). "Large-scale evolution of image classifiers". In: *The International Conference on Machine Learning (ICML)*. 2902–2911.

Ren, P., Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, X. Chen, and X. Wang (2021). "A comprehensive survey of neural architecture search: Challenges and solutions". *ACM Computing Surveys (CSUR)*. 54(4): 1–34. DOI: 10.1145/3447582.

Ren, M., W. Zeng, B. Yang, and R. Urtasun (2018). "Learning to reweight examples for robust deep learning". In: *The International Conference on Machine Learning (ICML)*. 4334–4343.

Ricci, F., L. Rokach, and B. Shapira (2011). "Introduction to recommender systems handbook". In: *Recommender Systems Handbook*. Springer. 1–35. DOI: 10.1007/978-0-387-85820-3_1.

Riedmiller, M. and H. Braun (1993). "A direct adaptive method for faster backpropagation learning: The RPROP algorithm". In: *IEEE International Conference on Neural Networks*. 586–591.

Romera-Paredes, B. and P. Torr (2015). "An embarrassingly simple approach to zero-shot learning". In: *The International Conference on Machine Learning (ICML)*. 2152–2161.

Ru, B., A. Alvi, V. Nguyen, M. A. Osborne, and S. Roberts (2020a). "Bayesian optimisation over multiple continuous and categorical inputs". In: *The International Conference on Machine Learning (ICML)*. 8276–8285.

Ru, B., P. Esperanca, and F. Carlucci (2020b). "Neural architecture generator optimization". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1985). "Learning internal representations by error propagation". *Technical Report*. California Univ San Diego La Jolla Inst for Cognitive Science. DOI: 10.21236/ADA164453.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, and L. Fei-Fei (2015). "Imagenet large scale visual recognition challenge". *International Journal of Computer Vision (IJCV)*. 115(3): 211–252. DOI: 10.1007/s11263-015-0816-y.

Rusu, A. A., N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell (2016). "Progressive neural networks". *arXiv preprint arXiv:1606.04671*.

Ruta, D., B. Gabrys, and C. Lemke (2011). "A generic multilevel architecture for time series prediction". *IEEE Transactions on Knowledge and Data Engineering (TKDE)*. 23(3): 350–359. DOI: 10.1109/TKDE.2010.137.

Salimans, T., J. Ho, X. Chen, S. Sidor, and I. Sutskever (2017). "Evolution strategies as a scalable alternative to reinforcement learning". *arXiv preprint arXiv:1703.03864*.

Salvador, M. M., M. Budka, and B. Gabrys (2016). "Towards automatic composition of multicomponent predictive systems". In: *International Conference on Hybrid Artificial Intelligence Systems*. 27–39. DOI: 10.1007/978-3-319-32034-2_3.

Salvador, M. M., M. Budka, and B. Gabrys (2019). "Automatic composition and optimization of multicomponent predictive systems with an extended auto-WEKA". *IEEE Transactions on Automation Science and Engineering (TASE)*. 16(2): 946–959. DOI: 10.1109/TASE.2018.2876430.

Sandler, M., A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen (2018). "Mobilenetv2: Inverted residuals and linear bottlenecks". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 4510–4520. DOI: 10.1109/CVPR.2018.00474.

Schaul, T. and J. Schmidhuber (2010). "Metalearning". *Scholarpedia*. 5(6): 4650. DOI: 10.4249/scholarpedia.4650.

Schmidhuber, J. (1987). "Evolutionary principles in self-referential learning". *PhD thesis*.

Schmidhuber, J. (1992). "Learning to control fast-weight memories: An alternative to dynamic recurrent networks". *Neural Computation*. 4(1): 131–139. DOI: 10.1162/neco.1992.4.1.131.

Schmidhuber, J. (2002). "Bias-optimal incremental problem solving". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Schmidhuber, J. (2015). "Deep learning in neural networks: An overview". *Neural Networks*. 61: 85–117. DOI: 10.1016/j.neunet.2014.09.003.

Schmidhuber, J., J. Zhao, and N. N. Schraudolph (1998). "Reinforcement learning with self-modifying policies". In: *Learning to Learn*. Springer. 293–309. DOI: 10.1007/978-1-4615-5529-2_12.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford, and O. Klimov (2017). "Proximal policy optimization algorithms". *arXiv preprint arXiv:1707.06347*.

Sculley, D., G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison (2015). "Hidden technical debt in machine learning systems". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 2503–2511.

Secret Labs AB Lundh, F., A. Clark Other Contributors (2011). "Python Imaging Library (PIL)".

Shaban, A., C.-A. Cheng, N. Hatch, and B. Boots (2019). "Truncated back-propagation for bilevel optimization". In: *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. 1723–1732.

Shorten, C. and T. M. Khoshgoftaar (2019). "A survey on image data augmentation for deep learning". *Journal of Big Data*. 6(1): 1–48. DOI: 10.1186/s40537-019-0197-0.

Shu, J., Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, and D. Meng (2019). "Meta-weight-net: Learning an explicit mapping for sample weighting". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Siems, J., L. Zimmer, A. Zela, J. Lukasik, M. Keuper, and F. Hutter (2020). "NAS-Bench-301 and the case for surrogate benchmarks for neural architecture search". *arXiv preprint arXiv:2008.09777*.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.* (2016). "Mastering the game of Go with deep neural networks and tree search". *Nature*. 529(7587): 484–489. DOI: 10.1038/nature16961.

Simonyan, K. and A. Zisserman (2015). "Very deep convolutional networks for large-scale image recognition". In: *International Conference on Learning Representations (ICLR)*.

Smith, S. L., P.-J. Kindermans, C. Ying, and Q. V. Le (2018). "Don't decay the learning rate, increase the batch size". In: *International Conference on Learning Representations (ICLR)*.

Smith-Miles, K. A. (2009). "Cross-disciplinary perspectives on meta-learning for algorithm selection". *ACM Computing Surveys (CSUR)*. 41(1): 1–25. DOI: 10.1145/1456650.1456656.

Snell, J., K. Swersky, and R. Zemel (2017). "Prototypical networks for few-shot learning". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 4077–4087.

Snoek, J., H. Larochelle, and R. P. Adams (2012). "Practical Bayesian optimization of machine learning algorithms". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 2951–2959.

So, D., Q. Le, and C. Liang (2019). "The evolved transformer". In: *The International Conference on Machine Learning (ICML)*. 5877–5886.

So, D. R., W. Mańke, H. Liu, Z. Dai, N. Shazeer, and Q. V. Le (2021). "Primer: Searching for efficient transformers for language modeling". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Song, Q., H. Jin, and X. Hu (2022). *Automated machine learning in action*. Manning Publications Co.

Sprechmann, P., S. M. Jayakumar, J. W. Rae, A. Pritzel, A. P. Badia, B. Uria, O. Vinyals, D. Hassabis, R. Pascanu, and C. Blundell (2018). "Memory-based parameter adaptation". In: *International Conference on Learning Representations (ICLR)*.

Strubell, E., A. Ganesh, and A. McCallum (2019). “Energy and policy considerations for deep learning in NLP”. In: *The Annual Meeting of the Association for Computational Linguistics*. DOI: 10.18653/v1/P19-1355.

Such, F. P., A. Rawal, J. Lehman, K. Stanley, and J. Clune (2020). “Generative teaching networks: Accelerating neural architecture search by learning to generate synthetic training data”. In: *The International Conference on Machine Learning (ICML)*. 9206–9216.

Sutton, R. S. (1992). “Adapting bias by gradient descent: An incremental version of delta-bar-delta”. In: *AAAI Conference on Artificial Intelligence (AAAI)*. 171–176.

Sutton, R. S. and A. G. Barto (2018). *Reinforcement Learning: An Introduction*. MIT press.

Süzen, M. (2020). “Equivalence in deep neural networks via conjugate matrix ensembles”. *arXiv preprint arXiv:2006.13687*.

Süzen, M., J. J. Cerdà, and C. Weber (2019). “Periodic spectral ergodicity: A complexity measure for deep neural networks and neural architecture search”. *arXiv preprint arXiv:1911.07831*.

Swersky, K., J. Snoek, and R. P. Adams (2014). “Freeze-thaw Bayesian optimization”. *arXiv preprint arXiv:1406.3896*.

Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich (2015). “Going deeper with convolutions”. In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 1–9. DOI: 10.1109/CVPR.2015.7298594.

Tan, M., B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, and Q. V. Le (2019). “MnasNet: Platform-aware neural architecture search for mobile”. In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 2820–2828.

Tan, M. and Q. V. Le (2021). “EfficientNetV2: Smaller models and faster training”. In: *The International Conference on Machine Learning (ICML)*. Vol. 139. 10096–10106.

Thornton, C., F. Hutter, H. H. Hoos, and K. Leyton-Brown (2013). "Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms". In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. 847–855. DOI: 10.1145/2487575.2487629.

Thrun, S. and T. M. Mitchell (1995). "Lifelong robot learning". *Robotics and Autonomous Systems*. 15(1-2): 25–46. DOI: 10.1016/0921-8890(95)00004-Y.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (2017). "Attention is all you need". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 6000–6010.

Veeriah, V., M. Hessel, Z. Xu, R. Lewis, J. Rajendran, J. Oh, H. van Hasselt, D. Silver, and S. Singh (2019). "Discovery of useful questions as auxiliary tasks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Vincent, P., H. Larochelle, Y. Bengio, and P.-A. Manzagol (2008). "Extracting and composing robust features with denoising autoencoders". In: *The International Conference on Machine Learning (ICML)*. 1096–1103. DOI: 10.1145/1390156.1390294.

Wan, A., X. Dai, P. Zhang, Z. He, Y. Tian, S. Xie, B. Wu, M. Yu, T. Xu, K. Chen, *et al.* (2020). "FBNetV2: Differentiable neural architecture search for spatial and channel dimensions". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 12965–12974. DOI: 10.1109/CVPR42600.2020.01298.

Wang, J. X., Z. Kurth-Nelson, H. Soyer, J. Z. Leibo, D. Tirumala, R. Munos, C. Blundell, D. Kumaran, and M. M. Botvinick (2017). "Learning to reinforcement learn". In: *Cognitive Science Society (CogSci)*.

Wang, J., C. Lan, C. Liu, Y. Ouyang, W. Zeng, and T. Qin (2021). "Generalizing to unseen domains: A survey on domain generalization". *arXiv preprint arXiv:2103.03097*. DOI: 10.24963/ijcai.2021/628.

Wang, K., Z. Liu, Y. Lin, J. Lin, and S. Han (2019b). "HAQ: Hardware-aware automated quantization with mixed precision". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 8612–8620. DOI: 10.1109/CVPR.2019.00881.

Wang, D., Q. Yang, A. Abdul, and B. Y. Lim (2019a). "Designing theory-driven user-centric explainable AI". In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15. DOI: 10.1145/3290605.3300831.

Wang, T., K. Wang, H. Cai, J. Lin, Z. Liu, H. Wang, Y. Lin, and S. Han (2020). "APQ: Joint search for network architecture, pruning and quantization policy". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 2078–2087. DOI: 10.1109/CVPR42600.2020.00215.

Wang, T., J.-Y. Zhu, A. Torralba, and A. A. Efros (2018). "Dataset distillation". *arXiv preprint arXiv:1811.10959*.

Watkins, C. J. C. H. (1989). "Learning from delayed rewards". *PhD thesis, University of Cambridge*.

Weigend, A. S. and N. A. Gershenfeld (1993). "Results of the time series prediction competition at the Santa Fe Institute". In: *IEEE International Conference on Neural Networks*. 1786–1793.

Wen, W., H. Liu, Y. Chen, H. Li, G. Bender, and P.-J. Kindermans (2020). "Neural predictor for neural architecture search". In: *European Conference on Computer Vision*. Springer. 660–676. DOI: 10.1007/978-3-030-58526-6_39.

White, C., W. Neiswanger, and Y. Savani (2021). "BANANAS: Bayesian optimization with neural architectures for neural architecture search". In: *AAAI Conference on Artificial Intelligence (AAAI)*. DOI: 10.1609/aaai.v35i12.17233.

White, M. and A. White (2016). "A greedy approach to adapting the trace parameter for temporal difference learning". In: *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. 557–565.

Williams, R. J. (1992). "Simple statistical gradient-following algorithms for connectionist reinforcement learning". *Machine Learning*. 8(3-4): 229–256. DOI: 10.1007/BF00992696.

Wistuba, M., A. Rawat, and T. Pedapati (2019). "A survey on neural architecture search". *arXiv preprint arXiv:1905.01392*.

Wolpert, D. H. and W. G. Macready (1997). "No free lunch theorems for optimization". *IEEE Transactions on Evolutionary Computation*. 1(1): 67–82. DOI: 10.1109/4235.585893.

Wu, B., X. Dai, P. Zhang, Y. Wang, F. Sun, Y. Wu, Y. Tian, P. Vajda, Y. Jia, and K. Keutzer (2019). "Fbnet: Hardware-aware efficient convnet design via differentiable neural architecture search". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 10734–10742.

Wu, B., Y. Wang, P. Zhang, Y. Tian, P. Vajda, and K. Keutzer (2018). "Mixed precision quantization of convnets via differentiable neural architecture search". *arXiv preprint arXiv:1812.00090*.

Xian, Y., C. H. Lampert, B. Schiele, and Z. Akata (2018). "Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. 41(9): 2251–2265. DOI: 10.1109/TPAMI.2018.2857768.

Xie, S., A. Kirillov, R. Girshick, and K. He (2019a). "Exploring randomly wired neural networks for image recognition". In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 1284–1293.

Xie, S., H. Zheng, C. Liu, and L. Lin (2019b). "SNAS: Stochastic neural architecture search". In: *International Conference on Learning Representations (ICLR)*.

Xiong, H., L. Huang, M. Yu, L. Liu, F. Zhu, and L. Shao (2020). "On the number of linear regions of convolutional neural networks". In: *The International Conference on Machine Learning (ICML)*. 10514–10523.

Xiong, Y., R. Mehta, and V. Singh (2019). "Resource constrained neural network architecture search: Will a submodularity assumption help?" In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 1901–1910. DOI: 10.1109/ICCV.2019.00199.

Xu, Z., H. van Hasselt, M. Hessel, J. Oh, S. Singh, and D. Silver (2020b). "Meta-gradient reinforcement learning with an objective discovered online". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Xu, Z., H. van Hasselt, and D. Silver (2018). "Meta-gradient reinforcement learning". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Xu, Y., L. Xie, X. Zhang, X. Chen, G.-J. Qi, Q. Tian, and H. Xiong (2020a). "PC-DARTS: Partial channel connections for memory-efficient architecture search". In: *International Conference on Learning Representations (ICLR)*.

Yan, S., C. White, Y. Savani, and F. Hutter (2021). "NAS-bench-x11 and the power of learning curves". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Yang, T.-J., Y.-H. Chen, and V. Sze (2017). "Designing energy-efficient convolutional neural networks using energy-aware pruning". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 5687–5695. DOI: 10.1109/CVPR.2017.643.

Yang, M., S. Ji, W. Xu, J. Wang, F. Lv, K. Yu, Y. Gong, M. Dikmen, D. J. Lin, and T. S. Huang (2009). "Detecting human actions in surveillance videos". In: *TREC Video Retrieval Evaluation Workshop*. Citeseer.

Yang, T.-J., A. Howard, B. Chen, X. Zhang, A. Go, M. Sandler, V. Sze, and H. Adam (2018). "NetAdapt: Platform-aware neural network adaptation for mobile applications". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 285–300. DOI: 10.1007/978-3-030-01249-6_18.

Yang, X., W. Liu, D. Zhou, J. Bian, and T.-Y. Liu (2020b). "Qlib: An AI-oriented quantitative investment platform". *arXiv preprint arXiv:2009.11189*.

Yang, L. and A. Shami (2020). "On hyperparameter optimization of machine learning algorithms: Theory and practice". *Neurocomputing*. 415: 295–316. DOI: 10.1016/j.neucom.2020.07.061.

Yang, Z., Y. Wang, K. Han, C. Xu, C. Xu, D. Tao, and C. Xu (2020c). "Searching for low-bit weights in quantized neural networks". In: *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*.

Yang, L., Z. Yan, M. Li, H. Kwon, L. Lai, T. Krishna, V. Chandra, W. Jiang, and Y. Shi (2020a). "Co-exploration of neural architectures and heterogeneous asic accelerator designs targeting multiple tasks". In: *The ACM/ESDA/IEEE Design Automation Conference (DAC)*. 1–6. DOI: 10.1109/DAC18072.2020.9218676.

Ying, C. (2019). "Enumerating unique computational graphs via an iterative graph invariant". *arXiv preprint arXiv:1902.06192*.

Ying, C., A. Klein, E. Christiansen, E. Real, K. Murphy, and F. Hutter (2019). "NAS-Bench-101: Towards reproducible neural architecture search". In: *The International Conference on Machine Learning (ICML)*. 7105–7114.

Yu, J., P. Jin, H. Liu, G. Bender, P.-J. Kindermans, M. Tan, T. Huang, X. Song, R. Pang, and Q. Le (2020a). "BigNAS: Scaling up neural architecture search with big single-stage models". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 702–717. DOI: 10.1007/978-3-030-58571-6_41.

Yu, K., C. Sciuto, M. Jaggi, C. Musat, and M. Salzmann (2020b). "Evaluating the search phase of neural architecture search". In: *International Conference on Learning Representations (ICLR)*.

Yu, J., L. Yang, N. Xu, J. Yang, and T. Huang (2019). "Slimmable neural networks". In: *International Conference on Learning Representations (ICLR)*.

Yu, T. and H. Zhu (2020). "Hyper-parameter optimization: A review of algorithms and applications". *arXiv preprint arXiv:2003.05689*.

Zaidi, S., A. Zela, T. Elsken, C. Holmes, F. Hutter, and Y. W. Teh (2020). "Neural ensemble search for performant and calibrated predictions". In: *The International Conference on Machine Learning (ICML) Workshop*.

Zela, A., T. Elsken, T. Saikia, Y. Marrakchi, T. Brox, and F. Hutter (2020a). "Understanding and robustifying differentiable architecture search". In: *International Conference on Learning Representations (ICLR)*.

Zela, A., A. Klein, S. Falkner, and F. Hutter (2018). "Towards automated deep learning: Efficient joint neural architecture and hyperparameter search". In: *The International Conference on Machine Learning (ICML) Workshop*.

Zela, A., J. Siems, and F. Hutter (2020b). "NAS-Bench-1Shot1: Benchmarking and dissecting one-shot neural architecture search". In: *International Conference on Learning Representations (ICLR)*.

Zhang, W. and T. G. Dietterich (1995). "A reinforcement learning approach to job-shop scheduling". In: *International Joint Conferences on Artificial Intelligence (IJCAI)*. Vol. 95. 1114–1120.

Zhang, Q., Z. Han, F. Yang, Y. Zhang, Z. Liu, M. Yang, and L. Zhou (2020). "Retiarii: A deep learning exploratory-training framework". In: *The USENIX Symposium on Operating Systems Design and Implementation (OSDI)*. 919–936.

Zhong, Z., J. Yan, W. Wu, J. Shao, and C.-L. Liu (2018). "Practical block-wise neural network architecture generation". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 2423–2432. DOI: 10.1109/CVPR.2018.00257.

Zhou, Y., X. Dong, B. Akin, M. Tan, D. Peng, T. Meng, A. Yazdanbakhsh, D. Huang, R. Narayanaswami, and J. Laudon (2022). "Rethinking co-design of neural architectures and hardware accelerators". In: *The Conference on Machine Learning and Systems (MLSys)*.

Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros (2017). "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *Proceedings of the IEEE International Conference Computer Vision (ICCV)*. 2223–2232.

Zhou, Z.-H., Y.-Y. Sun, and Y.-F. Li (2009). "Multi-instance learning by treating instances as non-iid samples". In: *The International Conference on Machine Learning (ICML)*. 1249–1256. DOI: 10.1145/1553374.1553534.

Zhou, H., M. Yang, J. Wang, and W. Pan (2019). "BayesNAS: A Bayesian approach for neural architecture search". In: *The International Conference on Machine Learning (ICML)*. 7603–7613.

Zimmer, L., M. Lindauer, and F. Hutter (2021). "Auto-PyTorch tabular: Multi-fidelity metalearning for efficient and robust AutoDL". *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*. DOI: 10.1109/TPAMI.2021.3067763.

Zliobaite, I., A. Bifet, M. Gaber, B. Gabrys, J. Gama, L. Minku, and K. Musial (2012). "Next challenges for adaptive learning systems". *ACM SIGKDD Explorations Newsletter*. 14(1): 48–55. DOI: 10.1145/2408736.2408746.

Zliobaite, I. and B. Gabrys (2014). "Adaptive preprocessing for streaming data". *IEEE Transactions on Knowledge and Data Engineering (TKDE)*. 26(2): 309–321. DOI: 10.1109/TKDE.2012.147.

Zoph, B., E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le (2020). "Learning data augmentation strategies for object detection". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer. 566–583. DOI: 10.1007/978-3-030-58583-9_34.

Zoph, B. and Q. V. Le (2017). "Neural architecture search with reinforcement learning". In: *International Conference on Learning Representations (ICLR)*.

Zoph, B., V. Vasudevan, J. Shlens, and Q. V. Le (2018). "Learning transferable architectures for scalable image recognition". In: *Proceedings of the IEEE Conference Computer Vision Pattern Recognition (CVPR)*. 8697–8710.