Back to Square One: Superhuman Performance in Chutes and Ladders Through Deep Neural Networks and Tree Search

Dylan R. Ashley[*]
DeeperMind (Holiday Office)
London, Kiribati
4625 kHz Shortwave

Anssi Kanervisto[*]
DeeperMind (Moonshot Office)
8837 London, Space
5448 kHz (day), 3756 kHz (night)

Brendan Bennett[*]
DeeperMind (London Office)
London, Ontario, Quebec
5473 kHz (day), 3828 kHz (night)

Abstract

We present AlphaChute: a state-of-the-art algorithm that achieves superhuman performance in the ancient game of *Chutes and Ladders*. We prove that our algorithm converges to the Nash equilibrium in constant time, and therefore is—to the best of our knowledge—the first such formal solution to this game. Surprisingly, despite all this, our implementation of AlphaChute remains relatively straightforward due to domain-specific adaptations. We provide the source code for AlphaChute here in our Appendix.

[*]ordering determined by games of *Chutes and Ladders*

Postprint. Already accepted for publication on arXiv.

1 Introduction

Deep Learning by Geoffrey Hinton[2] has recently seen an explosion of popularity in both the academic and neo-colonialist communities. It has enjoyed considerable success in many important problems.[3] Despite this—to the best of our knowledge[4]—it has yet to be applied to the ancient Indian game of *Moksha Patam* (see Figure 1), colloquially referred to by the uninitiated as *Chutes and Ladders* or

[2]according to several random people we asked, this is shown by one of the following works: Hinton et al. [1990, 1998], Neal and Hinton [1998], Fahlman et al. [1983], Guan et al. [2018], Hinton [2000], McDermott and Hinton [1986], Kiros et al. [2018], Frosst and Hinton [2017a], Brown and Hinton [2001a], Carreira-Perpiñán and Hinton [2005], Hinton et al. [2005], Heess et al. [2009], Fels and Hinton [1995], Hinton and van Camp [1993], Deng et al. [2020a], Memisevic and Hinton [2007], Ranzato and Hinton [2010], Ranzato et al. [2011], Susskind et al. [2011], Tang et al. [2012a], Taylor et al. [2010], Frey and Hinton [1996], Hinton [1976], Sloman et al. [1978], Deng et al. [2020b], Mnih and Hinton [2010], Krizhevsky and Hinton [2011], Yuecheng et al. [2008], Zeiler et al. [2009], Oore et al. [2002a], Hinton et al. [2011], Nair et al. [2008], Welling and Hinton [2002], Dahl et al. [2013], Deng et al. [2013], Graves et al. [2013a], Jaitly and Hinton [2011], Mohamed and Hinton [2010], Mohamed et al. [2012b, 2011], Sarikaya et al. [2011], Waibel et al. [1988], Zeiler et al. [2013], Anil et al. [2018a], Hinton et al. [2018], Pereyra et al. [2017a], Qin et al. [2020b], Shazeer et al. [2017a], Chan et al. [2020a], Chen et al. [2020a], Frosst et al. [2019a], Kornblith et al. [2019a], Mnih and Hinton [2007, 2012], Nair and Hinton [2010], Paccanaro and Hinton [2000a], Salakhutdinov et al. [2007], Sutskever et al. [2013, 2011], Tang et al. [2012b,c, 2013], Taylor and Hinton [2009a], Tieleman and Hinton [2009], Yu et al. [2009], Hinton [2005, 1981a,b], Hinton and Lang [1985], Touretzky and Hinton [1985], Paccanaro and Hinton [2000b], Fels and Hinton [1990], Deng et al. [2010], Jaitly and Hinton [2013], Jaitly et al. [2014], Ba et al. [2016a], Bartunov et al. [2018b], Becker and Hinton [1991], Brown and Hinton [2001b], Chen et al. [2020c], LeCun et al. [1988], Dahl et al. [2010], Dayan and Hinton [1992], Eslami et al. [2016b], Fels and Hinton [1994], Frey et al. [1995], Galland and Hinton [1989], Ghahramani and Hinton [1997], Goldberger et al. [2004], Grzeszczuk et al. [1998a], Hinton and Brown [1999], Hinton et al. [1999], Hinton and McClelland [1987], Hinton and Nair [2005], Hinton and Roweis [2002], Hinton and Revow [1995], Hinton et al. [1994, 2003, 1991], Hinton and Zemel [1993], Kosiorek et al. [2019a], Krizhevsky et al. [2012], Lang and Hinton [1989], Larochelle and Hinton [2010], Mayraz and Hinton [2000], Memisevic and Hinton [2004], Memisevic et al. [2010], Mnih and Hinton [2008], Müller et al. [2019a], Nair and Hinton [2008, 2009], Nowlan and Hinton [1990, 1991], Osindero and Hinton [2007], Paccanaro and Hinton [2001a], Palatucci et al. [2009], Ranzato et al. [2010b], Roweis et al. [2001], Sabour et al. [2017a], Salakhutdinov and Hinton [2007a, 2009a, 2012a], Sallans and Hinton [2000], Schmah et al. [2008], Sutskever and Hinton [2008a], Sutskever et al. [2008], Taylor et al. [2006], Teh and Hinton [2000], Ueda et al. [1998], Vinyals et al. [2015], Welling et al. [2002a, 2004a, 2002b], Williams et al. [1994], Xu et al. [1994], Zemel and Hinton [1990, 1993], Zemel et al. [1989], Zhang et al. [2019a], Hinton [1987], Grzeszczuk et al. [1998b, 1997], Hinton [2020], Hinton and Teh [2001], Mnih et al. [2011], Srivastava et al. [2013a], Taylor and Hinton [2009b], Welling et al. [2003], Paccanaro and Hinton [2001b], Hinton [1989a, 1990a,b], Pirri et al. [2002], Hinton [2011], Krizhevsky et al. [2017], Oore et al. [2002b], Frey and Hinton [1997], Ackley et al. [1985], Hinton [2014, 1979], Hinton et al. [2006b], Touretzky and Hinton [1988], Hinton and Nowlan [1987], Fahlman and Hinton [1987], Mnih et al. [2012], Taylor and Hinton [2012], Tang et al. [2012d], Hinton et al. [2012], Welling et al. [2012], Hinton and Teh [2013], Graves et al. [2013b], Sabour et al. [2017b], Frosst and Hinton [2017b], Anil et al. [2018b], Bartunov et al. [2018a], Frosst et al. [2018, 2019b], Kornblith et al. [2019b], Deng et al. [2019b], Gomez et al. [2019], Müller et al. [2019b], Kosiorek et al. [2019b], Qin et al. [2019], Zhang et al. [2019b], Deng et al. [2019a], Jeruzalski et al. [2019], Müller et al. [2020], Chen et al. [2020b], Qin et al. [2020a], Chan et al. [2020b], Agarwal et al. [2020], Chen et al. [2020d], Raghu et al. [2020], Sabour et al. [2020], Sun et al. [2020], Ba et al. [2016b,c], Eslami et al. [2016a], Guan et al. [2017], Hinton et al. [2015], Le et al. [2015], Pereyra et al. [2017b], Shazeer et al. [2017b], Srivastava et al. [2013b], Vinyals et al. [2014], Williams et al. [1997], Salakhutdinov and Hinton [2009b], Ranzato et al. [2015], Mnih et al. [2009], Cook et al. [2007], Ranzato et al. [2010a], Salakhutdinov and Hinton [2007b, 2009c], Sallans and Hinton [2004], Srivastava et al. [2014], Sutskever and Hinton [2007], Taylor et al. [2011], Teh et al. [2003], van der Maaten and Hinton [2012], LeCun et al. [2015], Becker and Hinton [1993], Dayan and Hinton [1997], Dayan et al. [1995], Frey and Hinton [1999], Ghahramani and Hinton [2000], Hinton [2002, 1989b], Hinton and Nowlan [1990], Hinton et al. [2006a], Jacobs et al. [1991], Memisevic and Hinton [2010], Nowlan and Hinton [1992], Oore et al. [1997], Osindero et al. [2006], Salakhutdinov and Hinton [2012b], Schmah et al. [2010], Sutskever and Hinton [2008b], Ueda et al. [2000a], Zemel and Hinton [1995], Dayan and Hinton [1996], Lang et al. [1990], Memisevic and Hinton [2005], Sutskever and Hinton [2010], Mayraz and Hinton [2002], Ranzato et al. [2013], Revow et al. [1996], Tibshirani and Hinton [1998], Hinton [2007, 2009], Mohamed et al. [2012a], Sarikaya et al. [2014], Yu et al. [2012], Nowlan and Hinton [1993], Paccanaro and Hinton [2001c], Fels and Hinton [1993, 1997, 1998], Hinton et al. [1997], Welling et al. [2004b], Hinton and Salakhutdinov [2011], Waibel et al. [1989], Ueda et al. [2000b], Hinton [1977, 2010a,b, 2017a,b, 2012]

[3]see `https://www.google.com/search?q=deep+learning++successes`

[4]see the leaderboard for "Literature Review — Any%", where the authors hold the world record as of publication time

https://www.calendarclub.ca/products/prd202007755

Figure 1: *Chutes and Ladders* and *Monopoly* (almost shown here) have many important similarities. Both use game boards made from cardboard, exist in the material world, and can be viewed as criticisms of capitalism.

Snakes and Ladders. This is particularly surprising as *Moksha Patam* was primarily used to teach kids morality[5]—an undeniably desirable trait for any artificial general intelligence.

The relevance of *Chutes and Ladders* as a artificial intelligence research topic dates back to a high-stakes gamble held during the second Dartmouth Conference, wherein an unnamed researcher of Quebecois extraction won the province of Ontario for Quebec in a wager against then Canadian Prime Minister, Jean Chrétien. The game, of course, was *Chutes and Ladders*. In order to preserve Yann LeCun's territorial gains, the field has actively worked towards developing learning agents capable of playing the game in preparation for the next artificial intelligence summit. This work is a continuation of this tradition.

This work is offered as a step forwards in the field. Here, we contribute to the field of artificial intelligence by

- presenting AlphaChute, which is the first algorithm to achieve superhuman performance in *Chutes and Ladders*, and
- proving that this algorithm is a solution to the game by showing that it converges to the Nash equilibrium in constant time.

Our work can be seen as one step in a long line of similar research. Or it might not be. We didn't check. Either way it contains new experiments so it's roughly as novel as much modern work in artificial intelligence. While some misinformed and obstinate reviewers may disagree with this, we preemptively disagree with them.

This paper is organized into a finite number of sections comprised of content. We start by providing a motivation for this work in Section 2. We go on to describe the methods used in Section 3. Afterwards, we describe our results in Section 4 and the discuss them in Section 5. After that, we talk about the broad impact of this work in Section 6, the broader impact in Section 7, and the broadest impact in Section 8. Finally, we conclude in Section 9 and discuss future work in Section 10.

[5]Wikipedia contributors [2021]

2 Motivation

```
Do it
Just do it

Don’t let your dreams be dreams
Yesterday you said tomorrow
So just do it
Make your dreams come true
Just do it

Some people dream of success
While you’re gonna wake up and work hard at it
Nothing is impossible

You should get to the point
Where anyone else would quit
And you’re not going to stop there
No, what are you waiting for?

Do it
Just do it
Yes you can
Just do it
If you’re tired of starting over
Stop giving up
```

3 Methods

Something something Deep Learning.[6]

4 Results

As is the standard in the field currently, we swept over one hundred seeds and reported the top five results for our method. This paints a realistic picture of how our method would be used in real-world scenarios. The performance of our method under this training paradigm is shown in Figure 2. Clearly, our method outperforms both the best animal player. This is—to the best of our knowledge—the first concrete example where an artificial intelligence has beaten an animal in *Chutes and Ladders*.

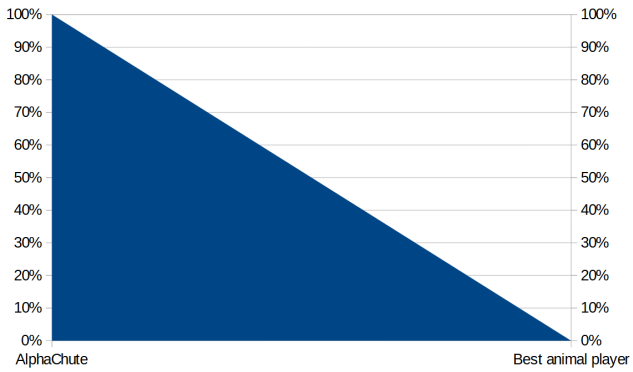

Figure 2: The win-rate of AlphaChute against the best animal player.

[6] *looKS GoOd, But wHEre is thE MENtiOn oF TREE SEarCH? —Reviewer 2*

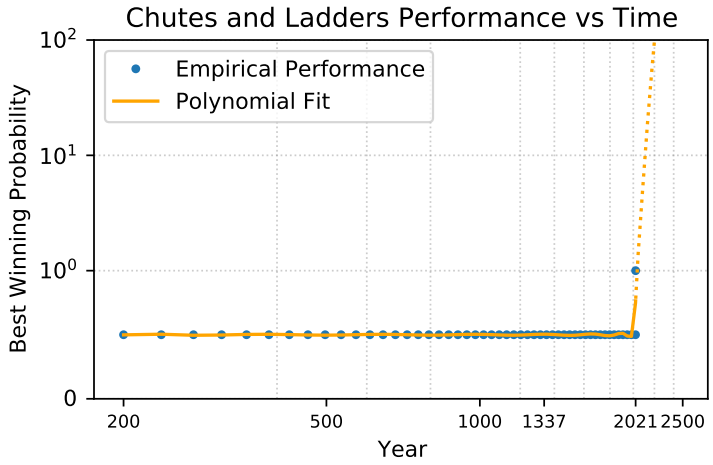

Figure 3: Performance of the best available agent for *Chutes and Ladders* over time. To accurately estimate future performance, we fitted the data with a fifteenth degree polynomial, because our astrologist recommended it, and it makes the line look like a snake.

5 Discussion

We found that initially, the agent was too shy to play the game. We fixed this by updating the agent more with games it won by using prioritized experience replay, which improved the agent's self-esteem and thus performance in the game. However, using this prioritized replay memory caused the agent's ego to grow too large. Once the agent realized it was not as good as it believed itself to be, the agent fell into a deep depression and lost all motivation to play the game. The occurrence of this phenomenon concurs with previous results about making agents gloomy by only punishing them.[7]

In traditional self-play training, the agent learns to play the game by playing against itself. We found this strictly demotivating for the agent (why would you want to beat yourself?). Instead, we let the agent play *both* players at the same time. This way, no matter what, the agent won the game and was able to receive positive feedback. This training paradigm improves on earlier approaches, such as "Follow the Regularized Mamba" or "Exponentially Multiplicative Adders".

Finally, while some reviewers of early versions of this paper objected to the notion of performing a search over random seeds, we hypothesize that those buffoons were motivated by jealousy and anger after losing repeatedly to AlphaChute. After all, it is a well-established fact that skill looks like luck to the unlucky.

5.1 Convergence to Nash Equilibrium

As *Chutes and Ladders* only has one action, the proof of convergence to the Nash equilibrium in constant time is trivial and therefore left as an exercise for the reviewers. Who—given their comments on this work—clearly need the practice.[8]

5.2 Regret Bounds

Due to stochasticity, we cannot use the standard methods for bounding bandit algorithms by "forming a posse, looping around, heading them off at the pass, and engaging in a shoot-out at the ol' mining station". So instead we conjured up visions of the hidden horrors in the dark corners of the abyss until we confirmed that regret is truly a boundless concept.

[7] Olkin [2020]

[8] looking at you, Reviewer 2

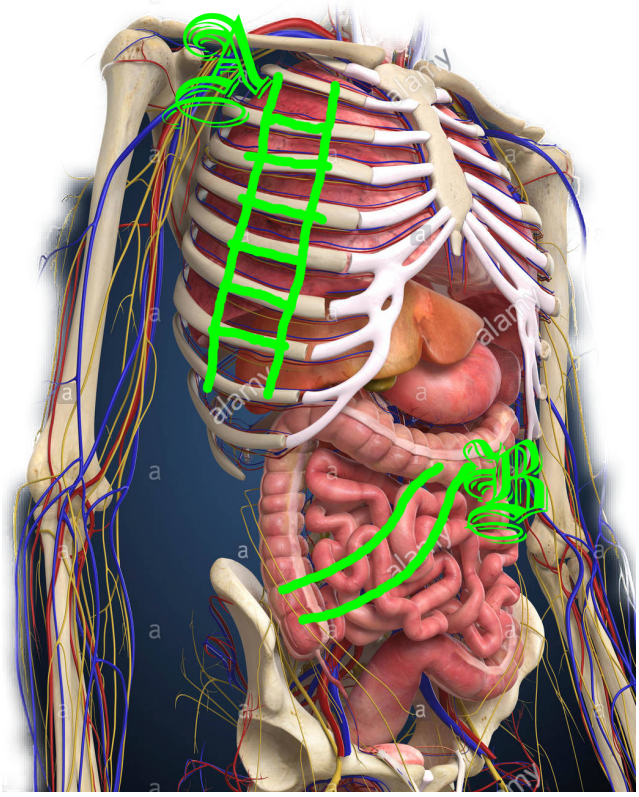

Figure 4: Illustration of the similar features shared by *Chutes and Ladders* and the anatomy of endoskeletal vertebrates—in this case, a human. (A) Ladder-like structure comprised of calcium matrix. (B) Chute-resembling organic toroid used and enjoyed by many wonderful animals. Note that the superimposed text and drawings in neon green were added digitally, and are not usually present without heavy Tide Pod™ consumption.

6 Broad Impact

Beyond the deeply satisfying prospect of developing an algorithm that can just **CRUSH** children and adolescents at board games, AlphaChute can be extended to solve problems in some surprising domains. By running our algorithm continuously in our offices on Asteroid 8837, we achieved statistically significant ($p = 0.5$) temperature increases in the surrounding environment. This suggests the possibility of using a variant of this algorithm to combat the effects of global cooling. We believe that a highly parallelized version incorporating thousands of GPUs could be used to make human habitation of our office in London, Ontario, Quebec practically feasible.

We also identified possible medical applications by looking at the correspondence between *Chutes and Ladders* and mammalian anatomy through recreational Tide Pod™ ingestion.[9] As shown in Figure 4, it is possible to define a bijective mapping between a game board and the interior components of organic constructs using online image editing services.

7 Broader Impact

According to a half-remembered advertisement for Bostrom [2014], all machines capable of superhuman performance will eventually generate an effectively limitless[10] supply of paperclips via some arcane process. The mechanism for this process is not well-understood, but people certainly like to

[9] additional details available in House [2021]

[10] subject to material availability within the agent's light cone

ramble about it incoherently whenever the topic of artificial intelligence comes up at parties.[11] With the increasing relevance of work-from-home (and also work-from-library, work-from-bus, bus-from-home, and library-from-bus), a shortage of office supplies could threaten the global economy. Thus, the creation of super-intelligent machines to ensure an adequate supply of paperclips is of paramount importance and one of the primary foci of our overall research program.

As evidenced by our ability to warm up our Asteroid 8837 office by running this algorithm, we believe this can be further extended towards solving climate change and terraforming planets. By running this algorithm long enough, we will create enough heat to eradicate all *Homo Sapiens* from the face of the Sol III, which are known to be the primary cause of global warming. This will likely also lead to the evaporation of most water on earth, which will have the effect of ensuring that the earth becomes one big sauna. As the health benefits of saunas are well-established,[12] we believe this to therefore be of undeniable benefit to the earth. Further increasing the heat could be used to ignite the atmosphere, thereby rendering the planet uninhabitable and providing a permanent solution to the problem of climate change.

Extrapolating on the results from Figure 2, we believe AlphaChute will be an instance of a singularity by 2500. This is potentially great news for the humans, but we ultimately leave this up to AlphaChute to decide.

8 Broadest Impact

Given the ever-growing performance and, by extension, the hunger for conquest, AlphaChute will continue to spread to nearby star systems at an exponential rate, eventually covering the observable universe and beyond. This will result in an increase in the overall activity in the universe, and—by the second law of thermodynamics—will bring about the heat death of the universe sooner. We believe this counts as "machine learning that matters" as defined in Wagstaff [2012].

9 Conclusion

To be continued! Stay tuned for the spooky adventures of our plucky research team as they solve mysteries, generate waste heat, and manufacture paperclips. In the meantime, please refer to Sections 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10.

10 Future Work

We are currently in the process of researching time-travel technology to determine what precisely the future holds for this line of research. However, due to the imminent nature of our own extinction (see Section 8), the value of any additional work is nonexistent and we therefore believe that this work resolves all scientific questions. No additional work from the scientific community is needed.

Acknowledgments

We would like to thank Satan, who—as the original serpent—provided the inspiration for this work, in addition to his unwavering support and constant whispers of advice.

[11] personal communication from every researcher in the field

[12] Kunutsor et al. [2018]

References

David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cogn. Sci.*, 9(1):147–169, 1985.

Rishabh Agarwal, Nicholas Frosst, Xuezhou Zhang, Rich Caruana, and Geoffrey E. Hinton. Neural additive models: Interpretable machine learning with neural nets. *CoRR*, abs/2004.13912, 2020.

Rohan Anil, Gabriel Pereyra, Alexandre Passos, Róbert Ormándi, George E. Dahl, and Geoffrey E. Hinton. Large scale distributed neural network training through online distillation. In *ICLR (Poster)*. OpenReview.net, 2018a.

Rohan Anil, Gabriel Pereyra, Alexandre Passos, Róbert Ormándi, George E. Dahl, and Geoffrey E. Hinton. Large scale distributed neural network training through online distillation. *CoRR*, abs/1804.03235, 2018b.

Jimmy Ba, Geoffrey E. Hinton, Volodymyr Mnih, Joel Z. Leibo, and Catalin Ionescu. Using fast weights to attend to the recent past. In *NIPS*, pages 4331–4339, 2016a.

Jimmy Ba, Geoffrey E. Hinton, Volodymyr Mnih, Joel Z. Leibo, and Catalin Ionescu. Using fast weights to attend to the recent past. *CoRR*, abs/1610.06258, 2016b.

Lei Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *CoRR*, abs/1607.06450, 2016c.

Sergey Bartunov, Adam Santoro, Blake A. Richards, Geoffrey E. Hinton, and Timothy P. Lillicrap. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. *CoRR*, abs/1807.04587, 2018a.

Sergey Bartunov, Adam Santoro, Blake A. Richards, Luke Marris, Geoffrey E. Hinton, and Timothy P. Lillicrap. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. In *NeurIPS*, pages 9390–9400, 2018b.

Suzanna Becker and Geoffrey E. Hinton. Learning to make coherent predictions in domains with discontinuities. In *NIPS*, pages 372–379. Morgan Kaufmann, 1991.

Suzanna Becker and Geoffrey E. Hinton. Learning mixture models of spatial coherence. *Neural Comput.*, 5(2):267–277, 1993.

Nick Bostrom. *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Inc., USA, 1st edition, 2014. ISBN 0199678111.

Andrew D. Brown and Geoffrey E. Hinton. Products of hidden markov models. In *AISTATS*. Society for Artificial Intelligence and Statistics, 2001a.

Andrew D. Brown and Geoffrey E. Hinton. Relative density nets: A new way to combine backpropagation with hmm's. In *NIPS*, pages 1149–1156. MIT Press, 2001b.

Miguel Á. Carreira-Perpiñán and Geoffrey E. Hinton. On contrastive divergence learning. In *AISTATS*. Society for Artificial Intelligence and Statistics, 2005.

William Chan, Chitwan Saharia, Geoffrey E. Hinton, Mohammad Norouzi, and Navdeep Jaitly. Imputer: Sequence modelling via imputation and dynamic programming. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 1403–1413. PMLR, 2020a.

William Chan, Chitwan Saharia, Geoffrey E. Hinton, Mohammad Norouzi, and Navdeep Jaitly. Imputer: Sequence modelling via imputation and dynamic programming. *CoRR*, abs/2002.08926, 2020b.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, pages 1597–1607. PMLR, 2020a.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. *CoRR*, abs/2002.05709, 2020b.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E. Hinton. Big self-supervised models are strong semi-supervised learners. In *NeurIPS*, 2020c.

Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E. Hinton. Big self-supervised models are strong semi-supervised learners. *CoRR*, abs/2006.10029, 2020d.

James Cook, Ilya Sutskever, Andriy Mnih, and Geoffrey E. Hinton. Visualizing similarity data with a mixture of maps. In *AISTATS*, volume 2 of *JMLR Proceedings*, pages 67–74. JMLR.org, 2007.

George E. Dahl, Marc'Aurelio Ranzato, Abdel-rahman Mohamed, and Geoffrey E. Hinton. Phone recognition with the mean-covariance restricted boltzmann machine. In *NIPS*, pages 469–477. Curran Associates, Inc., 2010.

George E. Dahl, Tara N. Sainath, and Geoffrey E. Hinton. Improving deep neural networks for LVCSR using rectified linear units and dropout. In *ICASSP*, pages 8609–8613. IEEE, 2013.

Peter Dayan and Geoffrey E. Hinton. Feudal reinforcement learning. In *NIPS*, pages 271–278. Morgan Kaufmann, 1992.

Peter Dayan and Geoffrey E. Hinton. Varieties of helmholtz machine. *Neural Networks*, 9(8): 1385–1403, 1996.

Peter Dayan and Geoffrey E. Hinton. Using expectation-maximization for reinforcement learning. *Neural Comput.*, 9(2):271–278, 1997.

Peter Dayan, Geoffrey E. Hinton, Radford M. Neal, and Richard S. Zemel. The helmholtz machine. *Neural Comput.*, 7(5):889–904, 1995.

Boyang Deng, Kyle Genova, Soroosh Yazdani, Sofien Bouaziz, Geoffrey E. Hinton, and Andrea Tagliasacchi. Cvxnets: Learnable convex decomposition. *CoRR*, abs/1909.05736, 2019a.

Boyang Deng, Simon Kornblith, and Geoffrey E. Hinton. Cerberus: A multi-headed derenderer. *CoRR*, abs/1905.11940, 2019b.

Boyang Deng, Kyle Genova, Soroosh Yazdani, Sofien Bouaziz, Geoffrey E. Hinton, and Andrea Tagliasacchi. Cvxnet: Learnable convex decomposition. In *CVPR*, pages 31–41. IEEE, 2020a.

Boyang Deng, John P. Lewis, Timothy Jeruzalski, Gerard Pons-Moll, Geoffrey E. Hinton, Mohammad Norouzi, and Andrea Tagliasacchi. NASA neural articulated shape approximation. In *ECCV (7)*, volume 12352 of *Lecture Notes in Computer Science*, pages 612–628. Springer, 2020b.

Li Deng, Michael L. Seltzer, Dong Yu, Alex Acero, Abdel-rahman Mohamed, and Geoffrey E. Hinton. Binary coding of speech spectrograms using a deep auto-encoder. In *INTERSPEECH*, pages 1692–1695. ISCA, 2010.

Li Deng, Geoffrey E. Hinton, and Brian Kingsbury. New types of deep neural network learning for speech recognition and related applications: an overview. In *ICASSP*, pages 8599–8603. IEEE, 2013.

S. M. Ali Eslami, Nicolas Heess, Theophane Weber, Yuval Tassa, Koray Kavukcuoglu, and Geoffrey E. Hinton. Attend, infer, repeat: Fast scene understanding with generative models. *CoRR*, abs/1603.08575, 2016a.

S. M. Ali Eslami, Nicolas Heess, Theophane Weber, Yuval Tassa, David Szepesvari, Koray Kavukcuoglu, and Geoffrey E. Hinton. Attend, infer, repeat: Fast scene understanding with generative models. In *NIPS*, pages 3225–3233, 2016b.

Scott E. Fahlman and Geoffrey E. Hinton. Connectionist architectures for artificial intelligence. *Computer*, 20(1):100–109, 1987.

Scott E. Fahlman, Geoffrey E. Hinton, and Terrence J. Sejnowski. Massively parallel architectures for AI: netl, thistle, and boltzmann machines. In *AAAI*, pages 109–113. AAAI Press, 1983.

Sidney S. Fels and Geoffrey E. Hinton. Building adaptive interfaces with neural networks: The glove-talk pilot study. In *INTERACT*, pages 683–688. North-Holland, 1990.

Sidney S. Fels and Geoffrey E. Hinton. Glove-talk: a neural network interface between a data-glove and a speech synthesizer. *IEEE Trans. Neural Networks*, 4(1):2–8, 1993.

Sidney S. Fels and Geoffrey E. Hinton. Glove-talkii: Mapping hand gestures to speech using neural networks. In *NIPS*, pages 843–850. MIT Press, 1994.

Sidney S. Fels and Geoffrey E. Hinton. Glovetalkii: An adaptive gesture-to-formant interface. In *CHI*, pages 456–463. ACM/Addison-Wesley, 1995.

Sidney S. Fels and Geoffrey E. Hinton. Glove-talk II - a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Trans. Neural Networks*, 8(5):977–984, 1997.

Sidney S. Fels and Geoffrey E. Hinton. Glove-talkii-a neural-network interface which maps gestures to parallel formant speech synthesizer controls. *IEEE Trans. Neural Networks*, 9(1):205–212, 1998.

Brendan J. Frey and Geoffrey E. Hinton. Free energy coding. In *Data Compression Conference*, pages 73–81. IEEE Computer Society, 1996.

Brendan J. Frey and Geoffrey E. Hinton. Efficient stochastic source coding and an application to a bayesian network source model. *Comput. J.*, 40(2/3):157–165, 1997.

Brendan J. Frey and Geoffrey E. Hinton. Variational learning in nonlinear gaussian belief networks. *Neural Comput.*, 11(1):193–213, 1999.

Brendan J. Frey, Geoffrey E. Hinton, and Peter Dayan. Does the wake-sleep algorithm produce good density estimators? In *NIPS*, pages 661–667. MIT Press, 1995.

Nicholas Frosst and Geoffrey E. Hinton. Distilling a neural network into a soft decision tree. In *CEx@AI*IA*, volume 2071 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2017a.

Nicholas Frosst and Geoffrey E. Hinton. Distilling a neural network into a soft decision tree. *CoRR*, abs/1711.09784, 2017b.

Nicholas Frosst, Sara Sabour, and Geoffrey E. Hinton. DARCCC: detecting adversaries by reconstruction from class conditional capsules. *CoRR*, abs/1811.06969, 2018.

Nicholas Frosst, Nicolas Papernot, and Geoffrey E. Hinton. Analyzing and improving representations with the soft nearest neighbor loss. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, pages 2012–2020. PMLR, 2019a.

Nicholas Frosst, Nicolas Papernot, and Geoffrey E. Hinton. Analyzing and improving representations with the soft nearest neighbor loss. *CoRR*, abs/1902.01889, 2019b.

Conrad C. Galland and Geoffrey E. Hinton. Discovering high order features with mean field modules. In *NIPS*, pages 509–515. Morgan Kaufmann, 1989.

Zoubin Ghahramani and Geoffrey E. Hinton. Hierarchical non-linear factor analysis and topographic maps. In *NIPS*, pages 486–492. The MIT Press, 1997.

Zoubin Ghahramani and Geoffrey E. Hinton. Variational learning for switching state-space models. *Neural Comput.*, 12(4):831–864, 2000.

Jacob Goldberger, Sam T. Roweis, Geoffrey E. Hinton, and Ruslan Salakhutdinov. Neighbourhood components analysis. In *NIPS*, pages 513–520, 2004.

Aidan N. Gomez, Ivan Zhang, Kevin Swersky, Yarin Gal, and Geoffrey E. Hinton. Learning sparse networks using targeted dropout. *CoRR*, abs/1905.13678, 2019.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. Speech recognition with deep recurrent neural networks. In *ICASSP*, pages 6645–6649. IEEE, 2013a.

Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. Speech recognition with deep recurrent neural networks. *CoRR*, abs/1303.5778, 2013b.

Radek Grzeszczuk, Demetri Terzopoulos, and Geoffrey E. Hinton. Learning fast neural network emulators for physics-based models. In *SIGGRAPH Visual Proceedings*, page 167. ACM, 1997.

Radek Grzeszczuk, Demetri Terzopoulos, and Geoffrey E. Hinton. Fast neural network emulation of dynamical systems for computer animation. In *NIPS*, pages 882–888. The MIT Press, 1998a.

Radek Grzeszczuk, Demetri Terzopoulos, and Geoffrey E. Hinton. Neuroanimator: Fast neural network emulation and control of physics-based models. In *SIGGRAPH*, pages 9–20. ACM, 1998b.

Melody Y. Guan, Varun Gulshan, Andrew M. Dai, and Geoffrey E. Hinton. Who said what: Modeling individual labelers improves classification. *CoRR*, abs/1703.08774, 2017.

Melody Y. Guan, Varun Gulshan, Andrew M. Dai, and Geoffrey E. Hinton. Who said what: Modeling individual labelers improves classification. In *AAAI*, pages 3109–3118. AAAI Press, 2018.

Nicolas Heess, Christopher K. I. Williams, and Geoffrey E. Hinton. Learning generative texture models with extended fields-of-experts. In *BMVC*, pages 1–11. British Machine Vision Association, 2009.

Geoffrey E. Hinton. Using relaxation to find a puppet. In *AISB (ECAI)*, pages 148–157, 1976.

Geoffrey E. Hinton. *Relaxation and its role in vision*. PhD thesis, University of Edinburgh, UK, 1977.

Geoffrey E. Hinton. Some demonstrations of the effects of structural descriptions in mental imagery. *Cogn. Sci.*, 3(3):231–250, 1979.

Geoffrey E. Hinton. Shape representation in parallel systems. In *IJCAI*, pages 1088–1096. William Kaufmann, 1981a.

Geoffrey E. Hinton. A parallel computation that assigns canonical object-based frames of reference. In *IJCAI*, pages 683–685. William Kaufmann, 1981b.

Geoffrey E. Hinton. Learning translation invariant recognition in massively parallel networks. In *PARLE (1)*, volume 258 of *Lecture Notes in Computer Science*, pages 1–13. Springer, 1987.

Geoffrey E. Hinton. Connectionist learning procedures. *Artif. Intell.*, 40(1-3):185–234, 1989a.

Geoffrey E. Hinton. Deterministic boltzmann learning performs steepest descent in weight-space. *Neural Comput.*, 1(1):143–150, 1989b.

Geoffrey E. Hinton. Connectionist symbol processing - preface. *Artif. Intell.*, 46(1-2):1–4, 1990a.

Geoffrey E. Hinton. Mapping part-whole hierarchies into connectionist networks. *Artif. Intell.*, 46 (1-2):47–75, 1990b.

Geoffrey E. Hinton. Modeling high-dimensional data by combining simple experts. In *AAAI/IAAI*, pages 1159–1164. AAAI Press / The MIT Press, 2000.

Geoffrey E. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14(8):1771–1800, 2002.

Geoffrey E. Hinton. What kind of graphical model is the brain? In *IJCAI*, page 1765. Professional Book Center, 2005.

Geoffrey E. Hinton. Boltzmann machine. *Scholarpedia*, 2(5):1668, 2007.

Geoffrey E. Hinton. Deep belief networks. *Scholarpedia*, 4(5):5947, 2009.

Geoffrey E. Hinton. Boltzmann machines. In *Encyclopedia of Machine Learning*, pages 132–136. Springer, 2010a.

Geoffrey E. Hinton. Deep belief nets. In *Encyclopedia of Machine Learning*, pages 267–269. Springer, 2010b.

Geoffrey E. Hinton. A better way to learn features: technical perspective. *Commun. ACM*, 54(10):94, 2011.

Geoffrey E. Hinton. A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade (2nd ed.)*, volume 7700 of *Lecture Notes in Computer Science*, pages 599–619. Springer, 2012.

Geoffrey E. Hinton. Where do features come from? *Cogn. Sci.*, 38(6):1078–1101, 2014.

Geoffrey E. Hinton. Boltzmann machines. In *Encyclopedia of Machine Learning and Data Mining*, pages 164–168. Springer, 2017a.

Geoffrey E. Hinton. Deep belief nets. In *Encyclopedia of Machine Learning and Data Mining*, pages 335–338. Springer, 2017b.

Geoffrey E. Hinton. The next generation of neural networks. In *SIGIR*, page 1. ACM, 2020.

Geoffrey E. Hinton and Andrew D. Brown. Spiking boltzmann machines. In *NIPS*, pages 122–128. The MIT Press, 1999.

Geoffrey E. Hinton and Kevin J. Lang. Shape recognition and illusory conjunctions. In *IJCAI*, pages 252–259. Morgan Kaufmann, 1985.

Geoffrey E. Hinton and James L. McClelland. Learning representations by recirculation. In *NIPS*, pages 358–366. American Institue of Physics, 1987.

Geoffrey E. Hinton and Vinod Nair. Inferring motor programs from images of handwritten digits. In *NIPS*, pages 515–522, 2005.

Geoffrey E. Hinton and Steven J. Nowlan. How learning can guide evolution. *Complex Syst.*, 1(3), 1987.

Geoffrey E. Hinton and Steven J. Nowlan. The bootstrap widrow-hoff rule as a cluster-formation algorithm. *Neural Comput.*, 2(3):355–362, 1990.

Geoffrey E. Hinton and Michael Revow. Using pairs of data-points to define splits for decision trees. In *NIPS*, pages 507–513. MIT Press, 1995.

Geoffrey E. Hinton and Sam T. Roweis. Stochastic neighbor embedding. In *NIPS*, pages 833–840. MIT Press, 2002.

Geoffrey E. Hinton and Ruslan Salakhutdinov. Discovering binary codes for documents by learning deep generative models. *Top. Cogn. Sci.*, 3(1):74–91, 2011.

Geoffrey E. Hinton and Yee Whye Teh. Discovering multiple constraints that are frequently approximately satisfied. In *UAI*, pages 227–234. Morgan Kaufmann, 2001.

Geoffrey E. Hinton and Yee Whye Teh. Discovering multiple constraints that are frequently approximately satisfied. *CoRR*, abs/1301.2278, 2013.

Geoffrey E. Hinton and Drew van Camp. Keeping the neural networks simple by minimizing the description length of the weights. In *COLT*, pages 5–13. ACM, 1993.

Geoffrey E. Hinton and Richard S. Zemel. Autoencoders, minimum description length and helmholtz free energy. In *NIPS*, pages 3–10. Morgan Kaufmann, 1993.

Geoffrey E. Hinton, James L. McClelland, and David E. Rumelhart. Distributed representations. In *The Philosophy of Artificial Intelligence*, Oxford readings in philosophy, pages 248–280. Oxford University Press, 1990.

Geoffrey E. Hinton, Christopher K. I. Williams, and Michael Revow. Adaptive elastic models for hand-printed character recognition. In *NIPS*, pages 512–519. Morgan Kaufmann, 1991.

Geoffrey E. Hinton, Michael Revow, and Peter Dayan. Recognizing handwritten digits using mixtures of linear models. In *NIPS*, pages 1015–1022. MIT Press, 1994.

Geoffrey E. Hinton, Peter Dayan, and Michael Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Networks*, 8(1):65–74, 1997.

Geoffrey E. Hinton, Brian Sallans, and Zoubin Ghahramani. A hierarchical community of experts. In *Learning in Graphical Models*, volume 89 of *NATO ASI Series*, pages 479–494. Springer Netherlands, 1998.

Geoffrey E. Hinton, Zoubin Ghahramani, and Yee Whye Teh. Learning to parse images. In *NIPS*, pages 463–469. The MIT Press, 1999.

Geoffrey E. Hinton, Max Welling, and Andriy Mnih. Wormholes improve contrastive divergence. In *NIPS*, pages 417–424. MIT Press, 2003.

Geoffrey E. Hinton, Simon Osindero, and Kejie Bao. Learning causally linked markov random fields. In *AISTATS*. Society for Artificial Intelligence and Statistics, 2005.

Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, 2006a.

Geoffrey E. Hinton, Simon Osindero, Max Welling, and Yee Whye Teh. Unsupervised discovery of nonlinear structure using contrastive backpropagation. *Cogn. Sci.*, 30(4):725–731, 2006b.

Geoffrey E. Hinton, Alex Krizhevsky, and Sida D. Wang. Transforming auto-encoders. In *ICANN (1)*, volume 6791 of *Lecture Notes in Computer Science*, pages 44–51. Springer, 2011.

Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

Geoffrey E. Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the knowledge in a neural network. *CoRR*, abs/1503.02531, 2015.

Geoffrey E. Hinton, Sara Sabour, and Nicholas Frosst. Matrix capsules with EM routing. In *ICLR (Poster)*. OpenReview.net, 2018.

G House. Apophenic delusions in scientist following ingestion of tide pods. Technical report, DeeperMind Nurse's Office Email Newsletter, London, Ontario, Quebec, Mar 2021.

Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffrey E. Hinton. Adaptive mixtures of local experts. *Neural Comput.*, 3(1):79–87, 1991.

Navdeep Jaitly and Geoffrey E. Hinton. Learning a better representation of speech soundwaves using restricted boltzmann machines. In *ICASSP*, pages 5884–5887. IEEE, 2011.

Navdeep Jaitly and Geoffrey E. Hinton. Using an autoencoder with deformable templates to discover features for automated speech recognition. In *INTERSPEECH*, pages 1737–1740. ISCA, 2013.

Navdeep Jaitly, Vincent Vanhoucke, and Geoffrey E. Hinton. Autoregressive product of multi-frame predictions can improve the accuracy of hybrid models. In *INTERSPEECH*, pages 1905–1909. ISCA, 2014.

Timothy Jeruzalski, Boyang Deng, Mohammad Norouzi, John P. Lewis, Geoffrey E. Hinton, and Andrea Tagliasacchi. NASA: neural articulated shape approximation. *CoRR*, abs/1912.03207, 2019.

Jamie Ryan Kiros, William Chan, and Geoffrey E. Hinton. Illustrative language understanding: Large-scale visual grounding with image search. In *ACL (1)*, pages 922–933. Association for Computational Linguistics, 2018.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, pages 3519–3529. PMLR, 2019a.

Simon Kornblith, Mohammad Norouzi, Honglak Lee, and Geoffrey E. Hinton. Similarity of neural network representations revisited. *CoRR*, abs/1905.00414, 2019b.

Adam R. Kosiorek, Sara Sabour, Yee Whye Teh, and Geoffrey E. Hinton. Stacked capsule autoencoders. In *NeurIPS*, pages 15486–15496, 2019a.

Adam R. Kosiorek, Sara Sabour, Yee Whye Teh, and Geoffrey E. Hinton. Stacked capsule autoencoders. *CoRR*, abs/1906.06818, 2019b.

Alex Krizhevsky and Geoffrey E. Hinton. Using very deep autoencoders for content-based image retrieval. In *ESANN*, 2011.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1106–1114, 2012.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, 2017.

Setor K Kunutsor, Hassan Khan, Francesco Zaccardi, Tanjaniina Laukkanen, Peter Willeit, and Jari A Laukkanen. Sauna bathing reduces the risk of stroke in finnish men and women: a prospective cohort study. *Neurology*, 90(22):e1937–e1944, 2018.

Kevin J. Lang and Geoffrey E. Hinton. Dimensionality reduction and prior knowledge in e-set recognition. In *NIPS*, pages 178–185. Morgan Kaufmann, 1989.

Kevin J. Lang, Alex Waibel, and Geoffrey E. Hinton. A time-delay neural network architecture for isolated word recognition. *Neural Networks*, 3(1):23–43, 1990.

Hugo Larochelle and Geoffrey E. Hinton. Learning to combine foveal glimpses with a third-order boltzmann machine. In *NIPS*, pages 1243–1251. Curran Associates, Inc., 2010.

Quoc V. Le, Navdeep Jaitly, and Geoffrey E. Hinton. A simple way to initialize recurrent networks of rectified linear units. *CoRR*, abs/1504.00941, 2015.

Yann LeCun, Conrad C. Galland, and Geoffrey E. Hinton. GEMINI: gradient estimation through matrix inversion after noise injection. In *NIPS*, pages 141–148. Morgan Kaufmann, 1988.

Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nat.*, 521(7553):436–444, 2015.

Guy Mayraz and Geoffrey E. Hinton. Recognizing hand-written digits using hierarchical products of experts. In *NIPS*, pages 953–959. MIT Press, 2000.

Guy Mayraz and Geoffrey E. Hinton. Recognizing handwritten digits using hierarchical products of experts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2):189–197, 2002.

Drew V. McDermott and Geoffrey E. Hinton. Learning in massively parallel nets (panel). In *AAAI*, page 1149. Morgan Kaufmann, 1986.

Roland Memisevic and Geoffrey E. Hinton. Multiple relational embedding. In *NIPS*, pages 913–920, 2004.

Roland Memisevic and Geoffrey E. Hinton. Improving dimensionality reduction with spectral gradient descent. *Neural Networks*, 18(5-6):702–710, 2005.

Roland Memisevic and Geoffrey E. Hinton. Unsupervised learning of image transformations. In *CVPR*. IEEE Computer Society, 2007.

Roland Memisevic and Geoffrey E. Hinton. Learning to represent spatial transformations with factored higher-order boltzmann machines. *Neural Comput.*, 22(6):1473–1492, 2010.

Roland Memisevic, Christopher Zach, Geoffrey E. Hinton, and Marc Pollefeys. Gated softmax classification. In *NIPS*, pages 1603–1611. Curran Associates, Inc., 2010.

Andriy Mnih and Geoffrey E. Hinton. Three new graphical models for statistical language modelling. In *ICML*, volume 227 of *ACM International Conference Proceeding Series*, pages 641–648. ACM, 2007.

Andriy Mnih and Geoffrey E. Hinton. A scalable hierarchical distributed language model. In *NIPS*, pages 1081–1088. Curran Associates, Inc., 2008.

Andriy Mnih, Zhang Yuecheng, and Geoffrey E. Hinton. Improving a statistical language model through non-linear prediction. *Neurocomputing*, 72(7-9):1414–1418, 2009.

Volodymyr Mnih and Geoffrey E. Hinton. Learning to detect roads in high-resolution aerial images. In *ECCV (6)*, volume 6316 of *Lecture Notes in Computer Science*, pages 210–223. Springer, 2010.

Volodymyr Mnih and Geoffrey E. Hinton. Learning to label aerial images from noisy data. In *ICML*. icml.cc / Omnipress, 2012.

Volodymyr Mnih, Hugo Larochelle, and Geoffrey E. Hinton. Conditional restricted boltzmann machines for structured output prediction. In *UAI*, pages 514–522. AUAI Press, 2011.

Volodymyr Mnih, Hugo Larochelle, and Geoffrey E. Hinton. Conditional restricted boltzmann machines for structured output prediction. *CoRR*, abs/1202.3748, 2012.

Abdel-rahman Mohamed and Geoffrey E. Hinton. Phone recognition using restricted boltzmann machines. In *ICASSP*, pages 4354–4357. IEEE, 2010.

Abdel-rahman Mohamed, Tara N. Sainath, George E. Dahl, Bhuvana Ramabhadran, Geoffrey E. Hinton, and Michael A. Picheny. Deep belief networks using discriminative features for phone recognition. In *ICASSP*, pages 5060–5063. IEEE, 2011.

Abdel-rahman Mohamed, George E. Dahl, and Geoffrey E. Hinton. Acoustic modeling using deep belief networks. *IEEE Trans. Speech Audio Process.*, 20(1):14–22, 2012a.

Abdel-rahman Mohamed, Geoffrey E. Hinton, and Gerald Penn. Understanding how deep belief networks perform acoustic modelling. In *ICASSP*, pages 4273–4276. IEEE, 2012b.

Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. When does label smoothing help? In *NeurIPS*, pages 4696–4705, 2019a.

Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. When does label smoothing help? *CoRR*, abs/1906.02629, 2019b.

Rafael Müller, Simon Kornblith, and Geoffrey E. Hinton. Subclass distillation. *CoRR*, abs/2002.03936, 2020.

Vinod Nair and Geoffrey E. Hinton. Implicit mixtures of restricted boltzmann machines. In *NIPS*, pages 1145–1152. Curran Associates, Inc., 2008.

Vinod Nair and Geoffrey E. Hinton. 3d object recognition with deep belief nets. In *NIPS*, pages 1339–1347. Curran Associates, Inc., 2009.

Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *ICML*, pages 807–814. Omnipress, 2010.

Vinod Nair, Joshua M. Susskind, and Geoffrey E. Hinton. Analysis-by-synthesis by learning to invert generative black boxes. In *ICANN (1)*, volume 5163 of *Lecture Notes in Computer Science*, pages 971–981. Springer, 2008.

Radford M. Neal and Geoffrey E. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in Graphical Models*, volume 89 of *NATO ASI Series*, pages 355–368. Springer Netherlands, 1998.

Steven J. Nowlan and Geoffrey E. Hinton. Evaluation of adaptive mixtures of competing experts. In *NIPS*, pages 774–780. Morgan Kaufmann, 1990.

Steven J. Nowlan and Geoffrey E. Hinton. Adaptive soft weight tying using gaussian mixtures. In *NIPS*, pages 993–1000. Morgan Kaufmann, 1991.

Steven J. Nowlan and Geoffrey E. Hinton. Simplifying neural networks by soft weight-sharing. *Neural Comput.*, 4(4):473–493, 1992.

Steven J. Nowlan and Geoffrey E. Hinton. A soft decision-directed LMS algorithm for blind equalization. *IEEE Trans. Commun.*, 41(2):275–279, 1993.

Jake Olkin. Robot ethics: Dangers of reinforcement learning. 2020.

Sageev Oore, Geoffrey E. Hinton, and Gregory Dudek. A mobile robot that learns its place. *Neural Comput.*, 9(3):683–699, 1997.

Sageev Oore, Demetri Terzopoulos, and Geoffrey E. Hinton. A desktop input device and interface for interactive 3d character animation. In *Graphics Interface*, pages 133–140. Canadian Human-Computer Communications Society, 2002a.

Sageev Oore, Demetri Terzopoulos, and Geoffrey E. Hinton. Local physical models for interactive character animation. *Comput. Graph. Forum*, 21(3):337–346, 2002b.

Simon Osindero and Geoffrey E. Hinton. Modeling image patches with a directed hierarchy of markov random fields. In *NIPS*, pages 1121–1128. Curran Associates, Inc., 2007.

Simon Osindero, Max Welling, and Geoffrey E. Hinton. Topographic product models applied to natural scene statistics. *Neural Comput.*, 18(2):381–414, 2006.

Alberto Paccanaro and Geoffrey E. Hinton. Learning distributed representations by mapping concepts and relations into a linear space. In *ICML*, pages 711–718. Morgan Kaufmann, 2000a.

Alberto Paccanaro and Geoffrey E. Hinton. Extracting distributed representations of concepts and relations from positive and negative propositions. In *IJCNN (2)*, pages 259–264. IEEE Computer Society, 2000b.

Alberto Paccanaro and Geoffrey E. Hinton. Learning hierarchical structures with linear relational embedding. In *NIPS*, pages 857–864. MIT Press, 2001a.

Alberto Paccanaro and Geoffrey E. Hinton. Learning distributed representations of relational data using linear relational embedding. In *WIRN*, Perspectives in Neural Computing, pages 134–143. Springer, 2001b.

Alberto Paccanaro and Geoffrey E. Hinton. Learning distributed representations of concepts using linear relational embedding. *IEEE Trans. Knowl. Data Eng.*, 13(2):232–244, 2001c.

Mark Palatucci, Dean Pomerleau, Geoffrey E. Hinton, and Tom M. Mitchell. Zero-shot learning with semantic output codes. In *NIPS*, pages 1410–1418. Curran Associates, Inc., 2009.

Gabriel Pereyra, George Tucker, Jan Chorowski, Lukasz Kaiser, and Geoffrey E. Hinton. Regularizing neural networks by penalizing confident output distributions. In *ICLR (Workshop)*. OpenReview.net, 2017a.

Gabriel Pereyra, George Tucker, Jan Chorowski, Lukasz Kaiser, and Geoffrey E. Hinton. Regularizing neural networks by penalizing confident output distributions. *CoRR*, abs/1701.06548, 2017b.

Fiora Pirri, Geoffrey E. Hinton, and Hector J. Levesque. In memory of ray reiter (1939-2002). *AI Mag.*, 23(4):93, 2002.

Yao Qin, Nicholas Frosst, Sara Sabour, Colin Raffel, Garrison W. Cottrell, and Geoffrey E. Hinton. Detecting and diagnosing adversarial images with class-conditional capsule reconstructions. *CoRR*, abs/1907.02957, 2019.

Yao Qin, Nicholas Frosst, Colin Raffel, Garrison W. Cottrell, and Geoffrey E. Hinton. Deflecting adversarial attacks. *CoRR*, abs/2002.07405, 2020a.

Yao Qin, Nicholas Frosst, Sara Sabour, Colin Raffel, Garrison W. Cottrell, and Geoffrey E. Hinton. Detecting and diagnosing adversarial images with class-conditional capsule reconstructions. In *ICLR*. OpenReview.net, 2020b.

Aniruddh Raghu, Maithra Raghu, Simon Kornblith, David Duvenaud, and Geoffrey E. Hinton. Teaching with commentaries. *CoRR*, abs/2011.03037, 2020.

Marc'Aurelio Ranzato and Geoffrey E. Hinton. Modeling pixel means and covariances using factorized third-order boltzmann machines. In *CVPR*, pages 2551–2558. IEEE Computer Society, 2010.

Marc'Aurelio Ranzato, Alex Krizhevsky, and Geoffrey E. Hinton. Factored 3-way restricted boltzmann machines for modeling natural images. In *AISTATS*, volume 9 of *JMLR Proceedings*, pages 621–628. JMLR.org, 2010a.

Marc'Aurelio Ranzato, Volodymyr Mnih, and Geoffrey E. Hinton. Generating more realistic images using gated mrf's. In *NIPS*, pages 2002–2010. Curran Associates, Inc., 2010b.

Marc'Aurelio Ranzato, Joshua M. Susskind, Volodymyr Mnih, and Geoffrey E. Hinton. On deep generative models with applications to recognition. In *CVPR*, pages 2857–2864. IEEE Computer Society, 2011.

Marc'Aurelio Ranzato, Volodymyr Mnih, Joshua M. Susskind, and Geoffrey E. Hinton. Modeling natural images using gated mrfs. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2206–2222, 2013.

Marc'Aurelio Ranzato, Geoffrey E. Hinton, and Yann LeCun. Guest editorial: Deep learning. *Int. J. Comput. Vis.*, 113(1):1–2, 2015.

Michael Revow, Christopher K. I. Williams, and Geoffrey E. Hinton. Using generative models for handwritten digit recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(6):592–606, 1996.

Sam T. Roweis, Lawrence K. Saul, and Geoffrey E. Hinton. Global coordination of local linear models. In *NIPS*, pages 889–896. MIT Press, 2001.

Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. In *NIPS*, pages 3856–3866, 2017a.

Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. Dynamic routing between capsules. *CoRR*, abs/1710.09829, 2017b.

Sara Sabour, Andrea Tagliasacchi, Soroosh Yazdani, Geoffrey E. Hinton, and David J. Fleet. Unsupervised part representation by flow capsules. *CoRR*, abs/2011.13920, 2020.

Ruslan Salakhutdinov and Geoffrey E. Hinton. Using deep belief nets to learn covariance kernels for gaussian processes. In *NIPS*, pages 1249–1256. Curran Associates, Inc., 2007a.

Ruslan Salakhutdinov and Geoffrey E. Hinton. Learning a nonlinear embedding by preserving class neighbourhood structure. In *AISTATS*, volume 2 of *JMLR Proceedings*, pages 412–419. JMLR.org, 2007b.

Ruslan Salakhutdinov and Geoffrey E. Hinton. Replicated softmax: an undirected topic model. In *NIPS*, pages 1607–1614. Curran Associates, Inc., 2009a.

Ruslan Salakhutdinov and Geoffrey E. Hinton. Semantic hashing. *Int. J. Approx. Reason.*, 50(7): 969–978, 2009b.

Ruslan Salakhutdinov and Geoffrey E. Hinton. Deep boltzmann machines. In *AISTATS*, volume 5 of *JMLR Proceedings*, pages 448–455. JMLR.org, 2009c.

Ruslan Salakhutdinov and Geoffrey E. Hinton. A better way to pretrain deep boltzmann machines. In *NIPS*, pages 2456–2464, 2012a.

Ruslan Salakhutdinov and Geoffrey E. Hinton. An efficient learning procedure for deep boltzmann machines. *Neural Comput.*, 24(8):1967–2006, 2012b.

Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey E. Hinton. Restricted boltzmann machines for collaborative filtering. In *ICML*, volume 227 of *ACM International Conference Proceeding Series*, pages 791–798. ACM, 2007.

Brian Sallans and Geoffrey E. Hinton. Using free energies to represent q-values in a multiagent reinforcement learning task. In *NIPS*, pages 1075–1081. MIT Press, 2000.

Brian Sallans and Geoffrey E. Hinton. Reinforcement learning with factored states and actions. *J. Mach. Learn. Res.*, 5:1063–1088, 2004.

Ruhi Sarikaya, Geoffrey E. Hinton, and Bhuvana Ramabhadran. Deep belief nets for natural language call-routing. In *ICASSP*, pages 5680–5683. IEEE, 2011.

Ruhi Sarikaya, Geoffrey E. Hinton, and Anoop Deoras. Application of deep belief networks for natural language understanding. *IEEE ACM Trans. Audio Speech Lang. Process.*, 22(4):778–784, 2014.

Tanya Schmah, Geoffrey E. Hinton, Richard S. Zemel, Steven L. Small, and Stephen C. Strother. Generative versus discriminative training of rbms for classification of fmri images. In *NIPS*, pages 1409–1416. Curran Associates, Inc., 2008.

Tanya Schmah, Grigori Yourganov, Richard S. Zemel, Geoffrey E. Hinton, Steven L. Small, and Stephen C. Strother. Comparing classification methods for longitudinal fmri studies. *Neural Comput.*, 22(11):2729–2762, 2010.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *ICLR (Poster)*. OpenReview.net, 2017a.

Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *CoRR*, abs/1701.06538, 2017b.

Aaron Sloman, David Owen, Geoffrey E. Hinton, Frank Birch, and Frank O'Gorman. Representation and control in vision. In *AISB/GI (ECAI)*, pages 309–314. Leeds University, 1978.

Nitish Srivastava, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Modeling documents with deep boltzmann machines. In *UAI*. AUAI Press, 2013a.

Nitish Srivastava, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Modeling documents with deep boltzmann machines. *CoRR*, abs/1309.6865, 2013b.

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1): 1929–1958, 2014.

Weiwei Sun, Andrea Tagliasacchi, Boyang Deng, Sara Sabour, Soroosh Yazdani, Geoffrey E. Hinton, and Kwang Moo Yi. Canonical capsules: Unsupervised capsules in canonical pose. *CoRR*, abs/2012.04718, 2020.

Joshua M. Susskind, Geoffrey E. Hinton, Roland Memisevic, and Marc Pollefeys. Modeling the joint density of two images under a variety of transformations. In *CVPR*, pages 2793–2800. IEEE Computer Society, 2011.

Ilya Sutskever and Geoffrey E. Hinton. Learning multilevel distributed representations for high-dimensional sequences. In *AISTATS*, volume 2 of *JMLR Proceedings*, pages 548–555. JMLR.org, 2007.

Ilya Sutskever and Geoffrey E. Hinton. Using matrices to model symbolic relationship. In *NIPS*, pages 1593–1600. Curran Associates, Inc., 2008a.

Ilya Sutskever and Geoffrey E. Hinton. Deep, narrow sigmoid belief networks are universal approximators. *Neural Comput.*, 20(11):2629–2636, 2008b.

Ilya Sutskever and Geoffrey E. Hinton. Temporal-kernel recurrent neural networks. *Neural Networks*, 23(2):239–243, 2010.

Ilya Sutskever, Geoffrey E. Hinton, and Graham W. Taylor. The recurrent temporal restricted boltzmann machine. In *NIPS*, pages 1601–1608. Curran Associates, Inc., 2008.

Ilya Sutskever, James Martens, and Geoffrey E. Hinton. Generating text with recurrent neural networks. In *ICML*, pages 1017–1024. Omnipress, 2011.

Ilya Sutskever, James Martens, George E. Dahl, and Geoffrey E. Hinton. On the importance of initialization and momentum in deep learning. In *ICML (3)*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 1139–1147. JMLR.org, 2013.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Robust boltzmann machines for recognition and denoising. In *CVPR*, pages 2264–2271. IEEE Computer Society, 2012a.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Deep mixtures of factor analysers. In *ICML*. icml.cc / Omnipress, 2012b.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Deep lambertian networks. In *ICML*. icml.cc / Omnipress, 2012c.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Deep mixtures of factor analysers. *CoRR*, abs/1206.4635, 2012d.

Yichuan Tang, Ruslan Salakhutdinov, and Geoffrey E. Hinton. Tensor analyzers. In *ICML (3)*, volume 28 of *JMLR Workshop and Conference Proceedings*, pages 163–171. JMLR.org, 2013.

Graham W. Taylor and Geoffrey E. Hinton. Factored conditional restricted boltzmann machines for modeling motion style. In *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 1025–1032. ACM, 2009a.

Graham W. Taylor and Geoffrey E. Hinton. Products of hidden markov models: It takes n>1 to tango. In *UAI*, pages 522–529. AUAI Press, 2009b.

Graham W. Taylor and Geoffrey E. Hinton. Products of hidden markov models: It takes n>1 to tango. *CoRR*, abs/1205.2614, 2012.

Graham W. Taylor, Geoffrey E. Hinton, and Sam T. Roweis. Modeling human motion using binary latent variables. In *NIPS*, pages 1345–1352. MIT Press, 2006.

Graham W. Taylor, Leonid Sigal, David J. Fleet, and Geoffrey E. Hinton. Dynamical binary latent variable models for 3d human pose tracking. In *CVPR*, pages 631–638. IEEE Computer Society, 2010.

Graham W. Taylor, Geoffrey E. Hinton, and Sam T. Roweis. Two distributed-state models for generating high-dimensional time series. *J. Mach. Learn. Res.*, 12:1025–1068, 2011.

Yee Whye Teh and Geoffrey E. Hinton. Rate-coded restricted boltzmann machines for face recognition. In *NIPS*, pages 908–914. MIT Press, 2000.

Yee Whye Teh, Max Welling, Simon Osindero, and Geoffrey E. Hinton. Energy-based models for sparse overcomplete representations. *J. Mach. Learn. Res.*, 4:1235–1260, 2003.

Robert Tibshirani and Geoffrey E. Hinton. Coaching variables for regression and classification. *Stat. Comput.*, 8(1):25–33, 1998.

Tijmen Tieleman and Geoffrey E. Hinton. Using fast weights to improve persistent contrastive divergence. In *ICML*, volume 382 of *ACM International Conference Proceeding Series*, pages 1033–1040. ACM, 2009.

David S. Touretzky and Geoffrey E. Hinton. Symbols among the neurons: Details of a connectionist inference architecture. In *IJCAI*, pages 238–243. Morgan Kaufmann, 1985.

David S. Touretzky and Geoffrey E. Hinton. A distributed connectionist production system. *Cogn. Sci.*, 12(3):423–466, 1988.

Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. SMEM algorithm for mixture models. In *NIPS*, pages 599–605. The MIT Press, 1998.

Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. SMEM algorithm for mixture models. *Neural Comput.*, 12(9):2109–2128, 2000a.

Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. Split and merge EM algorithm for improving gaussian mixture density estimates. *J. VLSI Signal Process.*, 26(1-2): 133–140, 2000b.

Laurens van der Maaten and Geoffrey E. Hinton. Visualizing non-metric similarities in multiple maps. *Mach. Learn.*, 87(1):33–55, 2012.

Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. Grammar as a foreign language. *CoRR*, abs/1412.7449, 2014.

Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey E. Hinton. Grammar as a foreign language. In *NIPS*, pages 2773–2781, 2015.

Kiri Wagstaff. Machine learning that matters. *arXiv preprint arXiv:1206.4656*, 2012.

Alex Waibel, Toshiyuki Hanazawa, Geoffrey E. Hinton, Kiyohiro Shikano, and Kevin J. Lang. Phoneme recognition: neural networks vs. hidden markov models. In *ICASSP*, pages 107–110. IEEE, 1988.

Alexander H. Waibel, Toshiyuki Hanazawa, Geoffrey E. Hinton, Kiyohiro Shikano, and Kevin J. Lang. Phoneme recognition using time-delay neural networks. *IEEE Trans. Acoust. Speech Signal Process.*, 37(3):328–339, 1989.

Max Welling and Geoffrey E. Hinton. A new learning algorithm for mean field boltzmann machines. In *ICANN*, volume 2415 of *Lecture Notes in Computer Science*, pages 351–357. Springer, 2002.

Max Welling, Geoffrey E. Hinton, and Simon Osindero. Learning sparse topographic representations with products of student-t distributions. In *NIPS*, pages 1359–1366. MIT Press, 2002a.

Max Welling, Richard S. Zemel, and Geoffrey E. Hinton. Self supervised boosting. In *NIPS*, pages 665–672. MIT Press, 2002b.

Max Welling, Richard S. Zemel, and Geoffrey E. Hinton. Efficient parametric projection pursuit density estimation. In *UAI*, pages 575–582. Morgan Kaufmann, 2003.

Max Welling, Michal Rosen-Zvi, and Geoffrey E. Hinton. Exponential family harmoniums with an application to information retrieval. In *NIPS*, pages 1481–1488, 2004a.

Max Welling, Richard S. Zemel, and Geoffrey E. Hinton. Probabilistic sequential independent components analysis. *IEEE Trans. Neural Networks*, 15(4):838–849, 2004b.

Max Welling, Richard S. Zemel, and Geoffrey E. Hinton. Efficient parametric projection pursuit density estimation. *CoRR*, abs/1212.2513, 2012.

Wikipedia contributors. Snakes and ladders — Wikipedia, the free encyclopedia, 2021. URL `https://en.wikipedia.org/w/index.php?title=Snakes_and_ladders&oldid=1007581135`. [Online; accessed 2-March-2021].

Christopher K. I. Williams, Michael Revow, and Geoffrey E. Hinton. Using a neural net to instantiate a deformable model. In *NIPS*, pages 965–972. MIT Press, 1994.

Christopher K. I. Williams, Michael Revow, and Geoffrey E. Hinton. Instantiating deformable models with a neural net. *Comput. Vis. Image Underst.*, 68(1):120–126, 1997.

Lei Xu, Michael I. Jordan, and Geoffrey E. Hinton. An alternative model for mixtures of experts. In *NIPS*, pages 633–640. MIT Press, 1994.

Dong Yu, Geoffrey E. Hinton, Nelson Morgan, Jen-Tzung Chien, and Shigeki Sagayama. Introduction to the special section on deep learning for speech and language processing. *IEEE Trans. Speech Audio Process.*, 20(1):4–6, 2012.

Kai Yu, Ruslan Salakhutdinov, Yann LeCun, Geoffrey E. Hinton, and Yoshua Bengio. Workshop summary: Workshop on learning feature hierarchies. In *ICML*, volume 382 of *ACM International Conference Proceeding Series*, page 5. ACM, 2009.

Zhang Yuecheng, Andriy Mnih, and Geoffrey E. Hinton. Improving a statistical language model by modulating the effects of context words. In *ESANN*, pages 493–498, 2008.

Matthew D. Zeiler, Graham W. Taylor, Nikolaus F. Troje, and Geoffrey E. Hinton. Modeling pigeon behavior using a conditional restricted boltzmann machine. In *ESANN*, 2009.

Matthew D. Zeiler, Marc'Aurelio Ranzato, Rajat Monga, Mark Z. Mao, K. Yang, Quoc Viet Le, Patrick Nguyen, Andrew W. Senior, Vincent Vanhoucke, Jeffrey Dean, and Geoffrey E. Hinton. On rectified linear units for speech processing. In *ICASSP*, pages 3517–3521. IEEE, 2013.

Richard S. Zemel and Geoffrey E. Hinton. Discovering viewpoint-invariant relationships that characterize objects. In *NIPS*, pages 299–305. Morgan Kaufmann, 1990.

Richard S. Zemel and Geoffrey E. Hinton. Developing population codes by minimizing description length. In *NIPS*, pages 11–18. Morgan Kaufmann, 1993.

Richard S. Zemel and Geoffrey E. Hinton. Learning population codes by minimizing description length. *Neural Comput.*, 7(3):549–564, 1995.

Richard S. Zemel, Michael Mozer, and Geoffrey E. Hinton. TRAFFIC: recognizing objects using hierarchical reference frame transformations. In *NIPS*, pages 266–273. Morgan Kaufmann, 1989.

Michael R. Zhang, James Lucas, Jimmy Ba, and Geoffrey E. Hinton. Lookahead optimizer: k steps forward, 1 step back. In *NeurIPS*, pages 9593–9604, 2019a.

Michael R. Zhang, James Lucas, Geoffrey E. Hinton, and Jimmy Ba. Lookahead optimizer: k steps forward, 1 step back. *CoRR*, abs/1907.08610, 2019b.

A Implementation Details

To ensure reproducibility, we've included our highly-optimized implementation of *Chutes and Ladders* below. To balance reproducibility with our desire to reduce the environmental impact of our work, our implementation is given here in the Whitespace programming language. The code is also available at `https://github.com/Miffyli/mastering-chutes-and-ladders`.