

1. [Neural Lattice Language Models](#) (Buckman and Neubig 2018)
2. [Quasi-Recurrent Neural Networks](#) (Bradbury et al. 2016)
3. [Learned in translation: Contextualized word vectors](#) (McCann et al. 2017)
4. [Deep contextualized word representations Blogpost](#) (Peters et al. 2018)
5. [Google Vizier: A Service for Black-Box Optimization](#) (Golovin et al. 2017)
6. [Zoneout: Regularizing RNNs by Randomly Preserving Hidden Activations](#) (Krueger et al. 2016)
7. [Regularization of neural networks using dropconnect](#) (Wan et al. 2013)
8. [Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1](#) (Courbariaux et al. 2016)
9. [Hierarchical multiscale recurrent neural networks](#) (Chung, Ahn, and Bengio 2016)
10. [Deliberation Networks: Sequence Generation Beyond One-Pass Decoding](#) (Xia et al. 2017)
11. [Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks](#) (Salimans and Kingma 2016)
12. [Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour](#) (P. Goyal et al. 2017)
13. [Deep learning with dynamic computation graphs](#) (Looks et al. 2017)
14. [Towards universal paraphrastic sentence embeddings](#) (Wieting et al. 2015)
15. [Hyperbolic Entailment Cones for Learning Hierarchical Embeddings](#) (Ganea, Bécigneul, and Hofmann 2018) [**Urgent!**]
16. [Neural Autoregressive Flows](#) (Huang et al. 2018) [**Urgent!**]
17. [Attentive Language Models](#) (Salton, Ross, and Kelleher 2017)
18. [Weight normalization: A simple reparameterization to accelerate training of deep neural networks](#) (Salimans and Kingma 2016)
19. [Accurate, large minibatch SGD: training imagenet in 1 hour](#) (P. Goyal et al. 2017)

## References

- Bradbury, James, Stephen Merity, Caiming Xiong, and Richard Socher. 2016. “Quasi-Recurrent Neural Networks.” *arXiv Preprint arXiv:1611.01576*.
- Buckman, Jacob, and Graham Neubig. 2018. “Neural Lattice Language Models.” *arXiv Preprint arXiv:1803.05071*.
- Chung, Junyoung, Sungjin Ahn, and Yoshua Bengio. 2016. “Hierarchical Multiscale Recurrent Neural Networks.” *arXiv Preprint arXiv:1609.01704*.
- Courbariaux, Matthieu, Itay Hubara, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. 2016. “Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained To+ 1 or-1.” *arXiv Preprint arXiv:1602.02830*.
- Ganea, Octavian-Eugen, Gary Bécigneul, and Thomas Hofmann. 2018. “Hyper-

bolic Entailment Cones for Learning Hierarchical Embeddings.” *arXiv Preprint arXiv:1804.01882*.

Golovin, Daniel, Benjamin Solnik, Subhodeep Moitra, Greg Kochanski, John Karro, and D Sculley. 2017. “Google Vizier: A Service for Black-Box Optimization.” In *Proceedings of the 23rd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 1487–95. ACM.

Goyal, Priya, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. 2017. “Accurate, Large Minibatch Sgd: Training Imagenet in 1 Hour.” *arXiv Preprint arXiv:1706.02677*.

Huang, Chin-Wei, David Krueger, Alexandre Lacoste, and Aaron Courville. 2018. “Neural Autoregressive Flows.” *arXiv Preprint arXiv:1804.00779*.

Krueger, David, Tegan Maharaj, János Kramár, Mohammad Pezeshki, Nicolas Ballas, Nan Rosemary Ke, Anirudh Goyal, Yoshua Bengio, Aaron Courville, and Chris Pal. 2016. “Zoneout: Regularizing Rnns by Randomly Preserving Hidden Activations.” *arXiv Preprint arXiv:1606.01305*.

Looks, Moshe, Marcello Herreshoff, DeLesley Hutchins, and Peter Norvig. 2017. “Deep Learning with Dynamic Computation Graphs.” *arXiv Preprint arXiv:1702.02181*.

McCann, Bryan, James Bradbury, Caiming Xiong, and Richard Socher. 2017. “Learned in Translation: Contextualized Word Vectors.” In *Advances in Neural Information Processing Systems*, 6297–6308.

Peters, Matthew E, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. “Deep Contextualized Word Representations.” *arXiv Preprint arXiv:1802.05365*.

Salimans, Tim, and Diederik P Kingma. 2016. “Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks.” In *Advances in Neural Information Processing Systems*, 901–9.

Salton, Giancarlo, Robert Ross, and John Kelleher. 2017. “Attentive Language Models.” In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1:441–50.

Wan, Li, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. 2013. “Regularization of Neural Networks Using Dropconnect.” In *International Conference on Machine Learning*, 1058–66.

Wieting, John, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2015. “Towards Universal Paraphrastic Sentence Embeddings.” *arXiv Preprint arXiv:1511.08198*.

Xia, Yingce, Fei Tian, Lijun Wu, Jianxin Lin, Tao Qin, Nenghai Yu, and Tie-Yan Liu. 2017. “Deliberation Networks: Sequence Generation Beyond One-Pass Decoding.” In *Advances in Neural Information Processing Systems*, 1782–92.