

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)

## 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*.
- 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.
- 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*.
- 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*.

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.

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.