1 Theory

1.1 Rectified linear unit(ReLU)

- Glorot, X., Bordes, A. & Bengio. Y. Deep sparse rectifier neural networks. In Proc. 14th International Conference on Artificial Intelligence and Statistics 315–323 (2011).

1.2 Local minima

- Dauphin, Y. et al. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. In Proc. Advances in Neural Information Processing Systems 27 2933–2941 (2014).

- Choromanska, A., Henaff, M., Mathieu, M., Arous, G. B. & LeCun, Y. The loss surface of multilayer networks. In Proc. Conference on AI and Statistics http://arxiv.org/abs/1412.0233 (2014).

1.3 Dropout

- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. J. Machine Learning Res. 15, 1929–1958 (2014).

1.4 Pre-training

- Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast learning algorithm for deep belief nets. Neural Comp. 18, 1527–1554 (2006).

  This paper introduced a novel and effective way of training very deep neural networks by pre-training one hidden layer at a time using the unsupervised learning procedure for restricted Boltzmann machines. Article

- Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. Greedy layerwise training of deep networks. In Proc. Advances in Neural Information Processing Systems 19 153–160 (2006).

1.5   Learning distributed representations

- Montufar, G. & Morton, J. When does a mixture of products contain a product of mixtures? J. Discrete Math. 29, 321–347 (2014).

1.6   RNN and Memory network

- Pascanu, R., Mikolov, T. & Bengio, Y. On the difficulty of training recurrent neural networks. In Proc. 30th International Conference on Machine Learning 1310–1318 (2013).

- Graves, A., Wayne, G. & Danihelka, I. Neural Turing machines. http://arxiv.org/abs/1410.5401 (2014).

- Weston, J. Chopra, S. & Bordes, A. Memory networks. http://arxiv.org/abs/1410.3916 (2014).

- Weston, J., Bordes, A., Chopra, S. & Mikolov, T. Towards AI-complete question answering: a set of prerequisite toy tasks. http://arxiv.org/abs/1502.05698 (2015).

- Graves, A., Mohamed, A.-R. & Hinton, G. Speech recognition with deep recurrent neural networks. In Proc. International Conference on Acoustics, Speech and Signal Processing 6645–6649 (2013).

    - S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735– 1780, 1997.

1.7   Miscellaneous

- Montufar, G. F., Pascanu, R., Cho, K. & Bengio, Y. On the number of linear regions of deep neural networks. In Proc. Advances in Neural Information Processing Systems 27 2924–2932 (2014).

- Bottou, L. From machine learning to machine reasoning. Mach. Learn. 94, 133–149 (2014).

- S Rifai, YN Dauphin, P Vincent. The manifold tangent classifier. In Proc. Advances in Neural Information Processing Systems (2011).

2 Application

2.1 Image recognition & Computer vision

- Vinyals, O., Toshev, A., Bengio, S. & Erhan, D. Show and tell: a neural image caption generator. In Proc. International Conference on Machine Learning http://arxiv.org/abs/1502.03044 (2014).

- Krizhevsky, A., Sutskever, I. & Hinton, G. ImageNet classification with deep convolutional neural networks. In Proc. Advances in Neural Information Processing Systems 25 1090–1098 (2012).

- Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Learning hierarchical features for scene labeling. IEEE Trans. Pattern Anal. Mach. Intell. 35, 1915–1929 (2013). ISIPubMedArticle

- Tompson, J., Jain, A., LeCun, Y. & Bregler, C. Joint training of a convolutional network and a graphical model for human pose estimation. In Proc. Advances in Neural Information Processing Systems 27 1799–1807 (2014).

- Szegedy, C. et al. Going deeper with convolutions. Preprint at http://arxiv.org/abs/1409.4842 (2014).

- Sermanet, P. et al. Overfeat: integrated recognition, localization and detection using convolutional networks. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1312.6229 (2014).

- Girshick, R., Donahue, J., Darrell, T. & Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proc. Conference on Computer Vision and Pattern Recognition 580–587 (2014).

- Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1409.1556 (2014).

2.2 Speech recognition

- Mikolov, T., Deoras, A., Povey, D., Burget, L. & Cernocky, J. Strategies for training large scale neural network language models. In Proc. Automatic Speech Recognition and Understanding 196–201 (2011).

- Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Processing Magazine 29, 82–97 (2012).

  This joint paper from the major speech recognition laboratories, summarizing the breakthrough achieved with deep learning on the task of phonetic classification for automatic speech recognition, was the first major industrial application of deep learning. ISIArticle

- Sainath, T., Mohamed, A.-R., Kingsbury, B. & Ramabhadran, B. Deep convolutional neural networks for LVCSR. In Proc. Acoustics, Speech and Signal Processing 8614–8618 (2013).

- Dahl, G. E., Yu, D., Deng, L. & Acero, A. Context-dependent pretrained deep neural networks for large vocabulary speech recognition. IEEE Trans. Audio Speech Lang. Process. 20, 33–42 (2012).

2.3 Drug molecules

- Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E. & Svetnik, V. Deep neural nets as a method for quantitative structure-activity relationships. J. Chem. Inf. Model. 55, 263–274 (2015).

2.4 Accelerator data

- Ciodaro, T., Deva, D., de Seixas, J. & Damazio, D. Online particle detection with neural networks based on topological calorimetry information. J. Phys. Conf. Series 368, 012030 (2012). CASArticle

2.5 Reconstructing brain circuits

- Helmstaedter, M. et al. Connectomic reconstruction of the inner plexiform layer in the mouse retina. Nature 500, 168–174 (2013).

2.6 DNA and Gene

- Leung, M. K., Xiong, H. Y., Lee, L. J. & Frey, B. J. Deep learning of the tissue-regulated splicing code. Bioinformatics 30, i121–i129 (2014).

- Xiong, H. Y. et al. The human splicing code reveals new insights into the genetic determinants of disease. Science 347, 6218 (2015).

2.7 Mobile robots and self-driving cars

- Hadsell, R. et al. Learning long-range vision for autonomous off-road driving. J. Field Robot. 26, 120–144 (2009). ISIArticle

- Farabet, C., Couprie, C., Najman, L. & LeCun, Y. Scene parsing with multiscale feature learning, purity trees, and optimal covers. In Proc. International Conference on Machine Learning http://arxiv.org/abs/1202.2160 (2012).

2.8 Natural language understanding

- Collobert, R., et al. Natural language processing (almost) from scratch. J. Mach. Learn. Res. 12, 2493–2537 (2011).

2.9 Topic classification, sentiment analysis and question answering

- Bordes, A., Chopra, S. & Weston, J. Question answering with subgraph embeddings. In Proc. Empirical Methods in Natural Language Processing http://arxiv.org/abs/1406.3676v3 (2014).

2.10 Language translation

- Jean, S., Cho, K., Memisevic, R. & Bengio, Y. On using very large target vocabulary for neural machine translation. In Proc. ACL-IJCNLP http://arxiv.org/abs/1412.2007 (2015).

- Sutskever, I. Vinyals, O. & Le. Q. V. Sequence to sequence learning with neural networks. In Proc. Advances in Neural Information Processing Systems 27 3104–3112 (2014).

- Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1409.0473 (2015).

2.11   Vector representations of words

- Cho, K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Proc. Conference on Empirical Methods in Natural Language Processing 1724–1734 (2014).

- Schwenk, H. Continuous space language models. Computer Speech Lang. 21, 492–518 (2007). ISIArticle

- Socher, R., Lin, C. C-Y., Manning, C. & Ng, A. Y. Parsing natural scenes and natural language with recursive neural networks. In Proc. International Conference on Machine Learning 129–136 (2011).

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed representations of words and phrases and their compositionality. In Proc. Advances in Neural Information Processing Systems 26 3111–3119 (2013).

- Bahdanau, D., Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1409.0473 (2015).

2.12   Miscellaneous

- Sutskever, I., Martens, J. & Hinton, G. E. Generating text with recurrent neural networks. In Proc. 28th International Conference on Machine Learning 1017–1024 (2011).

- Xu, K. et al. Show, attend and tell: Neural image caption generation with visual attention. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1502.03044 (2015).

3 The future of deep learning

3.1 Unsupervised Learning

- Hinton, G. E., Dayan, P., Frey, B. J. & Neal, R. M. The wake-sleep algorithm for unsupervised neural networks. Science 268, 1558–1161 (1995). PubMedArticle

- Salakhutdinov, R. & Hinton, G. Deep Boltzmann machines. In Proc. International Conference on Artificial Intelligence and Statistics 448–455 (2009).

- Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In Proc. 25th International Conference on Machine Learning 1096–1103 (2008).

- Kavukcuoglu, K. et al. Learning convolutional feature hierarchies for visual recognition. In Proc. Advances in Neural Information Processing Systems 23 1090–1098 (2010).

- Gregor, K. & LeCun, Y. Learning fast approximations of sparse coding. In Proc. International Conference on Machine Learning 399–406 (2010).

- Ranzato, M., Mnih, V., Susskind, J. M. & Hinton, G. E. Modeling natural images using gated MRFs. IEEE Trans. Pattern Anal. Machine Intell. 35, 2206–2222 (2013).

- Bengio, Y., Thibodeau-Laufer, E., Alain, G. & Yosinski, J. Deep generative stochastic networks trainable by backprop. In Proc. 31st International Conference on Machine Learning 226–234 (2014).

- Kingma, D., Rezende, D., Mohamed, S. & Welling, M. Semi-supervised learning with deep generative models. In Proc. Advances in Neural Information Processing Systems 27 3581–3589 (2014).

3.2 Reinforcement learning

- Ba, J., Mnih, V. & Kavukcuoglu, K. Multiple object recognition with visual attention. In Proc. International Conference on Learning Representations http://arxiv.org/abs/1412.7755 (2014).

- Mnih, V. et al. Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015).