Awesome - Most Cited Deep Learning Papers

👀 awesome

A curated list of the most cited deep learning papers (since 2012)

We believe that there exist *classic* deep learning papers which are worth reading regardless of their application domain. Rather than providing overwhelming amount of papers, We would like to provide a *curated list* of the awesome deep learning papers which are considered as *must-reads* in certain research domains.

Background

Before this list, there exist other *awesome deep learning lists*, for example, Deep Vision and Awesome Recurrent Neural Networks. Also, after this list comes out, another awesome list for deep learning beginners. called Deep Learning Papers Reading Roadmap, has been created and loved by many deep learning researchers.

Although the *Roadmap List* includes lots of important deep learning papers, it feels overwhelming for me to read them all. As I mentioned in the introduction, I believe that seminal works can give us lessons regardless of their application domain. Thus, I would like to introduce**top 100 deep learning papers** here as a good starting point of overviewing deep learning researches.

Awesome list criteria

1. A list of **top 100 deep learning papers** published from 2012 to 2016 is suggested.
2. If a paper is added to the list, another paper (usually from *More Papers from 2016" section) should be removed to keep top 100 papers. (Thus, removing papers is also important contributions as well as adding papers)
3. Papers that are important, but failed to be included in the list, will be listed in *More than Top 100* section.
4. Please refer to *New Papers* and *Old Papers* sections for the papers published in recent 6 months or before 2012.

(Citation criteria)

- **< 6 months** : *New Papers* (by discussion)
- **2016** : +60 citations or "More Papers from 2016"
- **2015** : +200 citations
- **2014** : +400 citations
- **2013** : +600 citations
- **2012** : +800 citations
- **~2012** : *Old Papers* (by discussion)

Please note that we prefer seminal deep learning papers that can be applied to various researches rather than application papers. For that reason, some papers that meet the criteria may not be accepted while others can be. It depends on the impact of the paper, applicability to other researches scarcity of the research domain, and so on.

We need your contributions!

If you have any suggestions (missing papers, new papers, key researchers or typos), please feel free to edit and pull a request. (Please read the contributing guide for further instructions, though just letting me know the title of papers can also be a big contribution to us.)

(Update) You can download all top-100 papers with this python code. Also, bib file for all top-100 papers are available. Thanks,doodhwala and grepinsight(https://github.com/grepinsight)!

- Can anyone contribute the code for obtaining the statistics of the authors of Top-100 papers?

Table of Contents

(More than Top 100)

Understanding / Generalization / Transfer

- **Distilling the knowledge in a neural network** (2015), G. Hinton et al. [pdf]
- **Deep neural networks are easily fooled: High confidence predictions for unrecognizable images** (2015), A. Nguyen et al. [pdf]
- **How transferable are features in deep neural networks?** (2014), J. Yosinski et al. [pdf]
- **CNN features off-the-Shelf: An astounding baseline for recognition** (2014), A. Razavian et al. [pdf]
- **Learning and transferring mid-Level image representations using convolutional neural networks** (2014), M. Oquab et al. [pdf]
- **Visualizing and understanding convolutional networks** (2014), M. Zeiler and R. Fergus [pdf]
- **Decaf: A deep convolutional activation feature for generic visual recognition** (2014), J. Donahue et al. [pdf]

<!---[Key researchers] [Geoffrey Hinton](#), [Yoshua Bengio](#), [Jason Yosinski](#) -->

Optimization / Training Techniques

- **Batch normalization: Accelerating deep network training by reducing internal covariate shift** (2015), S. Loffe and C. Szegedy[pdf]
- **Delving deep into rectifiers: Surpassing human-level performance on imagenet classification** (2015), K. He et al. [pdf]
- **Dropout: A simple way to prevent neural networks from overfitting** (2014), N. Srivastava et al. [pdf]
- **Adam: A method for stochastic optimization** (2014), D. Kingma and J. Ba [pdf]
- **Improving neural networks by preventing co-adaptation of feature detectors** (2012), G. Hinton et al. [pdf]
- **Random search for hyper-parameter optimization** (2012) J. Bergstra and Y. Bengio [pdf]

<!---[Key researchers] [Geoffrey Hinton](#), [Yoshua Bengio](#), [Christian Szegedy](#), [Sergey Ioffe](#), [Kaming He](#), [Diederik P. Kingma](#)-->

Unsupervised / Generative Models

- **Pixel recurrent neural networks** (2016), A. Oord et al. [pdf]
- **Improved techniques for training GANs** (2016), T. Salimans et al. [pdf]
- **Unsupervised representation learning with deep convolutional generative adversarial networks** (2015), A. Radford et al. [pdf]
- **DRAW: A recurrent neural network for image generation** (2015), K. Gregor et al. [pdf]
- **Generative adversarial nets** (2014), I. Goodfellow et al. [pdf]
- **Auto-encoding variational Bayes** (2013), D. Kingma and M. Welling [pdf]
- **Building high-level features using large scale unsupervised learning** (2013), Q. Le et al. [pdf]

<!---[Key researchers] [Yoshua Bengio](#), [Ian Goodfellow](#), [Alex Graves](#)-->

Convolutional Neural Network Models

- **Rethinking the inception architecture for computer vision** (2016), C. Szegedy et al. [pdf]
- **Inception-v4, inception-resnet and the impact of residual connections on learning** (2016), C. Szegedy et

al. [pdf]
- **Identity Mappings in Deep Residual Networks** (2016), K. He et al. [pdf]
- **Deep residual learning for image recognition** (2016), K. He et al. [pdf]
- **Going deeper with convolutions** (2015), C. Szegedy et al. [pdf]
- **Very deep convolutional networks for large-scale image recognition** (2014), K. Simonyan and A. Zisserman [pdf]
- **Spatial pyramid pooling in deep convolutional networks for visual recognition** (2014), K. He et al. [pdf]
- **Return of the devil in the details: delving deep into convolutional nets** (2014), K. Chatfield et al. [pdf]
- **OverFeat: Integrated recognition, localization and detection using convolutional networks** (2013), P. Sermanet et al. [pdf]
- **Maxout networks** (2013), I. Goodfellow et al. [pdf]
- **Network in network** (2013), M. Lin et al. [pdf]
- **ImageNet classification with deep convolutional neural networks** (2012), A. Krizhevsky et al. [pdf]

<!---[Key researchers] Christian Szegedy, Kaming He, Shaoqing Ren, Jian Sun, Geoffrey Hinton, Yoshua Bengio, Yann LeCun-->

Image: Segmentation / Object Detection

- **You only look once: Unified, real-time object detection** (2016), J. Redmon et al. [pdf]
- **Region-based convolutional networks for accurate object detection and segmentation** (2016), R. Girshick et al. [pdf]
- **Fully convolutional networks for semantic segmentation** (2015), J. Long et al. [pdf]
- **Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks** (2015), S. Ren et al. [[pdf]] (http://papers.nips.cc/paper/5638-faster-r-cnn-towards-real-time-object-detection-with-region-proposal-networks.pdf)
- **Fast R-CNN** (2015), R. Girshick [pdf]
- **Rich feature hierarchies for accurate object detection and semantic segmentation** (2014), R. Girshick et al. [pdf]
- **Semantic image segmentation with deep convolutional nets and fully connected CRFs**, L. Chen et al. [pdf]
- **Learning hierarchical features for scene labeling** (2013), C. Farabet et al. [pdf]

<!---[Key researchers] Ross Girshick, Jeff Donahue, Trevor Darrell-->

Image / Video / Etc

- **Image Super-Resolution Using Deep Convolutional Networks** (2016), C. Dong et al. [pdf]
- **A neural algorithm of artistic style** (2015), L. Gatys et al. [pdf]
- **Deep visual-semantic alignments for generating image descriptions** (2015), A. Karpathy and L. Fei-Fei [pdf]
- **Show, attend and tell: Neural image caption generation with visual attention** (2015), K. Xu et al. [pdf]
- **Show and tell: A neural image caption generator (2015)**, O. Vinyals et al. [pdf]
- **Long-term recurrent convolutional networks for visual recognition and description** (2015), J. Donahue et al. [pdf]
- **VQA: Visual question answering** (2015), S. Antol et al. [pdf]
- **DeepFace: Closing the gap to human-level performance in face verification** (2014), Y. Taigman et al. [pdf]:
- **Large-scale video classification with convolutional neural networks** (2014), A. Karpathy et al. [pdf]
- **DeepPose: Human pose estimation via deep neural networks** (2014), A. Toshev and C. Szegedy [pdf]
- **Two-stream convolutional networks for action recognition in videos** (2014), K. Simonyan et al. [pdf]
- **3D convolutional neural networks for human action recognition** (2013), S. Ji et al. [pdf]

<!---[Key researchers] Oriol Vinyals, Andrej Karpathy-->

Recurrent Neural Network Models

- **Conditional random fields as recurrent neural networks** (2015), S. Zheng and S. Jayasumana. [pdf]
- **Memory networks** (2014), J. Weston et al. [pdf]
- **Neural turing machines** (2014), A. Graves et al. [pdf]
- **Generating sequences with recurrent neural networks** (2013), A. Graves. [pdf]

<!---[Key researchers] Alex Graves-->

Natural Language Process

- **A character-level decoder without explicit segmentation for neural machine translation** (2016), J. Chung et al. [pdf]
- **Exploring the limits of language modeling** (2016), R. Jozefowicz et al. [pdf]
- **Teaching machines to read and comprehend** (2015), K. Hermann et al. [pdf]
- **Effective approaches to attention-based neural machine translation** (2015), M. Luong et al. [pdf]
- **Neural machine translation by jointly learning to align and translate** (2014), D. Bahdanau et al. [pdf]
- **Sequence to sequence learning with neural networks** (2014), I. Sutskever et al. [pdf]
- **Learning phrase representations using RNN encoder-decoder for statistical machine translation** (2014), K. Cho et al. [pdf]
- **A convolutional neural network for modeling sentences** (2014), N. Kalchbrenner et al. [pdf]
- **Convolutional neural networks for sentence classification** (2014), Y. Kim [pdf]
- **Glove: Global vectors for word representation** (2014), J. Pennington et al. [pdf]
- **Distributed representations of sentences and documents** (2014), Q. Le and T. Mikolov [pdf]
- **Distributed representations of words and phrases and their compositionality** (2013), T. Mikolov et al. [pdf]
- **Efficient estimation of word representations in vector space** (2013), T. Mikolov et al. [pdf]
- **Recursive deep models for semantic compositionality over a sentiment treebank** (2013), R. Socher et al. [pdf]

<!---[Key researchers] Kyunghyun Cho, Oriol Vinyals, Richard Socher, Tomas Mikolov, Christopher D. Manning, Yoshua Bengio-->

Speech / Other Domain

- **End-to-end attention-based large vocabulary speech recognition** (2016), D. Bahdanau et al. [pdf]
- **Deep speech 2: End-to-end speech recognition in English and Mandarin** (2015), D. Amodei et al. [pdf]
- **Speech recognition with deep recurrent neural networks** (2013), A. Graves [pdf]
- **Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups** (2012), G. Hinton et al. [pdf]
- **Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition** (2012) G. Dahl et al. [pdf]
- **Acoustic modeling using deep belief networks** (2012), A. Mohamed et al. [pdf]

<!---[Key researchers] Alex Graves, Geoffrey Hinton, Dong Yu-->

Reinforcement Learning / Robotics

- **End-to-end training of deep visuomotor policies** (2016), S. Levine et al. [pdf]
- **Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection** (2016), S. Levine et al. [pdf]
- **Asynchronous methods for deep reinforcement learning** (2016), V. Mnih et al. [pdf]
- **Deep Reinforcement Learning with Double Q-Learning** (2016), H. Hasselt et al. [pdf]
- **Mastering the game of Go with deep neural networks and tree search** (2016), D. Silver et al. [pdf]
- **Continuous control with deep reinforcement learning** (2015), T. Lillicrap et al. [pdf]
- **Human-level control through deep reinforcement learning** (2015), V. Mnih et al. [pdf]
- **Deep learning for detecting robotic grasps** (2015), I. Lenz et al. [pdf]
- **Playing atari with deep reinforcement learning** (2013), V. Mnih et al. [pdf])

<!---[Key researchers] Sergey Levine, Volodymyr Mnih, David Silver-->

More Papers from 2016

- **Layer Normalization** (2016), J. Ba et al. [pdf]
- **Learning to learn by gradient descent by gradient descent** (2016), M. Andrychowicz et al. [pdf]
- **Domain-adversarial training of neural networks** (2016), Y. Ganin et al. [pdf]
- **WaveNet: A Generative Model for Raw Audio** (2016), A. Oord et al. [pdf] [web]
- **Colorful image colorization** (2016), R. Zhang et al. [pdf]
- **Generative visual manipulation on the natural image manifold** (2016), J. Zhu et al. [pdf]

- **Texture networks: Feed-forward synthesis of textures and stylized images** (2016), D Ulyanov et al. [pdf]
- **SSD: Single shot multibox detector** (2016), W. Liu et al. [pdf]
- **SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 1MB model size** (2016), F. landola et al. [pdf]
- **Eie: Efficient inference engine on compressed deep neural network** (2016), S. Han et al. [pdf]
- **Binarized neural networks: Training deep neural networks with weights and activations constrained to+ 1 or-1** (2016), M. Courbariaux et al. [pdf]
- **Dynamic memory networks for visual and textual question answering** (2016), C. Xiong et al. [pdf]
- **Stacked attention networks for image question answering** (2016), Z. Yang et al. [pdf]
- **Hybrid computing using a neural network with dynamic external memory** (2016), A. Graves et al. [pdf]
- **Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation** (2016), Y. Wu et al.[pdf]

New papers

Newly published papers (< 6 months) which are worth reading

- Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models, S. Ioffe. [pdf]
- Wasserstein GAN, M. Arjovsky et al. [pdf]
- Understanding deep learning requires rethinking generalization, C. Zhang et al. [pdf]

Old Papers

Classic papers published before 2012

- An analysis of single-layer networks in unsupervised feature learning (2011), A. Coates et al. [pdf]
- Deep sparse rectifier neural networks (2011), X. Glorot et al. [pdf]
- Natural language processing (almost) from scratch (2011), R. Collobert et al. [pdf]
- Recurrent neural network based language model (2010), T. Mikolov et al. [pdf]
- Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion (2010), P. Vincent et al. [pdf]
- Learning mid-level features for recognition (2010), Y. Boureau [pdf]
- A practical guide to training restricted boltzmann machines (2010), G. Hinton [pdf]
- Understanding the difficulty of training deep feedforward neural networks (2010), X. Glorot and Y. Bengio [pdf]
- Why does unsupervised pre-training help deep learning (2010), D. Erhan et al. [pdf]
- Learning deep architectures for AI (2009), Y. Bengio. [pdf]
- Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations (2009), H. Lee et al. [pdf]
- Greedy layer-wise training of deep networks (2007), Y. Bengio et al. [pdf]
- Reducing the dimensionality of data with neural networks, G. Hinton and R. Salakhutdinov. [pdf]
- A fast learning algorithm for deep belief nets (2006), G. Hinton et al. [pdf]
- Gradient-based learning applied to document recognition (1998), Y. LeCun et al. [pdf]
- Long short-term memory (1997), S. Hochreiter and J. Schmidhuber. [pdf]

HW / SW / Dataset

- OpenAI gym (2016), G. Brockman et al. [pdf]
- TensorFlow: Large-scale machine learning on heterogeneous distributed systems (2016), M. Abadi et al. [pdf]
- Theano: A Python framework for fast computation of mathematical expressions, R. Al-Rfou et al.
- Torch7: A matlab-like environment for machine learning, R. Collobert et al. [pdf]
- MatConvNet: Convolutional neural networks for matlab (2015), A. Vedaldi and K. Lenc [pdf]
- Imagenet large scale visual recognition challenge (2015), O. Russakovsky et al. [pdf]
- Caffe: Convolutional architecture for fast feature embedding (2014), Y. Jia et al. [pdf]

Book / Survey / Review

- Deep learning (Book, 2016), Goodfellow et al. [html]
- LSTM: A search space odyssey (2016), K. Greff et al. [pdf]

- Deep learning (2015), Y. LeCun, Y. Bengio and G. Hinton [pdf]
- Deep learning in neural networks: An overview (2015), J. Schmidhuber [pdf]
- Representation learning: A review and new perspectives (2013), Y. Bengio et al. [pdf]

Video Lectures / Tutorials / Blogs

(Lectures)

- CS231n, Convolutional Neural Networks for Visual Recognition, Stanford University [web]
- CS224d, Deep Learning for Natural Language Processing, Stanford University [web]
- Oxford Deep NLP 2017, Deep Learning for Natural Language Processing, University of Oxford [web]

(Tutorials)

- NIPS 2016 Tutorials, Long Beach [web]
- ICML 2016 Tutorials, New York City [web]
- ICLR 2016 Videos, San Juan [web]
- Deep Learning Summer School 2016, Montreal [web]
- Bay Area Deep Learning School 2016, Stanford [web]

(Blogs)

- Andrej Karpathy Blog [web]
- Colah's Blog [Web]
- WildML [Web]
- FastML [web]

Appendix: More than Top 100

(2016)

- Dermatologist-level classification of skin cancer with deep neural networks (2017), A. Esteva et al. [html]
- Weakly supervised object localization with multi-fold multiple instance learning (2017), R. Gokberk et al. [pdf]
- Brain tumor segmentation with deep neural networks (2017), M. Havaei et al. [pdf]
- Professor Forcing: A New Algorithm for Training Recurrent Networks (2016), A. Lamb et al. [pdf]
- Adversarially learned inference (2016), V. Dumoulin et al. [web][pdf]
- Understanding convolutional neural networks (2016), J. Koushik [pdf]
- Taking the human out of the loop: A review of bayesian optimization (2016), B. Shahriari et al. [pdf]
- Adaptive computation time for recurrent neural networks (2016), A. Graves [pdf]
- Densely connected convolutional networks (2016), G. Huang et al. [pdf]
- Continuous deep q-learning with model-based acceleration (2016), S. Gu et al. [pdf]
- A thorough examination of the cnn/daily mail reading comprehension task (2016), D. Chen et al. [pdf]
- Achieving open vocabulary neural machine translation with hybrid word-character models, M. Luong and C. Manning. [pdf]
- Very Deep Convolutional Networks for Natural Language Processing (2016), A. Conneau et al. [pdf]
- Bag of tricks for efficient text classification (2016), A. Joulin et al. [pdf]
- Efficient piecewise training of deep structured models for semantic segmentation (2016), G. Lin et al. [pdf]
- Learning to compose neural networks for question answering (2016), J. Andreas et al. [pdf]
- Perceptual losses for real-time style transfer and super-resolution (2016), J. Johnson et al. [pdf]
- Reading text in the wild with convolutional neural networks (2016), M. Jaderberg et al. [pdf]
- What makes for effective detection proposals? (2016), J. Hosang et al. [pdf]
- Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks (2016), S. Bell et al. [pdf].
- Instance-aware semantic segmentation via multi-task network cascades (2016), J. Dai et al. [pdf]
- Conditional image generation with pixelcnn decoders (2016), A. van den Oord et al. [pdf]
- Deep networks with stochastic depth (2016), G. Huang et al., [pdf]
- Generative Short Term Stochastic Gibbs Networks 2016), I. Lenz et al. [pdf]

(2015)

- Spatial transformer network (2015), M. Jaderberg et al., [pdf]
- Ask your neurons: A neural-based approach to answering questions about images (2015), M. Malinowski et al. [pdf]
- Exploring models and data for image question answering (2015), M. Ren et al. [pdf]
- Are you talking to a machine? dataset and methods for multilingual image question (2015), H. Gao et al. [pdf]
- Mind's eye: A recurrent visual representation for image caption generation (2015), X. Chen and C. Zitnick. [pdf]
- From captions to visual concepts and back (2015), H. Fang et al. [pdf].
- Towards AI-complete question answering: A set of prerequisite toy tasks (2015), J. Weston et al. [pdf]
- Ask me anything: Dynamic memory networks for natural language processing (2015), A. Kumar et al. [pdf]
- Unsupervised learning of video representations using LSTMs (2015), N. Srivastava et al. [pdf]
- Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding (2015), S. Han et al.[pdf]
- Improved semantic representations from tree-structured long short-term memory networks (2015), K. Tai et al. [pdf]
- Character-aware neural language models (2015), Y. Kim et al. [pdf]
- Grammar as a foreign language (2015), O. Vinyals et al. [pdf]
- Trust Region Policy Optimization (2015), J. Schulman et al. [pdf]
- Beyond short snippets: Deep networks for video classification (2015) [pdf]
- Learning Deconvolution Network for Semantic Segmentation (2015), H. Noh et al. [pdf]
- Learning spatiotemporal features with 3d convolutional networks (2015), D. Tran et al. [pdf]
- Understanding neural networks through deep visualization (2015), J. Yosinski et al. [pdf]
- An Empirical Exploration of Recurrent Network Architectures (2015), R. Jozefowicz et al. [pdf]
- Training very deep networks (2015), R. Srivastava et al. [pdf]
- Deep generative image models using a laplacian pyramid of adversarial networks (2015), E.Denton et al. [pdf]
- Gated Feedback Recurrent Neural Networks (2015), J. Chung et al. [pdf]
- Fast and accurate deep network learning by exponential linear units (ELUS) (2015), D. Clevert et al. [pdf]
- Pointer networks (2015), O. Vinyals et al. [pdf]
- Visualizing and Understanding Recurrent Networks (2015), A. Karpathy et al. [pdf]
- Attention-based models for speech recognition (2015), J. Chorowski et al. [pdf]
- End-to-end memory networks (2015), S. Sukbaatar et al. [pdf]
- Describing videos by exploiting temporal structure (2015), L. Yao et al. [pdf]
- A neural conversational model (2015), O. Vinyals and Q. Le. [pdf]

(~2014)

- Learning a Deep Convolutional Network for Image Super-Resolution (2014, C. Dong et al. [pdf]
- Recurrent models of visual attention (2014), V. Mnih et al. [pdf]
- Empirical evaluation of gated recurrent neural networks on sequence modeling (2014), J. Chung et al. [pdf]
- Addressing the rare word problem in neural machine translation (2014), M. Luong et al. [pdf]
- On the properties of neural machine translation: Encoder-decoder approaches (2014), K. Cho et. al.
- Recurrent neural network regularization (2014), W. Zaremba et al. [pdf]
- Intriguing properties of neural networks (2014), C. Szegedy et al. [pdf]
- Towards end-to-end speech recognition with recurrent neural networks (2014), A. Graves and N. Jaitly. [pdf]
- Scalable object detection using deep neural networks (2014), D. Erhan et al. [pdf]
- On the importance of initialization and momentum in deep learning (2013), I. Sutskever et al. [pdf]
- Regularization of neural networks using dropconnect (2013), L. Wan et al. [pdf]
- Learning Hierarchical Features for Scene Labeling (2013), C. Farabet et al. [pdf]
- Linguistic Regularities in Continuous Space Word Representations (2013), T. Mikolov et al. [pdf]
- Large scale distributed deep networks (2012), J. Dean et al. [pdf]

Acknowledgement