

---

## neural-network-papers

### Table of Contents

1. Other Lists
2. Surveys
3. Books
4. Datasets
5. Pretrained Models
6. Programming Frameworks
7. Learning to Compute
8. Natural Language Processing
9. Convolutional Neural Networks
10. Recurrent Neural Networks
11. Convolutional Recurrent Neural Networks
12. Adversarial Neural Networks
13. Autoencoders
14. Restricted Boltzmann Machines
15. Biologically Plausible Learning
16. Supervised Learning
17. Unsupervised Learning
18. Reinforcement Learning
19. Theory
20. Quantum Computing
21. Training Innovations
22. Parallel Training
23. Weight Compression
24. Numerical Precision
25. Numerical Optimization
26. Motion Planning
27. Simulation
28. Hardware
29. Cognitive Architectures
30. Computational Creativity
31. Cryptography
32. Distributed Computing
33. Clustering

---

## Other Lists

- DeepLearning.University – An Annotated Deep Learning Bibliography | Memkite ([github.com/memkite/DeepLearningUniversity](https://github.com/memkite/DeepLearningUniversity))
- Deep Learning for NLP resources
- Reading List « Deep Learning
- Reading lists for new MILA students
- Awesome Recurrent Neural Networks
- Awesome Deep Learning
- Deep learning Reading List
- A curated list of speech and natural language processing resources ([github.com/edobashira/speech-language-processing](https://github.com/edobashira/speech-language-processing))
- CS089/CS189 | Deep Learning | Spring 2015

## Surveys

- Deep Learning
- Deep Learning in Neural Networks: An Overview
- Deep neural networks: a new framework for modelling biological vision and brain information processing
- A Primer on Neural Network Models for Natural Language Processing
- Natural Language Understanding with Distributed Representation

## Books

- Deep Learning
- Neural Networks and Deep Learning

## Datasets

- Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks ([fb.ai/babi](https://fb.ai/babi))
- Teaching Machines to Read and Comprehend ([github.com/deepmind/rc-data](https://github.com/deepmind/rc-data))
- One Billion Word Benchmark for Measuring Progress in Statistical Language Modeling ([github.com/ciprian-chelba/1-billion-word-language-modeling-benchmark](https://github.com/ciprian-chelba/1-billion-word-language-modeling-benchmark))
- The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems ([cs.mcgill.ca/~jpineau/datasets/ubuntu-corpus-1.0](https://cs.mcgill.ca/~jpineau/datasets/ubuntu-corpus-1.0))
- Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books (BookCorpus)

- 
- Every publicly available Reddit comment, for research.
  - Stack Exchange Data Dump
  - Europarl: A Parallel Corpus for Statistical Machine Translation ([www.statmt.org/europarl/](http://www.statmt.org/europarl/))
  - RTE Knowledge Resources

## Pretrained Models

- Model Zoo
- word2vec
  - GoogleNews-vectors-negative300.bin.gz
  - freebase-vectors-skipgram1000.bin.gz
- GloVe
- SENNA

## Programming Frameworks

- TensorFlow ([tensorflow.org](http://tensorflow.org)) ([github.com/tensorflow/tensorflow](https://github.com/tensorflow/tensorflow))
- Caffe: Convolutional Architecture for Fast Feature Embedding ([github.com/BVLC/caffe](https://github.com/BVLC/caffe)) ([github.com/amd/OpenCL-caffe](https://github.com/amd/OpenCL-caffe))
  - Improving Caffe: Some Refactoring ([github.com/Yangqing/caffe2](https://github.com/Yangqing/caffe2))
- Theano: A CPU and GPU Math Compiler in Python ([github.com/Theano/Theano](https://github.com/Theano/Theano))
  - Theano: new features and speed improvements
  - Blocks and Fuel: Frameworks for deep learning ([github.com/mila-udem/blocks](https://github.com/mila-udem/blocks)) ([github.com/mila-udem/blocks-examples](https://github.com/mila-udem/blocks-examples)) ([github.com/mila-udem/fuel](https://github.com/mila-udem/fuel))
  - Announcing Computation Graph Toolkit ([github.com/joschu/cgt](https://github.com/joschu/cgt))
- Torch7: A Matlab-like Environment for Machine Learning ([github.com/torch/distro](https://github.com/torch/distro))
- Brainstorm
- Deeplearning4j - Open-source, distributed deep learning for the JVM ([github.com/deeplearning4j/deeplearning4j](https://github.com/deeplearning4j/deeplearning4j))
  - ND4J: N-Dimensional Arrays for Java N-Dimensional Scientific Computing for Java ([github.com/deeplearning4j/nd4j](https://github.com/deeplearning4j/nd4j))
- linalg: Matrix Computations in Apache Spark
- cuDNN: Efficient Primitives for Deep Learning
- Fast Convolutional Nets With fbfft: A GPU Performance Evaluation ([github.com/facebook/fbcuda](https://github.com/facebook/fbcuda))
- Guide to NumPy
- Probabilistic Programming in Python using PyMC

---

## Learning to Compute

- Neural GPUs Learn Algorithms
- A Roadmap towards Machine Intelligence
- On Learning to Think: Algorithmic Information Theory for Novel Combinations of Reinforcement Learning Controllers and Recurrent Neural World Models
- Binding via Reconstruction Clustering
- Neural Random-Access Machines
- Learning Simple Algorithms from Examples
- Neural Programmer: Inducing Latent Programs with Gradient Descent
- Neural Programmer-Interpreters
- Neural Turing Machines
  - Reinforcement Learning Neural Turing Machines
  - Structured Memory for Neural Turing Machines
- Memory Networks ([github.com/facebook/MemNN](https://github.com/facebook/MemNN))
  - End-To-End Memory Networks
- Learning to Transduce with Unbounded Memory
- Inferring Algorithmic Patterns with Stack-Augmented Recurrent Nets ([github.com/facebook/Stack-RNN](https://github.com/facebook/Stack-RNN))
  - Learning Context Free Grammars: Limitations of a Recurrent Neural Network with an External Stack Memory
  - A connectionist symbol manipulator that discovers the structure of context-free languages
- Feedforward Sequential Memory Neural Networks without Recurrent Feedback
- Pointer Networks
- On End-to-End Program Generation from User Intention by Deep Neural Networks
- Deep Knowledge Tracing ([github.com/chrispiech/DeepKnowledgeTracing](https://github.com/chrispiech/DeepKnowledgeTracing))
- Learning to Execute
- Tree-structured composition in neural networks without tree-structured architectures
- Grammar as a Foreign Language
- Learning To Learn Using Gradient Descent
- Learning to control fast-weight memories: An alternative to recurrent nets ([ftp://ftp.idsia.ch/pub/juergen/fastweight.pdf](http://ftp.idsia.ch/pub/juergen/fastweight.pdf))
- An introspective network that can learn to run its own weight change algorithm ([ftp://ftp.idsia.ch/pub/juergen/introspective.pdf](http://ftp.idsia.ch/pub/juergen/introspective.pdf))
- Goedel Machines: Self-Referential Universal Problem Solvers Making Provably Optimal Self-Improvements
- Optimal Ordered Problem Solver ([ftp://ftp.idsia.ch/pub/juergen/oopsmlj.pdf](http://ftp.idsia.ch/pub/juergen/oopsmlj.pdf))

- 
- The Fastest and Shortest Algorithm for All Well-Defined Problems (<ftp://ftp.idsia.ch/pub/techrep/IDSIA-16-00.ps.gz>)
  - The Speed Prior: A New Simplicity Measure Yielding Near-Optimal Computable Predictions (<ftp://ftp.idsia.ch/pub/juergen/coltspeed.pdf>)
  - Learning Game of Life with a Convolutional Neural Network ([github.com/DanielRapp/cnn-gol](https://github.com/DanielRapp/cnn-gol))

## **Natural Language Processing**

- Deep Learning, NLP, and Representations
- Language Models for Image Captioning: The Quirks and What Works
- Zero-Shot Learning Through Cross-Modal Transfer
- On Using Very Large Target Vocabulary for Neural Machine Translation
- BlackOut: Speeding up Recurrent Neural Network Language Models With Very Large Vocabularies
- Deep Unordered Composition Rivals Syntactic Methods for Text Classification

## **Word Vectors**

- So similar and yet incompatible: Toward automated identification of semantically compatible words ([github.com/germank/compatibility-naacl2015](https://github.com/germank/compatibility-naacl2015))
- Controlled Experiments for Word Embeddings ([github.com/benjaminwilson/word2vec-norm-experiments](https://github.com/benjaminwilson/word2vec-norm-experiments))
- Natural Language Processing (almost) from Scratch
- Efficient Estimation of Word Representations in Vector Space
  - Distributed Representations of Words and Phrases and their Compositionality
  - Exploiting Similarities among Languages for Machine Translation
- GloVe: Global Vectors for Word Representation
- Learning to Understand Phrases by Embedding the Dictionary
- Inverted indexing for cross-lingual NLP
- Random walks on discourse spaces: a new generative language model with applications to semantic word embeddings
- Breaking Sticks and Ambiguities with Adaptive Skip-gram
- Language Recognition using Random Indexing

---

## Sentence and Paragraph Vectors

- Generating Sentences from a Continuous Space
- Distributed Representations of Sentences and Documents
- Document Embedding with Paragraph Vectors
- A Fixed-Size Encoding Method for Variable-Length Sequences with its Application to Neural Network Language Models
- Skip-Thought Vectors ([github.com/ryankiros/skip-thoughts](https://github.com/ryankiros/skip-thoughts))
- From Word Embeddings To Document Distances

## Character Vectors

- Alternative structures for character-level RNNs
- Character-based Neural Machine Translation
- Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation ([github.com/wlin12/JNN](https://github.com/wlin12/JNN))
- Character-Aware Neural Language Models ([github.com/yoonkim/lstm-char-cnn](https://github.com/yoonkim/lstm-char-cnn))
- Modeling Order in Neural Word Embeddings at Scale
- Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs

## Attention Mechanisms

- Neural Machine Translation by Jointly Learning to Align and Translate
- Ask Me Anything: Dynamic Memory Networks for Natural Language Processing
- Attention with Intention for a Neural Network Conversation Model

## Sequence-to-Sequence Learning

- Multi-task Sequence to Sequence Learning
- Order Matters: Sequence to sequence for sets
- Task Loss Estimation for Sequence Prediction
- Semi-supervised Sequence Learning
- A Hierarchical Neural Autoencoder for Paragraphs and Documents ([github.com/jiweil/Hierarchical-Neural-Autoencoder](https://github.com/jiweil/Hierarchical-Neural-Autoencoder))
- Sequence to Sequence Learning with Neural Networks
- Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation

- 
- Neural Transformation Machine: A New Architecture for Sequence-to-Sequence Learning
  - On Using Monolingual Corpora in Neural Machine Translation

## **Language Understanding**

- Reasoning about Entailment with Neural Attention
- The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations
- Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding
- Language Understanding for Text-based Games Using Deep Reinforcement Learning ([github.com/karthikncode/text-world-player](https://github.com/karthikncode/text-world-player))

## **Question Answering, and Conversing**

- A Cognitive Neural Architecture Able to Learn and Communicate through Natural Language ([github.com/golosio/annabell](https://github.com/golosio/annabell))
- Large-scale Simple Question Answering with Memory Networks
- Reasoning in Vector Space: An Exploratory Study of Question Answering
- Deep Learning for Answer Sentence Selection
- Neural Responding Machine for Short-Text Conversation
- A Neural Conversational Model
- VQA: Visual Question Answering
- Question Answering with Subgraph Embeddings
- Hierarchical Neural Network Generative Models for Movie Dialogues
- Ask Your Neurons: A Neural-based Approach to Answering Questions about Images
- Are You Talking to a Machine? Dataset and Methods for Multilingual Image Question Answering

## **Convolutional**

- Basset: Learning the regulatory code of the accessible genome with deep convolutional neural networks. ([github.com/davek44/Basset](https://github.com/davek44/Basset))
- A Convolutional Neural Network for Modelling Sentences
- Convolutional Neural Networks for Sentence Classification ([github.com/yoonskim/CNN\\_sentence](https://github.com/yoonskim/CNN_sentence))
- Text Understanding from Scratch
  - Character-level Convolutional Networks for Text Classification
- DeepWriterID: An End-to-end Online Text-independent Writer Identification System

- 
- Encoding Source Language with Convolutional Neural Network for Machine Translation
  - Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling
  - Convolutional Neural Network Architectures for Matching Natural Language Sentences

## **Recurrent**

- Long Short-Term Memory Over Tree Structures
- Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks
- CCG Supertagging with a Recurrent Neural Network

## **Convolutional Neural Networks**

- Spatial Transformer Networks
- SimNets: A Generalization of Convolutional Networks
- Fast Algorithms for Convolutional Neural Networks
- Striving for Simplicity: The All Convolutional Net
- Very Deep Convolutional Networks for Large-Scale Image Recognition
- Very Deep Multilingual Convolutional Neural Networks for LVCSR
- Network In Network
- Going Deeper with Convolutions ([github.com/google/inception](https://github.com/google/inception))
- Convolutional Networks on Graphs for Learning Molecular Fingerprints ([github.com/HIPS/neural-fingerprint](https://github.com/HIPS/neural-fingerprint))
- Deep Learning for Single-View Instance Recognition
- Learning to Generate Chairs with Convolutional Neural Networks ([github.com/stokasto/caffe/tree/chairs\\_deconv](https://github.com/stokasto/caffe/tree/chairs_deconv))
- Deep Convolutional Inverse Graphics Network
- Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks
- Long-term Recurrent Convolutional Networks for Visual Recognition and Description
- A Machine Learning Approach for Filtering Monte Carlo Noise
- Image Super-Resolution Using Deep Convolutional Networks
- Learning to Deblur
- Monocular Object Instance Segmentation and Depth Ordering with CNNs
- FlowNet: Learning Optical Flow with Convolutional Networks
- DeepStereo: Learning to Predict New Views from the World's Imagery
- Deep convolutional filter banks for texture recognition and segmentation
- FaceNet: A Unified Embedding for Face Recognition and Clustering ([github.com/cmusatyalab/openface](https://github.com/cmusatyalab/openface))
- DeepFace: Closing the Gap to Human-Level Performance in Face Verification



- 
- Deep Karaoke: Extracting Vocals from Musical Mixtures Using a Convolutional Deep Neural Network
  - 3D ConvNets with Optical Flow Based Regularization
  - DeepPose: Human Pose Estimation via Deep Neural Networks
  - Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification
  - Rotation-invariant convolutional neural networks for galaxy morphology prediction
  - Deep Fried Convnets
  - Fractional Max-Pooling
  - Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps
  - Learning FRAME Models Using CNN Filters for Knowledge Visualization (code)
  - Invariant backpropagation: how to train a transformation-invariant neural network
  - Recommending music on Spotify with deep learning
  - Conv Nets: A Modular Perspective
  - Learning 3D Shape (1) ([github.com/danfischetti/shape-classifier](https://github.com/danfischetti/shape-classifier))

## **Recurrent Neural Networks**

- Unitary Evolution Recurrent Neural Networks
- Regularizing RNNs by Stabilizing Activations
- Training recurrent networks online without backtracking
- Modeling sequential data using higher-order relational features and predictive training ([github.com/memisevic/grammar-cells](https://github.com/memisevic/grammar-cells))
- Recurrent Neural Network Regularization
- How to Construct Deep Recurrent Neural Networks
- DAG-Recurrent Neural Networks For Scene Labeling
- Long Short-Term Memory (<ftp://ftp.idsia.ch/pub/juergen/lstm.pdf>)
  - LSTM: A Search Space Odyssey
  - Grid Long Short-Term Memory
  - Depth-Gated LSTM
- Learning Longer Memory in Recurrent Neural Networks
- A Simple Way to Initialize Recurrent Networks of Rectified Linear Units
- A Clockwork RNN
- DRAW: A Recurrent Neural Network For Image Generation
- Gated Feedback Recurrent Neural Networks
- A Recurrent Latent Variable Model for Sequential Data
- ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks

- 
- Translating Videos to Natural Language Using Deep Recurrent Neural Networks
  - Unsupervised Learning of Video Representations using LSTMs
  - Visualizing and Understanding Recurrent Networks
  - Advances in Optimizing Recurrent Networks
  - Learning Stochastic Recurrent Networks
  - Understanding LSTM Networks
  - Optimizing RNN performance
  - Mastering the Game of Go with Deep Neural Networks and Tree Search

### **Convolutional Recurrent Neural Networks**

- Recurrent Spatial Transformer Networks ([github.com/skaae/recurrent-spatial-transformer-code](https://github.com/skaae/recurrent-spatial-transformer-code))
- Recurrent Models of Visual Attention
  - Multiple Object Recognition with Visual Attention
- Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting
- Describing Multimedia Content using Attention-based Encoder–Decoder Networks

### **Adversarial Neural Networks**

- Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
- Unsupervised and Semi-supervised Learning with Categorical Generative Adversarial Networks
- Adversarial Autoencoders

### **Autoencoders**

- Correlational Neural Networks
- Optimizing Neural Networks that Generate Images ([github.com/mrkulk/Unsupervised-Capsule-Network](https://github.com/mrkulk/Unsupervised-Capsule-Network))
- Auto-Encoding Variational Bayes
- Analyzing noise in autoencoders and deep networks
- MADE: Masked Autoencoder for Distribution Estimation ([github.com/mgermain/MADE](https://github.com/mgermain/MADE))
- Winner-Take-All Autoencoders ([github.com/stephenbalaban/convnet](https://github.com/stephenbalaban/convnet))
- k-Sparse Autoencoders ([github.com/stephenbalaban/convnet](https://github.com/stephenbalaban/convnet))
- Zero-bias autoencoders and the benefits of co-adapting features

- 
- Importance Weighted Autoencoders ([github.com/yburda/iwae](https://github.com/yburda/iwae))
  - Generalized Denoising Auto-Encoders as Generative Models
  - Marginalized Denoising Auto-encoders for Nonlinear Representations
    - Marginalized Denoising Autoencoders for Domain Adaptation
  - Real-time Hebbian Learning from Autoencoder Features for Control Tasks
  - Procedural Modeling Using Autoencoder Networks (pdf) ([youtu.be/wl3h4S1g2u4](https://youtu.be/wl3h4S1g2u4))
  - Is Joint Training Better for Deep Auto-Encoders?
  - Towards universal neural nets: Gibbs machines and ACE
  - Transforming Auto-encoders
  - Discovering Hidden Factors of Variation in Deep Networks

## **Restricted Boltzmann Machines**

- The wake-sleep algorithm for unsupervised neural networks
  - A simple algorithm that discovers efficient perceptual codes
  - Reweighted Wake-Sleep
- An Infinite Restricted Boltzmann Machine
- Quantum Inspired Training for Boltzmann Machines
- Training Bidirectional Helmholtz Machines

## **Biologically Plausible Learning**

- How Auto-Encoders Could Provide Credit Assignment in Deep Networks via Target Propagation
  - Difference Target Propagation
  - Towards Biologically Plausible Deep Learning
- How Important is Weight Symmetry in Backpropagation?
- Random feedback weights support learning in deep neural networks

## **Supervised Learning**

- Fast Label Embeddings via Randomized Linear Algebra
  - Fast Label Embeddings for Extremely Large Output Spaces
- Locally Non-linear Embeddings for Extreme Multi-label Learning
- Efficient and Parsimonious Agnostic Active Learning

---

## Unsupervised Learning

- Towards Principled Unsupervised Learning
- Index-learning of unsupervised low dimensional embedding
- An Analysis of Unsupervised Pre-training in Light of Recent Advances ([github.com/ifp-uiuc/an-analysis-of-unsupervised-pre-training-iclr-2015](https://github.com/ifp-uiuc/an-analysis-of-unsupervised-pre-training-iclr-2015))
- Is Joint Training Better for Deep Auto-Encoders?
- Unsupervised Feature Learning from Temporal Data
- Learning to Linearize Under Uncertainty
- Semi-Supervised Learning with Ladder Network ([github.com/arasmus/ladder](https://github.com/arasmus/ladder))
  - Denoising autoencoder with modulated lateral connections learns invariant representations of natural images
  - Lateral Connections in Denoising Autoencoders Support Supervised Learning
- Semi-Supervised Learning with Deep Generative Models
- Rectified Factor Networks
- An Analysis of Single-Layer Networks in Unsupervised Feature Learning
- Deep Unsupervised Learning using Nonequilibrium Thermodynamics
- On-the-Fly Learning in a Perpetual Learning Machine

## Reinforcement Learning

- Variational Information Maximisation for Intrinsically Motivated Reinforcement Learning
- Prioritized Experience Replay
- Human-level control through deep reinforcement learning ([sites.google.com/a/deepmind.com/dqn](https://sites.google.com/a/deepmind.com/dqn))
- Playing Atari with Deep Reinforcement Learning
- Universal Value Function Approximators
- Giraffe: Using Deep Reinforcement Learning to Play Chess ([bitbucket.org/waterreaction/giraffe](https://bitbucket.org/waterreaction/giraffe))

## Theory

- The Human Kernel
- Why Neurons Have Thousands of Synapses, A Theory of Sequence Memory in Neocortex
- Deep Manifold Traversal: Changing Labels with Convolutional Features
- On the Expressive Power of Deep Learning: A Tensor Analysis
- $\ell_1$ -regularized Neural Networks are Improperly Learnable in Polynomial Time
- Provable approximation properties for deep neural networks
- Steps Toward Deep Kernel Methods from Infinite Neural Networks

- 
- On the Number of Linear Regions of Deep Neural Networks
  - On the saddle point problem for non-convex optimization
    - Identifying and attacking the saddle point problem in high-dimensional non-convex optimization
  - The Loss Surfaces of Multilayer Networks
  - Statistical mechanics of complex neural systems and high dimensional data
  - Qualitatively characterizing neural network optimization problems
  - An average-case depth hierarchy theorem for Boolean circuits
  - An exact mapping between the Variational Renormalization Group and Deep Learning
  - Why does Deep Learning work? - A perspective from Group Theory
  - A Group Theoretic Perspective on Unsupervised Deep Learning
  - Exact solutions to the nonlinear dynamics of learning in deep linear neural networks
  - On the Stability of Deep Networks
  - Over-Sampling in a Deep Neural Network
  - A theoretical argument for complex-valued convolutional networks
  - Spectral Representations for Convolutional Neural Networks
  - A Probabilistic Theory of Deep Learning
  - Deep Convolutional Networks on Graph-Structured Data ([github.com/mbhenaff/spectral-lib](https://github.com/mbhenaff/spectral-lib))
  - Learning with Group Invariant Features: A Kernel Perspective
  - Randomized algorithms for matrices and data
  - Calculus on Computational Graphs: Backpropagation
  - Understanding Convolutions
  - Groups & Group Convolutions
  - Neural Networks, Manifolds, and Topology
  - Neural Networks, Types, and Functional Programming
  - Causal Entropic Forces
  - On the Computability of AIXI
  - Physics, Topology, Logic and Computation: A Rosetta Stone

## **Quantum Computing**

- Analyzing Big Data with Dynamic Quantum Clustering
- Quantum algorithms for supervised and unsupervised machine learning
- Entanglement-Based Machine Learning on a Quantum Computer
- A quantum speedup in machine learning: Finding a N-bit Boolean function for a classification
- Application of Quantum Annealing to Training of Deep Neural Networks
- Quantum Deep Learning

- 
- Experimental Realization of Quantum Artificial Intelligence

## **Training Innovations**

- Adding Gradient Noise Improves Learning for Very Deep Networks
- Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)
- Net2Net: Accelerating Learning via Knowledge Transfer
- Learning the Architecture of Deep Neural Networks
- GradNets: Dynamic Interpolation Between Neural Architectures
- Reducing the Training Time of Neural Networks by Partitioning
- The Effects of Hyperparameters on SGD Training of Neural Networks
- Gradient-based Hyperparameter Optimization through Reversible Learning ([github.com/HIPS/hypergrad](https://github.com/HIPS/hypergrad))
- Learning Ordered Representations with Nested Dropout
- Learning Compact Convolutional Neural Networks with Nested Dropout
- Reducing Overfitting in Deep Networks by Decorrelating Representations
- Efficient Exact Gradient Update for training Deep Networks with Very Large Sparse Targets
- Efficient Per-Example Gradient Computations
- Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks
- Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift
  - Batch Normalized Recurrent Neural Networks
- Highway Networks
  - Training Very Deep Networks
- Random Walk Initialization for Training Very Deep Feedforward Networks
- Deeply-Supervised Nets
- Improving neural networks by preventing co-adaptation of feature detectors
  - Efficient batchwise dropout training using submatrices ([github.com/btgraham/Batchwise-Dropout](https://github.com/btgraham/Batchwise-Dropout))
  - Dropout Training for Support Vector Machines
  - Dropout as data augmentation
  - Partitioning Large Scale Deep Belief Networks Using Dropout
- Maxout Networks
- Regularization of Neural Networks using DropConnect
- Distilling the Knowledge in a Neural Network
- Domain-Adversarial Neural Networks
- Weight Uncertainty in Neural Networks

- 
- Notes on Noise Contrastive Estimation and Negative Sampling
  - Scale-invariant learning and convolutional networks
  - Empirical Evaluation of Rectified Activations in Convolutional Network
  - Deep Boosting ([github.com/google/deepboost](https://github.com/google/deepboost))
  - No Regret Bound for Extreme Bandits

## **Parallel Training**

- Scalable Distributed DNN Training Using Commodity GPU Cloud Computing
- AdaDelay: Delay Adaptive Distributed Stochastic Convex Optimization
- Large Scale Distributed Deep Networks
- HOGWILD!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent

## **Weight Compression**

- Tensorizing Neural Networks ([github.com/Bihaqo/TensorNet](https://github.com/Bihaqo/TensorNet)) ([github.com/vadimkantorov/tensornet.torch](https://github.com/vadimkantorov/tensornet.torch))
  - Tensorizing Neural Networks presentation
  - Tensor-Train Decomposition (pdf) ([github.com/oseledets/TT-Toolbox](https://github.com/oseledets/TT-Toolbox))
  - Spectral tensor-train decomposition
- Structured Transforms for Small-Footprint Deep Learning
- An exploration of parameter redundancy in deep networks with circulant projections
- A Deep Neural Network Compression Pipeline: Pruning, Quantization, Huffman Encoding
- Learning both Weights and Connections for Efficient Neural Networks
- Compressing Neural Networks with the Hashing Trick
- Flattened Convolutional Neural Networks for Feedforward Acceleration ([github.com/jhjin/flattened-cnn](https://github.com/jhjin/flattened-cnn))
- Predicting Parameters in Deep Learning

## **Numerical Precision**

- Neural Networks with Few Multiplications
- Deep Learning with Limited Numerical Precision
- Low precision storage for deep learning
- 1-Bit Stochastic Gradient Descent and Application to Data-Parallel Distributed Training of Speech DNNs

---

## Numerical Optimization

- Recursive Decomposition for Nonconvex Optimization ([github.com/afriesen/rdis](https://github.com/afriesen/rdis))
  - Recursive Decomposition for Nonconvex Optimization: Supplementary Material
- Beating the Perils of Non-Convexity: Guaranteed Training of Neural Networks using Tensor Methods
- Adapting Resilient Propagation for Deep Learning
- Accelerating Stochastic Gradient Descent via Online Learning to Sample
- Preconditioned Spectral Descent for Deep Learning
  - Preconditioned Spectral Descent for Deep Learning: Supplemental Material
- Probabilistic Backpropagation for Scalable Learning of Bayesian Neural Networks
- Beyond Convexity: Stochastic Quasi-Convex Optimization
- Graphical Newton
- Gradient Estimation Using Stochastic Computation Graphs
- Equilibrated adaptive learning rates for non-convex optimization
- Path-SGD: Path-Normalized Optimization in Deep Neural Networks
- Deep learning via Hessian-free optimization
- On the importance of initialization and momentum in deep learning
- Adaptive Subgradient Methods for Online Learning and Stochastic Optimization
- ADADELTA: An Adaptive Learning Rate Method
- ADASECANT: Robust Adaptive Secant Method for Stochastic Gradient
- Adam: A Method for Stochastic Optimization
- A sufficient and necessary condition for global optimization
  - A Level-Value Estimation Method and Stochastic Implementation for Global Optimization
- Unit Tests for Stochastic Optimization
- A\* Sampling
- Solving Random Quadratic Systems of Equations Is Nearly as Easy as Solving Linear Systems
- When Are Nonconvex Problems Not Scary?
- Automatic differentiation in machine learning: a survey

## Motion Planning

- Interactive Control of Diverse Complex Characters with Neural Networks (video)
- Continuous control with deep reinforcement learning
- Continuous Character Control with Low-Dimensional Embeddings



- 
- Supersizing Self-supervision: Learning to Grasp from 50K Tries and 700 Robot Hours ([youtu.be/oSqHc0nLkm8](https://youtu.be/oSqHc0nLkm8))
  - End-to-End Training of Deep Visuomotor Policies ([youtu.be/Q4bMcUk6pcw](https://youtu.be/Q4bMcUk6pcw))
  - Deep Spatial Autoencoders for Visuomotor Learning([youtu.be/TsPpoxKST2A](https://youtu.be/TsPpoxKST2A))
  - From Pixels to Torques: Policy Learning with Deep Dynamical Models (thesis) ([github.com/iassael/torch-ddcnn](https://github.com/iassael/torch-ddcnn))
  - Sampling-based Algorithms for Optimal Motion Planning ([youtu.be/r34XWEZ41HA](https://youtu.be/r34XWEZ41HA))
    - Informed RRT\*: Optimal Sampling-based Path Planning Focused via Direct Sampling of an Admissible Ellipsoidal Heuristic ([youtu.be/nsl-5MZfwu4](https://youtu.be/nsl-5MZfwu4))
    - Batch Informed Trees (BIT\*): Sampling-based Optimal Planning via the Heuristically Guided Search of Implicit Random Geometric Graphs ([youtu.be/TQloCC48gp4](https://youtu.be/TQloCC48gp4)) ([github.com/utiasASRL/batch-informed-trees](https://github.com/utiasASRL/batch-informed-trees))
  - Planning biped locomotion using motion capture data and probabilistic roadmaps ([youtu.be/cKrcjrdnD-M](https://youtu.be/cKrcjrdnD-M))
  - Stability of Surface Contacts for Humanoid Robots: Closed-Form Formulae of the Contact Wrench Cone for Rectangular Support Areas ([github.com/Tastalian/surface-contacts-icra-2015](https://github.com/Tastalian/surface-contacts-icra-2015))

## Simulation

- Data-Driven Fluid Simulations using Regression Forests ([vimeo.com/144267433](https://vimeo.com/144267433)) ([vimeo.com/144266101](https://vimeo.com/144266101))

## Hardware

- Towards Trainable Media: Using Waves for Neural Network-Style Training
- Random Projections through multiple optical scattering: Approximating kernels at the speed of light
- VLSI Implementation of Deep Neural Network Using Integral Stochastic Computing
- Training and operation of an integrated neuromorphic network based on metal-oxide memristors
- AHaH Computing–From Metastable Switches to Attractors to Machine Learning
- Finding a roadmap to achieve large neuromorphic hardware systems

## Cognitive Architectures

- A Large-Scale Model of the Functioning Brain
  - How to Build a Brain: A Neural Architecture for Biological Cognition

- 
- Derivation of a novel efficient supervised learning algorithm from cortical-subcortical loops
  - A Minimal Architecture for General Cognition ([github.com/mikegashler/manic](https://github.com/mikegashler/manic))

## **Computational Creativity**

- Inceptionism: Going Deeper into Neural Networks
  - DeepDream - a code example for visualizing Neural Networks ([github.com/google/deepdream](https://github.com/google/deepdream))
- A Neural Algorithm of Artistic Style
- The Unreasonable Effectiveness of Recurrent Neural Networks ([github.com/karpathy/char-rnn](https://github.com/karpathy/char-rnn))
- GRUV: Algorithmic Music Generation using Recurrent Neural Networks ([github.com/MattVitelli/GRUV](https://github.com/MattVitelli/GRUV))
- Composing Music With Recurrent Neural Networks ([github.com/hexahedria/biaxial-rnn-music-composition](https://github.com/hexahedria/biaxial-rnn-music-composition))

## **Cryptography**

- Crypto-Nets: Neural Networks over Encrypted Data

## **Distributed Computing**

- Dimension Independent Similarity Computation
  - Dimension Independent Matrix Square using MapReduce
  - All-pairs similarity via DIMSUM
- A Fast, Minimal Memory, Consistent Hash Algorithm

## **Clustering**

- Convolutional Clustering for Unsupervised Learning
- Deep clustering: Discriminative embeddings for segmentation and separation
- Clustering is Easy When ....What?
- Clustering by fast search and find of density peaks