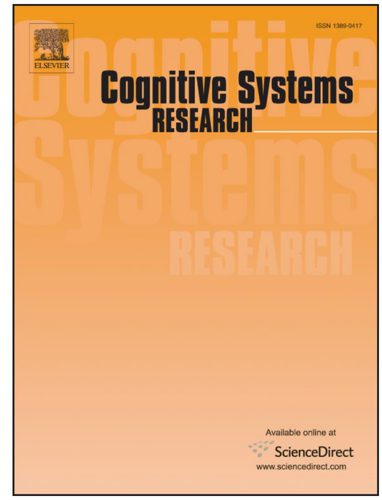
Accepted Manuscript

Deep Learning: Evolution and Expansion

Ritika Wason



PII: S1389-0417(17)30354-6
DOI: https://doi.org/10.1016/j.cogsys.2018.08.023
Reference: COGSYS 709

To appear in: *Cognitive Systems Research*

Received Date: 22 December 2017
Revised Date: 3 April 2018
Accepted Date: 24 August 2018

Please cite this article as: Wason, R., Deep Learning: Evolution and Expansion, *Cognitive Systems Research* (2018), doi: https://doi.org/10.1016/j.cogsys.2018.08.023

Deep Learning: Evolution and Expansion

Dr. Ritika Wason
Associate Professor, Bharati Vidyapeeth's Institute of Computer Applications and Management
rit_2282@yahoo.co.in

Deep Learning: Evolution and Expansion

Abstract
This paper historically attempts to map the significant success of deep neural networks in notably varied classification problems and application domains with near human-level performance. The paper also addresses the various doubts surrounding the acceptance of deep learning as a science of future. The manuscript attempts to unveil the hidden capabilities of deep neural networks in enabling machines perform the human way tasks which can be learned through what we call observation and experience.

Keywords
Deep Learning; Multi-Layer Neural Networks;

1. INTRODUCTION

Enabling computing systems to what human's term intelligence has been researcher's aspiration for over half a century. Late 1950's Rosenblatt et. al. aimed to create some brain analogue useful for analytical tasks[1]. This marked the inception of the research in the field of enabling machines to learn and classify the human way. The primary requirement of any such system shall be twofold: firstly, the ability to recognize and process complex patterns of related information from every dimension just as the human brain. Secondly, to achieve human-like intelligence a machine shall require access to large volume of information [2]. A system that could operate with some of the above notions, named perceptron[1] was suggested in [1]. This perceptron went on to be the basis for creating multi-layer learning networks which have formed the basis of what is popularly called Deep Learning[2], [3].

Deep Learning has lately emerged as the science of choice to learn convoluted structures in large, real-world data sets. Heavily researched in academia to study intelligence and industry to create intelligent systems, it applies back-propagation algorithm to determine how a machine should alter its parameters to correctly compute the output in each layer from the previous layer. The technology has succeeded in making breakthrough impact in varied domains like image, video, speech and audio processing[4]. At its simplest, deep learning is an application of neural networks with several layers of nodes(4 or more) between input and output[3].

Traditional machine learning mechanisms had confined ability of processing real data in its actual raw form. Hence, constructing a feature extractor involved decades of hard work and expertise to transform raw data to its suitable representation or feature vector from which the learner machine could classify patterns in the input[4].The multiple layers between input and output are used to perform feature identification and processing through a series of stages like the human brain[3].

Deep Learning methods are a kind of representation learning methods where non-linear modules are organized into multiple layers to transform the representation at a lower layer to a slightly higher abstract level. Such compositions of transformations are capable of learning complex functions[4].The noticeable feature of deep learning is that the feature layers are not designed or determined by humans but are learned automatically through some generalised learning procedure[2]. Through this ability of itself, deep learning has achieved significant success over counterpart machine learning and AI technologies in varied fields of human endeavour[3], [5].

In this paper, we aim to present a comprehensive survey of the growth and expansion of deep learning. The rest of the paper is organized as follows. Section II highlights the current research applications of deep learning. It is important to gain an insight into how one unified science has succeeded in expanding itself to varied domains utilizing varied volume, velocity and veracity of data. Section III highlights the characteristics of deep learning algorithms that have led to their widespread applications. Section IV, discusses the current challenges faced by Deep Learning. Section V concludes with an analysis of this whole discussion.

2. THE DEEP LEARNING REVOLUTION

In the last decade deep learning has been successfully applied to various technologies and mechanisms that require large volumes of digital data for training and providing useful information[6]. Table 1 below lists some considerable applications of deep learning:

Table 1: Deep Learning Application Domains

S. No	Ref.	Application	Description	Application Domain
1	[7] 1989	**Use of Multi-Layered Networks for Coding Speech with Phonetic Features**	A novel method combining expertise of neural networks with speech recognition is used to realize a speech recognition system.	Speaker-Independent Speech Recognition
2	[8] 1992	**Local Feedback Multilayered Networks**	Analyzes the limitations and characteristics of a local feedback multi-layered network with feedback connections allowed only from neurons to itself.	Recurrent Networks
3	[9] 1998	**Gradient-Based Learning Applied to Document Recognition**	Multilayer neural network trained through backpropogation algorithm applied to create a complex decision surface to classify high-dimensional patterns like handwritten characters.	Handwriting Recognition
4	[10] 2006	**A Fast Learning Algorithm for Deep Belief Nets**	Complimentary Priors applied to eliminate the explaining effects that make presumption complicated in densely linked belief nets with numerous hidden layers.	Deep-Belief Network
5	[11] 2006	**Reducing the Dimensionality of Data with Neural Networks**	High –Dimensional Data converted to low-dimensional codes through training a MNN.	Autoencoders
6	[12] 2010	**Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion**	Innovative framework of comprehending a deep neural network through layers of denoising encoders trained to denoise corrupted versions of their inputs	Autoencoders
7	[13] 2012	**ImageNet Classification with Deep Convolutional Neural Networks**	Breakthrough testimony that applied convolutional nets to halve the error rate for object recognition, resulting in brisk implementation of deep learning by the computer vision commune.	Object Classification
8	[14] 2012	**Large Scale Distributed Deep Networks**	DistBelief software framework developed to train large, distributed models. Obtains significant results about large-scale non-convex optimizations.	Speech Recognition service at both moderate and large-scale level
9	[15] 2012	**Deep Neural Networks for Acoustic Modeling in Speech Recognition**	Utilizes a feed-forward neural network for speech recognition.	Acoustic Modeling
10	[16] 2012	**Building High-level Features Using Large Scale Unsupervised Learning**	Prepares a 9- layered locally connected thin autoencoder with pooling and local contrast normalization on a large dataset of images.	High Level, Class-Specific Feature Detectors
11	[17] 2012	**Improving Neural networks by preventing**	Reduces Overfitting on large feed-forward neural networks through randomly omitting	Speech and Object Recognition Model

		Co-Adaptation of Feature Detectors	half feature detectors on each training set. Overcomes complex co-adaptations for many routine tasks in speech and object recognition.	Development
12	[18] 2013	**Speech-Recognition with Deep Recurrent Neural Networks**	Train Deep Long Short-term Memory RNN on the TIMIT phoneme recognition benchmark for speech recognition.	Speech Recognition
13	[19] 2013	**Visualizing and Understanding Convolutional Neural network**	Network by Zeiler et. al., winner of ILSVRC 2013 achieved top 5 error rate of 11.2%. More of AlexNet fine tuning to improve performance. Examined different feature activations and their relations to the input space.	Deconvolutional network
14	[20] 2013	**Learning Hierarchical Features for Scene Labelling**	Applied convolutional nets to scene label through multi-scale convolutional nets. The method succeeded in extracting an optimal set of segmentation components that could best explain a scene.	Scene Labelling
15	[21] 2013	**Generating Sequences with Recurrent Neural Networks**	Long-Short Term Memory Recurrent Neural Networks are applied to produce complex sequences with long range structures by predicting one data point at a time.	Handwriting Synthesis
16	[22] 2013	**Multiframe Deep Neural Networks for Acoustic Modeling**	Achieves reduction of the neural network computational cost using speech signal stationarity for tying neural network parameters across frames.	Acoustic Modeling
17	[23] 2013	**On the Importance of Initialization and Momentum in Deep learning**	Stochastic Gradient Descent with Momentum applied to train DNNs as well as RNNs on datasets with long term dependencies to achieve considerable performance.	Model Optimization
18	[24] 2014	**Learning Phrase Representations using RNN encoder-decoder for statistical machine translation**	Utilizes two recurrent neural networks to maximize the conditional probability of a target sequence given a source sequence.	Sequence-to-Sequence Modelling
19	[25] 2014	**Dropout: A Simple Way to Prevent Neural Networks from Overfitting**	Overfitting is a grave concern in deep neural nets with huge amount of parameters. Dropout technique uses random unit dropping during training to avoid co-adaption. These resultant thinned networks optimize neural net performance on supervised learning tasks in vision, speech recognition, document categorization as well as computational biology.	Model Optimizations
20	[26] 2014	**Very Deep Convolution Networks for Large-Scale Image Recognition**	19-layer CNN using 3X3 filters with stride and pad of 1 as compared to 11X11 of Alexnet and 7X7 of ZFNet.Reinforced notion that CNNs apply deep-layered network .	ImageNet
21	[27] 2014	**One BillionWord Benchmark for Measuring Progress in Statistical Language Modeling**	Proposes a novel benchmark for estimating progress in statistical language modelling.	Language Modeling
22	[28] 2014	**Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation**	Proposes novel hybrid framework through a deep convolutional network and Markov Random Field. The framework is applied for human pose evaluation in monocular images.	Human Pose Identification

23	[29] 2014	**Neural Turing Machines**	Expand neural network capabilities by coupling them to external memory resources through attentional processes. System is analogous to a Turing machine but end-to-end differentiable. Efficient training with gradient descent possible.	Basic Future Computer Prototype
24	[30] 2014	**Going Deeper with Convolutions**	GoogLeNet a 22-layer CNN, winner of ILSVRC 2014 with 6.7% error rate. Utilized 9 inception modules with over 100 layers. No sequential stacking of layers.	ImageNet
25	[31] 2015	**DRAW: A Recurrent Neural Network for Image Generation**	Deep Recurrent Attentive Writer (DRAW) utilizing spatial attention mechanism to mimic the foveation of human eye with a sequential variational auto-encoding framework to construct complex images.	Deep Image Generative Model
26	[32] 2015	**End-to-End Memory Networks**	Neural Network with recurrent attention model over a large external memory. A type of memory network trained end-to-end possibly without supervision. Can be applied for a diverse variety of tasks from question answering to language modelling.	
27	[33] 2015	**A Neural Conversational Model**	Approach for conversational modelling based on sequence to sequence framework. Used to predict the next sentence in a conversation.	Natural Language Interception
28	[34] 2015	**Deep Residual Learning for Image Recognition**	Microsoft ResNet, a 152-layer network architecture winner of ILSVRC 2015 with error rate 3.6%. An ultra-deep network spatial size compression from 224X224 to 56X56 right after first two layers.	ImageNet
29	[35] 2015	**Deep Visual-Semantic Alignments for Generating Image Descriptions**	Aligns natural language descriptions of images and their regions. Learns inter-modal correspondence between language and visual data through convolutional neural networks over image parts and bidirectional recurrent neural networks over sentences.	Inter-modal correspondence between Language and Visual Data
30	[36] 2015	**Algorithmic Composition of Melodies with Deep Recurrent Neural Networks**	Artificial neural networks are trained on a large range of melodies to be capable of reproducing long-range temporal dependencies typical of music.	Automated Music Composers
31	[37] 2015	**Deep Speech 2: End-to End Speech Recognition in English and Mandarin**	End-to-End Deep learning approach applied to recognize either English/Mandarin Chinese speech. Applied HPC techniques to iterate quickly to identify superior architecture and algorithms.	Baidu Speech Recognition System
32	[38] 2015	**Net2Net: Accelerating Learning via Knowledge Transfer**	Proposes techniques for rapidly transferring information from one neural net to another in order to accelerate training of a significantly large neural net.	Model Improvement
33	[39] 2015	**Pointer Networks**	Novel NN architecture to learn the conditional probability of an output sequence on elements that are discrete tokens to positions in an input sequence.	Model Optimization
34	[40] 2016	**Achieving Human Parity in Convolutional Speech Recognition**	Computes the error rate on NIST 2000 test set to conform if automated system has achieved human parity. First recognition of human parity for conversational speech.	Speech recognition
35	[41] 2016	**Decoupled Neural Interfaces using**	Forward data propagation amalgamated with back propagating error signal is required for	Model Improvement

		Synthetic Ingredients	training directed neural networks for weight updates. In such scenario, network layers remain locked till weight updation is completed. Work proposes a model for future computation of network graph.	
36	[42] 2016	**Binarized Neural Networks: Training Neural Networks with Weights and Activation Constrained to +1 and -1**	Proposes a scheme to train a BNN using binary weights and activations at runtime and when computing the parameters gradient at train-time.Torch7 and Theano framework used to train BNN on MNIST, CIFAR-10 and SVHN.	Model Improvement
37	[43] 2016	**Network Morphism**	Analyzes morphing a well-trained neural network to a novel one in order to preserve its network function. Enables child network to inherit parent knowledge and grow to be a more powerful one.	Model improvement
38	[44] 2016	**Pixel Recurrent Neural Networks**	DNN that sequentially predicts raw pixels in an image along the two-dimensional space.	Image Generation
39	[45] 2016	**Deep Learning for Social Media Analytics in Crises Situations**	Proposes application of Deep Learning to extract text related features and patterns from crises related social media posts for use in handling crises situations.	Social Media Analytics
40	[46] 2016	**Deep Learning Techniques in Big Data Analytics**	Investigates role of Deep Learning in addressing Big Data Analytics issues like, mining composite patterns from substantial volumes of data, semantic indexing, data tagging, rapid information reclamation, and shortening discriminative tasks.	Big data Analytics
41	[47] 2016	**Mobile Big Data Analytics using Deep Learning and Apache spark**	Executes distributed deep learning as an iterative MapReduce on multiple Spark co-workers. Each worker is trained on a partial model and a master deep model is then built through parameter averaging of all partial models.	Big data Analytics
42	[48] 2016	**Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs**	Applies deep learning to propose an algorithm for automated detection of diabetic retinopathy and diabetic retinal fundus photographs.	HealthCare
43	[49] 2017	**Synthesizing Obama: Learning Lip Sync from Audio Output Obama Video**	A Recurrent Neural Network trained to map raw audio characteristics to mouth shapes. High Quality video of President Barack Obama generated from his audio.	Audio to Visual Speech Synthesis
44	[50] 2017	**Enabling large-scale viscoelastic calculations via neural network acceleration**	Trains a deep neural network with a computationally able illustration of viscoelastic solutions, at any time, location, and for a huge array of rheological structures.	Viscoelastic Calculations
45	[51] 2017	**Using Deep Learning and Google Street View to Estimate the De- mographic Makeup of the US**	Mechanism to determine socio-economic trends from almost 50 million images of street scenes collected from across 200 American cities.	Demographic Prediction
46	[52] 2017	**Mastering Chess and Shogi by Self-Play with a General Reinforcement**	Proposes AlphaZero algorithm that can achieve, tabula rasa, superhuman performance in many challenging games like Chess, Shogi and Go.	Game Playing

		Learning Algorithm		
47	[53] 2017	**CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning**	Algorithm applying 121-layer CNN to detect pneumonia from chest x-rays proposed. Algorithm tested and found to exceed average radiologist performance on pneumonia detection.	Detection
48	[54] 2017	**Improving Palliative Care with Deep Learning**	Deep Learning amalgamated with Electronic Health Record Data to guess patients that are to be benefitted by palliative care. Enables palliative care team to proactively reach out to patients.	Healthcare monitoring
49	[55] 2017	**VoxelNet: End-to-End Learning for Point Cloud-Based 3D Object Detection**	Generic 3D detection network for efficient 3D detection of pedestrians and cyclists.	Object Detection
50	[56] 2017	**Asymmetric Actor Critic for Image-Based robot Learning**	Deep Reinforcement Learning offers significant challenges in robotics. Actor critic training algorithm applied to train robots on tasks like picking, pushing and moving a block.	Robotics
51	[57] 2017	**A Hybrid DSP/DEEP Learning Approach to Real-Time Full-Band Speech Enhancement**	A hybrid DSP/deep learning approach proposed for noise suppression	Speech Signal Processing
52	[58] 2017	**Globally Normalized Reader**	Extractive Question Answering used as an iterative search problem to reduce the space of each search step. The representation is proved to be viable and more learning efficient.	
53	[59] 2017	**Dermatologist-level Classification of Skin Cancer with Deep Neural Networks**	Applies single CNN for classification of skin lesions. The tests were performed for identification of most common cancers to the most malignant cancers. CNN achieved performance on par with all experts across both cases.	Diagnostic Healthcare
54	[60] 2018	**Emergent Translation in Multi-Agent Communication**	Communication game where two native speakers of a language jointly learn to solve a visual referential problem.	Robotics

Table 1 above is not a comprehensive summary of the different applications of deep learning. However, it is a sneak peek into the different domains into which deep learning has been successfully applied to extract useful insights from the large volumes of available data. What is interesting to note from the above table is the fact that deep learning has been successfully applied to a variety of domains only in a very recent span of let's say the past five-six years. Before2006, though deep learning existed researchers were sceptical about its viability and success[61]–[63].

3. DEEP LEARNING EXPANSION

The rapidly emerging science of deep learning is an excellent resource for many powerful, future strategies and domains. As can be noticed from Table 1 in the previous section, many strategies and technologies have effectively used deep learning for attaining volume, variety and veracity in extracting useful information from data and making a machine self-reliant to mimic some human task.

However, the scene for deep architectures has never been the same since inception. Notably before 2006, attempts at training such architectures failed. In 2006, certain considerable work on deep belief nets spearheaded by Hinton [63], [64], [61]changed the scenario. 2006 thus was a remarkable year in the field of deep learning as since then a plethora of work has been undertaken in the domain[2]. The strong points of deep learning that have further promoted its widespread application in a number of domains are:

i. They can work in a noisy environment to filter and extract information hidden within noise.
ii. The algorithms train through examples to identify patterns and integrate the information into some sort of visual analytics displays.
iii. Deep Learning Algorithms can apply discrimination to data to reveal patterns and valuable information.
iv. Can easily classify unstructured as well as structured data through strategies like Deep Belief Method (DBM) or Convolutional Neural Networks (CNN) etc [6]. It applies a combination of un-supervised training and supervised fine-tuning to construct models.
v. They mimic the human brain through artificial neural networks and progressively learn how to solve a given problem the human way.

4. CHALLENGES OF DEEP LEARNING

The present decade is a remarkable one for the human race. Not only are we progressing towards a smarter universe we are also generating and accumulating vast data repository whose number shall soon outgrow our computational range[65]. In such a scenario to avoid the data chaos deep learning is being projected as the science capable of handling this vast data and putting it to effective use.

Despite its widespread application, Deep Learning science is still in its nascent stages. However, it has quickly become a primary research interest for developing and realizing smart, intelligent, autonomous machines. As per Table 1 above, almost all AI applications irrespective of the field or domain are driven by Deep Learning. However, the science is not all that rosy and still far from being flawless. There are still significant challenges of deep learning that we need to overcome effectively[6], [66]–[69].

i. Big Data: As per National Security Agency [6] world processes more than 1.8 Exabytes of data per day. Further the amount of data being generated by the human race across the globe is expanding at an exponential rate each day. Deep Learning when used with Big Data etc has the potential to manage and analyze this large amount of supervised or unsupervised information in a short time. However, training deep learning algorithms on such massive amounts of data with a single processor is a challenging task. Hence, clusters of CPU or graphic processing unit (GPUs) have been applied to increase the training speed of deep learning algorithms. However, though many optimizations have been achieved, the complete process is still time consuming and requires high data processing capabilities.

ii. Massive Datasets[67], [70]:Deep learning has found successful application in varied domains like computer vision, natural language processing, robotics etc. However, notably the number of data samples for an efficient learning should be 10X the number of parameters in deepnet. Hence, large volume of data is a pre-requisite for the success of such networks.

iii. Neural Network Over fitting: There can be a significant difference in error reported in training data set and error encountered in real data set. This can be a common issue in large networks with multiple parameters thus affecting model efficacy.

iv. Hyper-Parameter Optimization: there are certain parameters whose values are defined prior to commencement of learning. Such parameters are called as hyper-parameters. A minor change in value of these parameters can lead to a significant change in model performance.

v. Trial-and-Error Learning (How much depth is sufficient?): Neural networks by nature are a black box as their operations are opaque to the humans[70]. Deep Learning creates computational models composed of multiple processing layers to learn data representation through multi-level abstraction. However, notably in this case the layer of abstractions are not decided by some human engineer but are learned from data through some generic learning algorithm. Hence, deep nets process inputs in a layered, non-linear mechanism to initialize hidden layer nodes to learn generic structures and representations. These representations are then submitted to a supervised layer to fine tune the deep net through backpropogation algorithm towards optimized representations for desired task.

vi. Brittle Nature: Deep learning networks are brittle in the sense that a trained network can only perform on the task it is trained for and performs poorly on any new task.

vii. Ex post High-Dimensional Path Attribution: Deep Learning takes raw data as an input and the machine learns from it how to achieve the desirable outcome. This is a deeper challenge that involves linking series of actions over time and synthesizing them into useful lessons. This link between the attribution of actions (ex: hiding, movement etc) and outcomes is based on complex temporal relationships and objective functions.

Many different systems are being worked upon to overcome the above limitations [70]–[73].

5. CONCLUSION

Deep Learning science has been existing for decades and its genesis can be traced back to artificial neural networks (ANN) way back from the late 50's[1]. However, it was grounded in deep sceptics at inception. However, it overcame rose above all uncertainties to be a technology applied to many significant domains of human life. The tremendous progress made by deep systems in a short span of a decade proves beyond doubt that the impact of this science is actually not overhyped[74]. However, the science is still young and there are a number of challenges to be overcome. Expecting deep learning combined with improved data processing being a solution to computers gaining generic human-like intelligence (human equivalent AI) is still a distant dream.

BIBLIOGRAPHY

[1] F. Rosenblatt, “The Perceptron - A Perceiving and Recognizing Automaton,” *Report 85, Cornell Aeronautical Laboratory*. pp. 460–1, 1957.

[2] Y. Bengio, “Learning Deep Architectures for AI,” *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[3] I. Arel, D. C. Rose, and T. P. Karnowski, “Deep Machine Learning - A New Frontier in Artificial Intelligence Research,” *IEEE Comput. Intell. Mag.*, vol. 5, no. 4, pp. 13–18, 2010.

[4] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[5] T. Chen, R. Xu, Y. He, and X. Wang, “Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN,” *Expert Syst. Appl.*, vol. 72, pp. 221–230, 2017.

[6] Xue-Wen Chen and Xiaotong Lin, “Big Data Deep Learning: Challenges and Perspectives,” *IEEE Access*, vol. 2, pp. 514–525, 2014.

[7] “Ð Ð □ ` a ä íUí Å m‡Yqu(ø … Ð¾P“f Y f P u T ™‡¦ R f ` (&.”

[8] P. Frasconi, M. Gori, and G. Soda, “Local Feedback Multilayered Networks,” *Neural Comput.*, vol. 4, no. 1, pp. 120–130, 1992.

[9] “lecun- at el - Gradient-based learning applied to document.”

[10] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets.,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–54, 2006.

[11] A. J. Holden, D. J. Robbins, W. J. Stewart, D. R. Smith, S. Schultz, M. Wegener, S. Linden, C. Hormann, C. Enkrich, M. Wegener, C. M. Soukoulis, S. Linden, D. Schurig, D. R. Smith, C. Enkrich, M. Wegener, C. M. Soukoulis, S. Linden, A. J. Taylor, C. Highstrete, M. Lee, R. D. Averitt, S. Schultz, P. Markos, C. M. Soukoulis, D. Mcpeake, S. A. Ramakrishna, J. B. Pendry, V. M. Shalaev, M. Maksimchuk, D. Umstadter, W. Chen, Y. R. Shen, and J. V Moloney, “Science,” vol. 313, no. July, pp. 504–507, 2006.

[12] P. Vincent PASCALVINCENT and H. Larochelle LAROCHEH, “Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol,” *J. Mach. Learn. Res.*, vol. 11, pp. 3371–3408, 2010.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.

[14] J. Dean, G. S. Corrado, R. Monga, K. Chen, M. Devin, Q. V Le, M. Z. Mao, M. A. Ranzato, A. Senior, P. Tucker, K. Yang, and A. Y. Ng, “Large Scale Distributed Deep Networks,” *NIPS 2012 Neural Inf. Process. Syst.*, pp. 1–11, 2012.

[15] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, “Deep Neural Networks for Acoustic Modeling in Speech Recognition,” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 82–97, 2012.

[16] Q. V. Le, M. A. Ranzato, M. Devin, G. S. Corrado, and A. Y. Ng, “Building High-level Features Using Large Scale Unsupervised Learning,” *Arxiv*, vol. 28, no. 4, pp. 61–76, 2012.

[17] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” pp. 1–18, 2012.

[18] A. Graves, A. Mohamed, and G. Hinton, “Speech Recognition with Deep Recurrent Neural Networks,” no. 3, 2013.

[19] M. D. Zeiler and R. Fergus, “Visualizing and Understanding Convolutional Networks arXiv:1311.2901v3 [cs.CV] 28 Nov 2013,” *Comput. Vision–ECCV 2014*, vol. 8689, pp. 818–833, 2014.

[20] M. Mohri and A. Rostamizadeh, “Perceptron Mistake Bounds,” *arXiv Prepr. arXiv1305.0208*, 2013.

[21] A. Graves, “Generating Sequences With Recurrent Neural Networks,” pp. 1–43, 2013.

[22] V. Vanhoucke, M. Devin, and G. Heigold, “Multiframe deep neural networks for acoustic modeling,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 7582–7585, 2013.

[23] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton, “On the importance of initialization and momentum in deep learning,” *Jmlr W&Cp*, vol. 28, no. 2010, pp. 1139–1147, 2013.

[24] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation,” 2014.

[25] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[26] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” pp. 1–14, 2014.

[27] C. Chelba, T. Mikolov, M. Schuster, Q. Ge, T. Brants, P. Koehn, and T. Robinson, “One billion word benchmark for measuring progress in statistical language modeling,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, pp. 2635–2639, 2014.

[28] J. Tompson, A. Jain, Y. LeCun, and C. Bregler, “Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation,” *Adv. Neural Inf. Process. Syst.*, pp. 1799--1807, 2014.

[29] A. Graves, G. Wayne, and I. Danihelka, “Neural Turing Machines,” pp. 1–26, 2014.

[30] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, C. Hill, and A. Arbor, “Going Deeper with Convolutions,” pp. 1–9, 2014.

[31] K. Gregor, I. Danihelka, A. Graves, and D. Wierstra, “DRAW: A Recurrent Neural Network For Image Generation,” *Icml-2015*, pp. 1–16, 2014.

[32] S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus, “End-To-End Memory Networks,” pp. 1–9, 2015.

[33] O. Vinyals and Q. Le, “A Neural Conversational Model,” vol. 37, 2015.

[34] S. Wu, S. Zhong, and Y. Liu, “Deep residual learning for image steganalysis,” *Multimed. Tools Appl.*, pp. 1–17, 2017.

[35] A. Karpathy and F. F. Li, “Deep visual-semantic alignments for generating image

descriptions," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2015.

[36] F. Colombo, S. P. Muscinelli, A. Seeholzer, J. Brea, and W. Gerstner, "Algorithmic Composition of Melodies with Deep Recurrent Neural Networks," 2016.

[37] D. Amodei, R. Anubhai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, J. Chen, M. Chrzanowski, A. Coates, G. Diamos, E. Elsen, J. Engel, L. Fan, C. Fougner, T. Han, A. Hannun, B. Jun, P. LeGresley, L. Lin, S. Narang, A. Ng, S. Ozair, R. Prenger, J. Raiman, S. Satheesh, D. Seetapun, S. Sengupta, Y. Wang, Z. Wang, C. Wang, B. Xiao, D. Yogatama, J. Zhan, and Z. Zhu, "Deep Speech 2: End-to-End Speech Recognition in English and Mandarin," pp. 1–28, 2015.

[38] T. Chen, I. Goodfellow, and J. Shlens, "Net2Net: Accelerating Learning via Knowledge Transfer," pp. 1–12, 2015.

[39] O. Vinyals, M. Fortunato, and N. Jaitly, "Pointer Networks," pp. 1–9, 2015.

[40] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "Achieving Human Parity in Conversational Speech Recognition," 2016.

[41] M. Jaderberg, W. M. Czarnecki, S. Osindero, O. Vinyals, A. Graves, D. Silver, and K. Kavukcuoglu, "Decoupled Neural Interfaces using Synthetic Gradients," vol. 1, 2016.

[42] M. Courbariaux, I. Hubara, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized Neural Networks: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1," 2016.

[43] T. Wei, C. Wang, Y. Rui, and C. W. Chen, "Network Morphism," 2016.

[44] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, "Pixel Recurrent Neural Networks," vol. 48, 2016.

[45] M. Ben Lazreg and O. Granmo, "Deep Learning for Social Media Analysis in Crises Situations," no. June, pp. 1–6, 2016.

[46] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagc, "Deep Learning Techniques in Big Data Analytics," in *Big Data Technologies and Applications*, Cham: Springer International Publishing, 2016, pp. 133–156.

[47] M. A. Alsheikh, D. Niyato, S. Lin, H. Tan, and Z. Han, "Mobile big data analytics using deep learning and apache spark," *IEEE Netw.*, vol. 30, no. 3, pp. 22–29, May 2016.

[48] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 22, pp. 2402–2410, 2016.

[49] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing Obama: Learning Lip Sync from Audio Output Obama Video," *ACM Trans. Graph. Artic.*, vol. 36, no. 95, 2017.

[50] P. M. R. DeVries, T. Ben Thompson, and B. J. Meade, "Enabling large-scale viscoelastic calculations via neural network acceleration," *Geophys. Res. Lett.*, vol. 44, no. 6, pp. 2662–2669, 2017.

[51] T. Gebru, J. Krause, Y. Wang, D. Chen, J. Deng, E. L. Aiden, and L. Fei-Fei, "Using Deep Learning and Google Street View to Estimate the Demographic Makeup of the US," pp. 1–41,

2017.

[52] D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, and D. Hassabis, “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm,” pp. 1–19, 2017.

[53] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, “CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,” 2017.

[54] A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, and N. H. Shah, “Improving Palliative Care with Deep Learning,” 2017.

[55] Y. Zhou and O. Tuzel, “VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection,” 2017.

[56] L. Pinto, M. Andrychowicz, P. Welinder, W. Zaremba, and P. Abbeel, “Asymmetric Actor Critic for Image-Based Robot Learning,” 2017.

[57] J.-M. Valin, “A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement,” 2017.

[58] J. Raiman and J. Miller, “Globally Normalized Reader,” 2017.

[59] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[60] J. Lee, K. Cho, J. Weston, and D. Kiela, “Emergent Translation in Multi-Agent Communication,” pp. 1–16, 2017.

[61] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy Layer-Wise Training of Deep Networks,” *Adv. Neural Inf. Process. Syst.*, vol. 19, no. 1, p. 153, 2007.

[62] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[63] M. A. Ranzato, C. Poultney, S. Chopra, and Y. Lecun, “Efficient Learning of Sparse Representations with an Energy-Based Model,” *Adv. Neural Inf. Process. Syst.*, vol. 19, pp. 1137–1134, 2007.

[64] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, *Continuous control with deep reinforcement learning*, vol. 2, no. 1. 2015.

[65] J. Gantz and D. Reinsel, “The Digital Universe Decade – Are You Ready?,” *Idc*, vol. 2009, no. May, p. 16, 2010.

[66] V. Dhar, “The Scope and Challenges for Deep Learning,” *Big Data*, vol. 3, no. 3, pp. 127–129, 2015.

[67] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, “Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.*, no. December 2016, pp. 1–11, 2017.

[68] P. Angelov and A. Sperduti, “Challenges in Deep Learning,” *Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn.*, no. April, pp. 27–29, 2016.

[69] G. Mesnil, Y. Dauphin, X. Glorot, S. Rifai, Y. Bengio, I. J. Goodfellow, E. Lavoie, X. Muller, G. Desjardins, D. Warde-Farley, P. Vincent, A. Courville, and J. Bergstra, “Unsupervised and

Transfer Learning Challenge: a Deep Learning approach," *JMLR W& CP Proc. Unsupervised Transf. Learn. Chall. Work.*, vol. 27, pp. 97–110, 2012.

[70] M. Garnelo, K. Arulkumaran, and M. Shanahan, "Towards Deep Symbolic Reinforcement Learning," pp. 1–13, 2016.

[71] A. A. Rusu, N. C. Rabinowitz, G. Desjardins, H. Soyer, J. Kirkpatrick, K. Kavukcuoglu, R. Pascanu, and R. Hadsell, "Progressive Neural Networks," 2016.

[72] R. M. Seraj, "Multi-task Learning," 2014.

[73] L. Kaiser, A. N. Gomez, N. Shazeer, A. Vaswani, N. Parmar, L. Jones, and J. Uszkoreit, "One Model To Learn Them All," 2017.

[74] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 427–436, 2015.