INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

THE IMPLICATIONS OF DEEP LEARNING ALGORITHMS AND TOOLS: A REVIEW

Deepali Vora[*1] & Dr Kamatchi Iyer[2]
[*1&2]Amity University, Mumbai

ABSTRACT

Deep Learning, viewed as a part of machine learning,is capable of learning the essence of data from a small data set by learning a nonlinear network structure.Deep Learning algorithms have shown wide use in speech recognition, image processing, video processing and has wide application in various areas which uses any of these techniques. Still many application areas are emerging and are untouched by Deep Learning Architectures. This paper gives an overview of various application areas where Deep Learning algorithms or variants are applied and proved beneficial. As well it gives a summary of the top five popular Deep Learning tools available for research and implementation of Deep Learning algorithms.

I.   INTRODUCTION

Learning is required where we cannot directly write a computer program to do a task, but the task is accomplished using the past experiences or history. Machine learning is nothing but writing an application which learns from past experiences. We require machine learning where we wish the system should adapt to changes in environment and act accordingly. Many such examples exist like conversion of speech into text, best path determination by the router based on quality of Service etc. Machine learning is that branch of Computer Science which focuses on development of programs that can learn by themselves. [1]

The digital data is increasing day by day thereby posing challenges to machine learning. The term Big Data has become buzzword because of its distinct characteristics as volume, veracity and velocity.The data collected today are complex, huge and unlabelled. To analyse such data simple machine learning algorithms are not sufficient. Deep Learning algorithms are one of the such avenues which helps in processing of complex,huge and unlabelled data [2]. Deep Learning was first time proposed by Hinton et al in 2006. It is a new field of machine learning. Algorithms in Deep Learning aim at learning the features from the hierarchies of features which exists at various levels of abstraction [3]; thus making the analysis of complex and huge data possible.

The paper is organised as: Section 2 gives an overview of Deep Learning and basic algorithms, Section 3 talks about the application area where these deep learning algorithms are widely used and tested. It talks about the algorithms used in these application areas. Section 4 gives an overview of the tools available for deep learning.

II.   DEEP LEARNING AND ALGORITHMS

Normal machine learning algorithms work on a Neural Network with 1 or 2 hidden layers. The network architecture in deep learning is created with the input layer, the hidden layers of which the number is normally 5 or 6 or even more, and the output layer. [2].
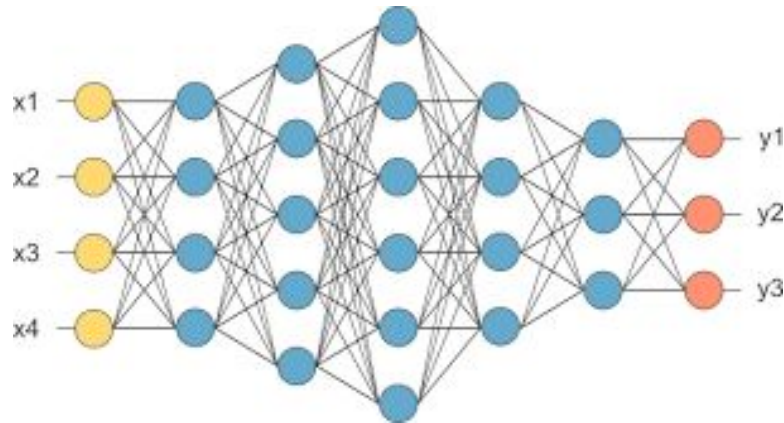
Fig1:A Deep Architecture

The network transforms low-level features of images into high level features which are more abstract and thus makes classification and prediction easier. Generally deep architectures have more units in hidden layers than normal shallow architectures.

Though the deep networks were available but were difficult to train. In 2007 a greedy layarwise training of deep network was proposed [4].This algorithms made deep architectures for use in real world scenarios.
Deep learning has numerous algorithms of machine learning. These algorithms attempts to model high-level abstractions in data. They create or design architectures which are composed of many non-linear transformations.Deep architectures can be modelled using any combinations of layers of network, but still it has set traditional algorithmssuch as Stacked AutoEncoder, Deep Boltzman Machines, Deep Convolutional Networks and Deep Belief Networks.

A. Stacked Auto Encoder

An autoencoder is an artificial 3 or more layer neural network. It is used for unsupervised learning and output units are connected back to input units. The main purpose of an autoencoder is to learn a representation (encoding) for a set of data, typically for the purpose of dimensionality reduction. The other variant, called as stacked denoising auto encoders, works on cleaning(de-noising) the partially corrupted output. This algorithm was introduced in 2008 by Vincent et al. [5]to provide a specific approach to good representation. If we can obtain output robustly from a corrupted input then it is termed as a good representation. Such a good representation is useful for recovering the corresponding clean input.

The autoencoder network is trained to get the correct output. Now this cleaned output is used as a input to a supervised machine learning algorithm for example; Support Vector Machine or Logistic Regression.
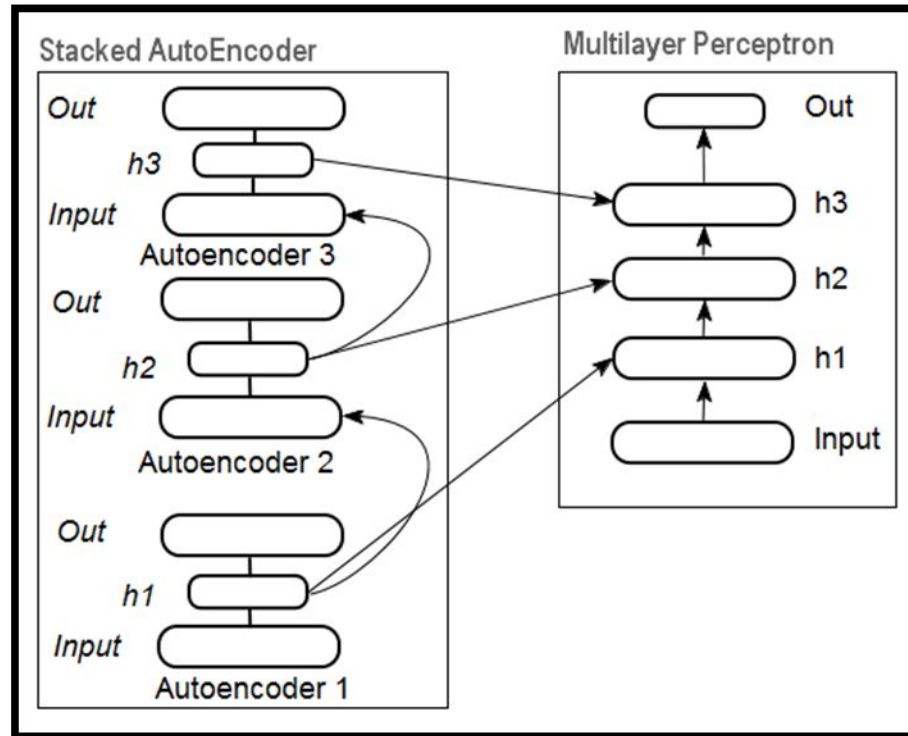
.

Fig 2:StackedAutoencoder

B. Deep Boltzmann Machines

The basic unit of Deep Boltzmann machines (DBM) are boltzmann machines which is constructed using stochastic binary nits. These units are symmetrically coupled and are stochastics. A Deep Boltzmann machine is a created with layers of boltzmann machines. It is created with a set of visible units and many layers of hidden units. DBMs are very good in learning complex and abstract internal representations of the input. They can be effectively used in tasks such as object or speech recognition. One advantage of DBM is that they use limited labelled data. They canget the exact representations constructed using a very large supply of unlabelled sensory input data. The performance and functionality of DBMs are limited because of speed[6].

C. Deep Belief Networks

A deep belief network (DBN) is a probabilistic and generative model. They are made up of multiple layers of latent units. The latent units are binary and input and output units are real or binary. One characteristics of BDN is that the layers are not intra connected. One layer units are connected with units of neighbouring layer. DBN has undirected connections in its deep layers while for the first two layers the connections are directed.

DBNs are useful when limited training data are available. Because of limited data weights can be poorly initialized and thus can significantly affect the model's performance. The weights obtained are in a region of the weight space that is near to the optimal weights. Thus it is easy to have improved modelling and faster convergence of the model.
A DBN can be efficiently trained in an unsupervised, layer-by-layer manner. The layers here are made of restricted Boltzmann machines (RBM). [7]
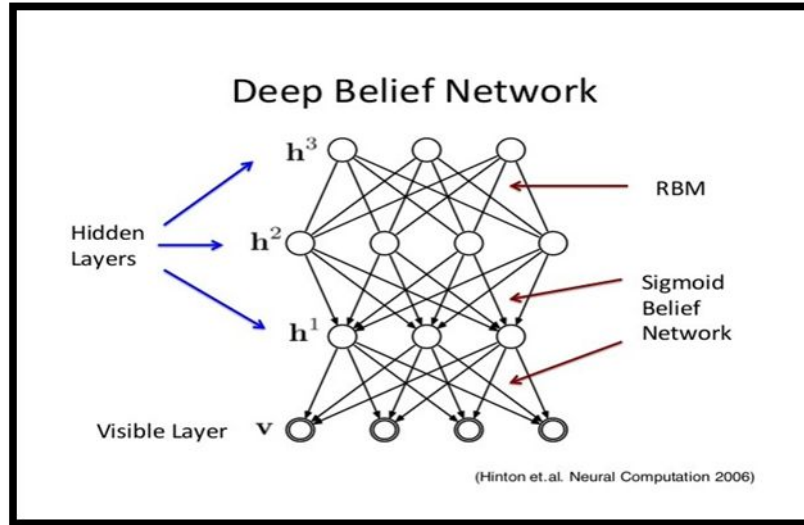
Fig3:Deep Belief Network

D. Deep Convolutional Networks

Deep Convolutional networks(DCN) can be constructed using layers of convolutional restricted Boltzmann machines. These machines are stacked together to form the network. In recent research convolutional deep belief networks are used for deep learning. The recent work includes using these networks on standard datasets to achieve many good results. Convolutional neural networks are used for image processing, on the same line; DCN can be effectively used for the task of image processing and object recognition.Also being a generative model it is used in other domains also to get best results. The main advantage of using DCN is scalability for high dimensional images and the processing is translation-invariant. [8]

III. APPLICATION AREAS

Deep learning architectures and algorithms have wide application area right from computer vision, automatic speech recognition, audio recognition, bioinformatics, natural language processing etc. These algorithms have shown remarkable results on various tasks. Following table gives a summary of various major application areas and deep learning algorithms used in these areas. Still there are many new areas where deep learning algorithms are not applied. The major research lies in testing these algorithms in these new area and finding out their usability.

Table 1: Deep Learning in various Application Areas

Application Area	Deep Learning Algorithms used	Paper	Inferences
Security	DBN with Dropout	DeepSign: Deep Learning for Automatic Malware Detection[9]	Deep learning is applied mainly in speech, image and video domain. New domain is tried as malware detection
Image Processing	CNN	Improving Deep Convolutional Neural Networks with Unsupervised Feature Learning [8]	Worked on image dataset. Next step is to investigate the application to other network architectures to gain further insight into the proposed approach.
	Stacked Denoising auto encoders	Stacked DenoisingAutoencoders: Learning useful representations in a Deep Network with a local Denoising	Classification error is reduced. Performance is equal to performance of DBN. Experimental results show that DAE are helpful for learning of higher level representations.

		criterion[10]	
	Autoencoder, Deep SVM and GMM	A novel deep learning by combining discriminative model with generative model [11]	Output is created using SVM and GMM and then are combined and given to another layer. Easier classification of the uncertain data.
	SVR with Linear Rectifier Units	Estimating crop yields with deep learning and remotely sensed data[12]	Using Satellite images crop yield estimation is done. Has better accuracy then normal SVM.
	Deep NuralNetwork with SVM	Deep Learning using Linear Support Vector Machines[13]	On standard datasets like MINST, CIFAR-10 proved to be more accurate. Simple for implementation.
	SDA and SVM	An introduction to deep learning[14]	Stacked auto encoder with SVM. Proved to be more accurate than SDA and logistic regression. Preprocessing is not required. While in regular SVM and KNN the preprocessing is required.
Regression Analysis	Deep SVM	Deep Support Vector Machines for regression problems[15]	Experimented on 10 regression datasets. Performance and accuracy is improved.
Finance	Deep SVM with Fuzzy	A Hierarchical Fused Fuzzy Deep Neural Network for data classification[16]	Deep learning SVM with fuzzy system is applied on financial trading system. Compared with normal SVM and normal Fuzzy neural network accuracy and performance is improved.
	Deep Neural Network with Genetic Algorithm	Genetic Deep Neural Networks Using Different Activation Functions for Financial Data Mining[17]	Two different activation functions are used. Used optimization method as GA for activation functions, future work is to use Ant Colony Optimization and bees algorithm. Better accuracy.
Big Data : Traffic Network Analysis	SAE with Logistic regression	Traffic Flow prediction with Big Data: A deep Learning Approach [18]	Stacked auto encoders are used and compared with SVM, logistic regression is used for prediction. Other Deep learning algorithms can be used. As well other prediction models can be used.
Big Data and Image Processing	SAE and SVM	A Deep Learning Method Combined Sparse Autoencoder with SVM[19]	SVM with Sparseautoencoder is used. Here a 2 hidden layer network is used to determine the features and then output is given to SVM for classification. Better results than SVM. 8 standard datasets were used.
Big Data and Education	SAE	Predicting Students performance in education data mining[20]	Used deep learning to predict students' performance. Used sparse Auto Encoders and back propagation for learning. Applied to 9th school children data. Used 4 hidden layers and GPU processor.
Music Genere Classification	Deep Feed forward network and LRU	Deep Neural Networks: A Case Study for Music Genere Classification[21]	Created a 2 hidden layer deep feed forward neural network for classification of music. Used rectifier linear units as the activation function. Used preprocessing on data to reduce the dimension.
Handwritten character recognition	DBM	Comparison of Different Variants of Restricted Boltzmann Machines[22]	DBM is the best RBM variant in terms of the classification errors. However, training a DBM is time consuming and the tuning of parameters is very difficult to the users.

Network Analysis	Restricted Boltzmann Machines	LRBM: A Restricted Boltzmann Machine based approach for representation learning on Linked Data [23]	RBM are used to do link prediction and node classification on linked data that is network data like Facebook network.
Sentiment Classification	Deep network using RBM	Active deep learning method for semi-supervised sentiment classification[24]	Used RBM with unsupervised learning. The last layer is the linear RBM. The hidden layer with linear RBM has more nodes as compared to other hidden layers.
Credit Card Analysis	Deep Neural Network and SVM	Deep Learning for Credit Card Data Analysis [25]	In this paper, two major applications are introduced to develop advanced deep learning methods for credit-card data analysis. The proposed methods are validated using various experiments with other machine learning algorithms. At the end it is observed that deep learning algorithms exhibits similar accuracy to the Gaussian kernel SVM.
Big Data	Tensor Auto Encoder	Deep Computation Model for Unsupervised feature learning on Big Data[26]	The tensor deep learning model takes more times to train the parameters than that of the stacked auto-encoder and the multimodal deep learning model, especially for feature learning on heterogeneous data.
	Deep SVM	Deep Twin Support Vector Machine[27]	Used twin SVM; that is first SVM for feature selection in hidden layer and another for classification as a main SVM. Accuracy is increased. Now these twin SVM are put in two layers - one in hidden layer whose output is given to main twin SVM.
	Deep Neural Network	Big data and deep learning[28]	Common training algorithm for deep networks is suggested, new visualization method is demonstrated.
Time Series Forecasting	DBN with SVR	Ensemble Deep Learning for Regression and Time Series Forecasting[29]	Ensemble deep belief network is proposed. Aggregated the outputs from various DBNs by a support vector regression model. Provided best training construction result and most accurate prediction results. Standard power and time series dataset are used.
Multimedia	DCNN	Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks [30]	Deep networks are used for structure output problem. The input is mapped to output which has its own structure. Encoder decoder architecture is used to describe multimedia content.

IV. DEEP LEARNING TOOLS

While Deep Learning algorithms seem attractive; choosing the correct Deep Learning tool for implementation is important. There are various tools available. This section gives a comparative overview of top Deep Learning Tools. All these tools are open source tools. [31,32,33,34,35]

Table 2: Summary of popular Deep Learning Tools

Tools → Features ↓	Pylearn2	Theano	Caffe	Torch	Cuda-convnet
Hardware	CPU , GPU	CPU, GPU	CPU,GPU	CPU,GPU	GPU
Architecture	Three key components: Dataset, Model, and Training Algorithm classes	Fairly Hacky	Standard layer wise design	Well designed with modular interface	Convolutional feed-forward neural networks
Ecosystem	Python	Python	C++	Lua	Matlab, C++/CUDA
Platform	Cross Platform	Cross Platform	Cross Platform	Linux based	Cross Platform
Performance	Good	Good	Good	Good	Best
Interfaces	Github	Python	Pycaffe	LuaJIT	CUDA
Advantages	Pylearn2 is designed for flexibility and extensibility of machine learning algorithms	Quickest, Computational graph is nice abstraction, High level wrappers ease the pain	Train models without writing any code, feedforward networks and image processing	Easy to write your own layer types and run on GPU, Lots of pretrained models	Efficient implementation of convolution in CUDA
Disadvantages	Load all data set to main memory	Error messages can be unhelpful, Large models can have long compile times	Need to write C++ / CUDA for new GPU layers,Not good for recurrent networks, Not extensible	Not good for recurrent neural networks	It requires High performance GPU.

V. CONCLUSION

Though Deep Networks are proposed in 1996 their use has gained popularity from 2006. Because of advancements in processing capabilities of machine and parallel processing framework, implementation of Deep Networks and Deep Learning Algorithms seems possible. Deep Learning algorithms are found to be very useful in the area of Image Processing, Signal Processing and Natural Language Processing. But now there is a need to apply these

concepts in new areas. With the advancement of Big Data new challenges are emerging and Deep Learning is proving useful in handling these challenges. From the table it is evident that new application area challenges are solved effectively with the help of Deep Learning algorithms. It will be very interesting to test these algorithms in still more new areas to get accurate results and prediction

REFERENCES

[1] EthemAlpaydin ,"Introduction to Machine Learning," Prentice Hall of India, 2004

[2] Y. Bengio, "Learning Deep Architectures for AI,"Foundations and Trends in Machine Learning , 2009, pages 1-127.

[3] Maryam M Najafabadi, Flavio Villansutre, Taghi M Khoshgoftaar, NaeemSeliya, Randall Wald and EdinMuharemagic , "Deep learning applications and challenges in big data analyatics," Spinger Open Journal of Big Data,2015

[4] Y Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy Layer-Wise Training of Deep Networks," Advances in Neural Iriformation Processing Systems 19 MIT Press, 2007, pages 153-160.

[5] P.Vincent, H. Larochelle Y. Bengio and P.A. Manzagol, "Extracting and Composing Robust Features with DenoisingAutoencoders," Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML '08), pages 1096 - 1103, ACM, 2008.

[6] RuslanSalakhutdinov, Geoffrey Hinton , "Deep Boltzman machines",12th International Conference on Artificial Intelligence and Statistics (AISTATS) ,Clearwater Beach, Florida, USA, 2009, Volume 5 of JMLR: W&CP 5

[7] Larochelle, Hugo, Bengio, Yoshua "Classification using Discriminative Restricted Boltzmann Machine," Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08),2008 ,Pages:536—543

[8] Kien Nguyen, Clinton Fookes, SridhaSridharan , "Improving Deep Convolutional Neural Networks With Unsupervised Feature Learning,"ICIC2015, 978-1-4799-8339-1/15,2015

[9] Omid E. Davidt,Nathan S. Netanyahu , "DeepSign: Deep Learning for Automatic Malware Detection", IEEE ,2015, 978-1-4799-1959-8/15

[10] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie,YoshuaBengio, Pierre-Antoine Manzagol , "Stacked DenoisingAutoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion", Journal of Machine Learning Research 11 (2010) 3371-3408

[11] Sangwook Kim ,Minho Lee,Jixiang Shen , "A novel deep learning by combining discriminative model with generative model, " IEEE, 2015

[12] KentaroKuwata and RyosukeShibasaki , "Estimating crop yields with deep learning and remotely sensed data," IGARSS , 2015

[13] M.A. Wiering, M. Schutten, A. Millea, A. Meijster, and L.R.B. Schomaker , "Deep Learning using Linear Support Vector Machines," International Conference on Machine Learning 2013: Challenges in Representation Learning Workshop. Atlanta, Georgia, USA, 2013

[14] Francis Quintal Lauzon, "An introduction to deep learning,"The 11th International Conference on Information Sciences, Signal Processing and their Applications: Special Sessions

[15] M.A. Wiering, M. Schutten, A. Millea, A. Meijster, and L.R.B. Schomaker ,"Deep Support Vector Machines for Regression Problems"

[16] Yue Deng, Zhiquan Ren, Youyong Kong, Feng Bao and Qionghai Dai, "A Hierarchical Fused Fuzzy Deep Neural Network for data classification,"IEEE,2016

[17] Luna M. Zhang, "Genetic Deep Neural Networks Using Different Activation Functions for Financial Data Mining," IEEE International Conference on Big Data (Big Data),2015

[18] YishengLv, YanjieDuan, Wenwen Kang, Zhengxi Li, and Fei-Yue Wang, Fellow , "Traffic Flow Prediction With Big Data:A Deep Learning Approach," IEEE Transactions On Intelligent Transportation Systems, VOL. 16, NO. 2, APRIL 2015

[19] Yao Ju, Jun Guo, Shuchun Liu , "A Deep Learning Method Combined Sparse Autoencoder with SVM," IEEE, 2015

[20] *Bo Guo∗, Rui Zhang, Guang Xu, Chuangming Shi and Li Yang , "Predicting Students performance in education*

data mining,"   IEEE-ISET 2015

[21] *Arjun Raj Rajanna, KameliaAryafar, Ali Shokoufandeh and Raymond Ptucha, "Deep Neural Networks: A Case Study for Music Genere Classification," IEEE 14th International Conference on Machine Learning and Applications,2015*

[22] *XiaoweiGuo , Haiying Huang , Jason Zhang , "Comparison of Different Variants of Restricted Boltzmann Machines," 2nd International Conference on Information Technology and Electronic Commerce(ICITEC), 2014*

[23] *Kang Li, Jing Gao, SuxinGuo, Nan Du, Xiaoyi Li and Aidong Zhang , "LRBM: A Restricted Boltzmann Machine based Approach for Representation Learning on Linked Data," IEEE International Conference on Data Mining 2014*

[24] *Shusen Zhou a,n, QingcaiChen b, XiaolongWang b , "Active deep learning method for semi-supervised sentiment classification," Elsevier, Neurocomputing, 2013*

[25] *AyahikoNiimi , "Deep Learning for Credit Card Data Analysis," World   Congress   on   Internet   Security (WorldCIS-2015) 2015*

[26] *Qingchen Zhang, Laurence T. Yang, and ZhikuiChen , "Deep Computation Model for Unsupervised feature learning on Big Data," IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 9, NO. 1, JANUARY/FEBRUARY 2016*

[27] *DeweiLi,YingjieTian,Honggui Xu , "Deep Twin Support Vector Machine," IEEE International Conference on Data Mining Workshop, 2014*

[28] *B.M. Wilamowaski, Bo Wu., JanuszKorniak , "Big data and deep learning," 20th Jubilee IEEE International Conference on Intelligent Engineering Systems , June 30-July 2, 2016*

[29] *XuehengQiu, Le Zhang, Ye Ren and P. N. Suganthan , "Ensemble Deep Learning for Regression and Time Series Forecasting," IEEE 978-1-4799-4512-2, 2014*

[30] *Kyunghyun Cho, Aaron Courville, and YoshuaBengio, "Describing Multimedia Content Using Attention-Based Encoder-Decoder Networks," IEEE Transactions On Multimedia, VOL. 17, NO. 11, NOVEMBER 2015*

[31] *"PYLEARN", http://deeplearning.net/software /pylearn2/ , December 2016*

[32] *Theano team, "Theano: A Python framework for fast computation of mathematical expressions," https://arxiv.org/pdf/1605.02688.pdf, May 2016*

[33] *"CudaConvnet", https://code.google.com/p/cuda-convnet/ , December 2016*

[34] *Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor, "Caffe: Convolutional Architecture for Fast Feature Embedding ,"arXiv preprint arXiv:1408.5093, 2014*

[35] *"Torch", https://deeplearning4j.org/compare-dl4j-torch7-pylearn , December 2016*