A Review on Conventional Machine Learning vs Deep Learning

Conference Paper · September 2018

DOI: 10.1109/GUCON.2018.8675097

2 authors:


Nitin Kumar Chauhan
Guru Gobind Singh Indraprastha University


Krishna Singh
G B PANT ENGINEERING COLLEGE NE DELHI

A Review on Conventional Machine Learning vs Deep Learning

Nitin Kumar Chauhan
USICT, GGSIPU
New Delhi, India
nitinchauhan7201@gmail.com

Dr. Krishna Singh
Department of Electronics and Communication Engg.
GBPEC, New Delhi
singhkrishna5@gmail.com

***Abstract*— In now days, deep learning has become a prominent and emerging research area in computer vision applications. Deep learning permits the multiple layers models for computation to learn representations of data by processing in their original form while it is not possible in conventional machine learning. These methods surprisingly improved the accuracy of various image processing domains such as speech recognition, face recognition, object detection and in biomedical applications. Deep neural networks (DNN) such as convolutional neural network (CNN) provide tremendous results in processing of images and videos, while another approach of deep network i.e. recurrent neural network (RNN) gives better performance with sequential data such as text and speech.**

Keywords—ANN, CNN, RNN, DNN, DT, SVM, PCA, LDA, QDA, Pooling layers, Fully connected layers, RBM.

I. INTRODUCTION

Conventional machine learning and deep learning both are the subfield of artificial intelligence. Artificial intelligence is the today's preferred research of interest of researchers in computer vision application. Both machine learning and deep learning fall under the wide-ranging class of artificial intelligence.

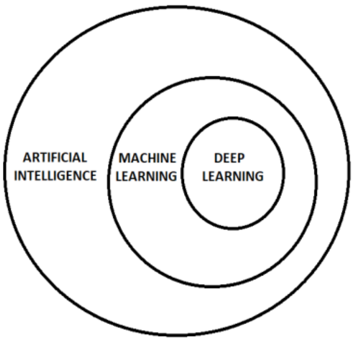

Fig. 1. Domains in Artificial Intelligence

Machine-learning, a computer vision algorithm is automatic learning of machines to classify and identify the data such as images, videos, objects, scenes, etc. Machine learning makes use of algorithms to analyze, learn, and decision making from the raw data [1]. Conventional machine-learning techniques have limited in capability of processing the data in their original form. These methods required considerable understanding and expertise for representation i.e. selection of features required careful engineering. It refers to a set of algorithms that permits a system to be input with dataset and it automatically realizes the representations required for decision making i.e. detection or classification [2, 3].

Deep learning is a subfield of machine learning. Deep learning is an advance machine learning approach that is used to make computers able to automatically extract, analyze and understand the useful information from the raw data. The results obtained from the deep learning are much improved than conventional machine learning approach [4]. Deep neural network uses non-linear model of multiple hidden layer architecture that make the system learn about the complex relationship between input and output. Advantage of deep learning over machine learning is that it does not require manually extracted or handcrafted features as in machine learning. Deep learning automatically extracts the features from the raw data, process it and make the decision based on this [5, 6].

Both the conventional machine learning and deep learning uses two types of learning methodology-supervised and unsupervised learning. In supervised learning the target value is assigned to train the data while in unsupervised learning target value is not provided. Supervised learning approach makes use of solving regression and classification problems. The unsupervised learning approach is used for making decision of association and clustering problems [4].

II. CONVENTIONAL MACHINE LEARNING

Machine learning is field of computer vision that, according to Arthur Samuel (1959) defined as "computers the ability to learn without being explicitly programmed." The computational theory of learning in artificial intelligence developed machine learning that explores analysis and building of algorithms that learn from the raw data, train the system and make predictions based on this train data [2, 3]. Conventional machine learning algorithms are based on learning of system by training set to develop a trained model as shown in fig 2. This pre-trained model is used to classify or recognize the test dataset shown in fig 3. Classification of various conventional machine learning algorithms for supervised and unsupervised learning is shown in fig 4.

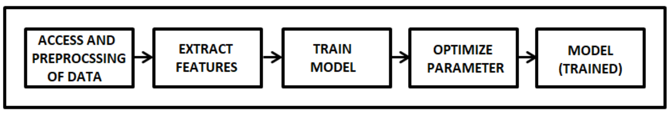

Fig. 2. Training of a model in machine learning

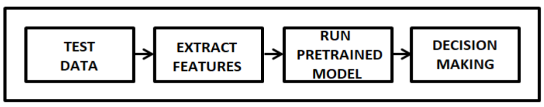

Fig. 3. Decision making for Test data

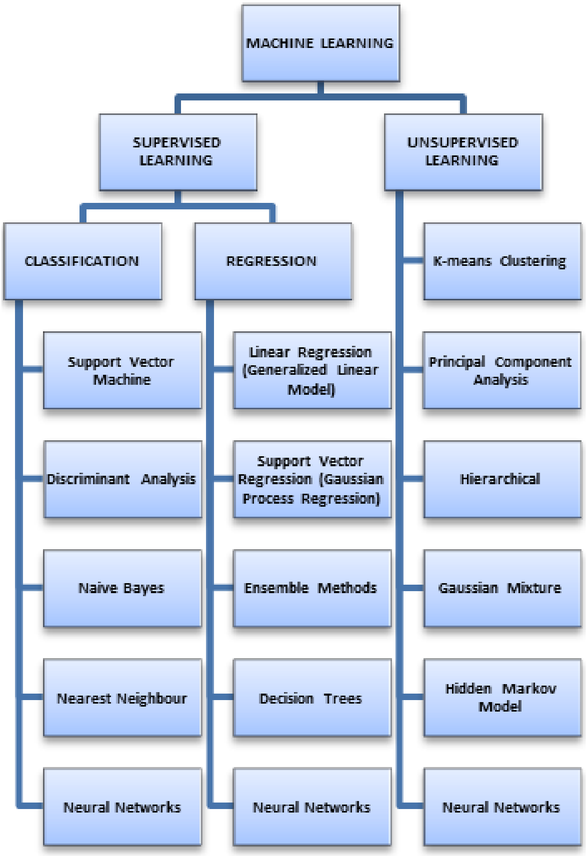

Fig. 4. Classification of conventional machine learning approaches

Some of the popular algorithms of machine learning are reviewed here:

A. Support vector machine (SVM)

SVMs are most efficient classification method [7]. SVM is a type of discriminative classifier well-defined through a separating hyperplane. Initially SVM model provides only binary classification, but later the extensions are developed to handle multiclass classification problems [8, 9]. They get added some additional constraint and parameters for separation of classes. Since performance of system is depends on selection of features, a hybrid model PSO-SVM is developed for optimized feature selection [10].

B. Discriminative analysis

Discriminate analysis is based on classification of objects where two or more objects are grouped or clustered based on the measured features that are used to describe that object. Linear discriminative analysis (LDA) is used generally for dimensionality reduction while quadratic discriminative analysis (QDA) uses quadratic surface for separation of classes [11].

C. Naive Bayes classifier

Naive Bayesian classifiers are Bayesian networks that make use of directed acyclic graphs containing only one unobserved (parent) node and several observed (children) nodes having an assumption of independency among them that is given by Naive Bayes independency model [12]. There is no need to set the free parameters as required in SVM and Neural Networks that simplifies the Naive Bayes. It results in probability that makes Naive Bayes easy to apply on wise range of tasks [13].

D. Decision Trees

Decision tree approach is another supervised learning algorithm in which all the dataset is labeled. Decision trees are used to classification by sorting the classes based on parameters values. ID3, C4.5, CHAID and CART are some of the algorithms belong to decision tree. Major advantage of this approach is that it is able to handle numerical as well as categorical attributes. This method holds good for small datasets but causes lagging for large datasets. However building of decision tree required considerable time, although time required for computation is less [14].

E. K-means Clustering

Clustering is an example of unsupervised learning approach in which test dataset are not labeled. Hierarchical clustering is based on building hierarchy that uses two types of a clustering technique: Agglomerative and Divisive. Agglomerative Clustering pair up to create big cluster in bottom-up manner. Divisive Clustering uses broken of a big cluster dividing them into the small clusters in a top-down manner. Partitioning Clustering is a technique that makes use of partitioning of the datasets into equal or unequal sets in which each is characterized in the form of cluster mean. In K-means clustering the dataset is a partitioned into set of K-small clusters in which each is represented through its cluster mean [15].

F. Principal component analysis (PCA)

PCA is used for dimensionality reduction by transforming the high dimensional data to low dimensional data. If the learning algorithms have m inputs, m outputs and n hidden layers then m>n. However PCA capability is limited only upto linear transformation of one space (m) to another space (n) [16].

G. Neural Networks (NN)

The neural network also referred as artificial neural network (ANN) is originated from biological theory of neurons. It consists of input layers, hidden layers and output layers. The hidden layers are used to process the input received and send them to output layers. The basic architecture of neural network is shown in fig 5. NN approach provides more accuracy and efficiency than pre-existing methods. Various approaches of NN are back propagation algorithm (BPNN), Radial basis function (RBF), Complementary NN (CMPNN) and Probabilistic approach. BPNN is the most widely used NN approach however it is slower in training when large dataset is used [11].

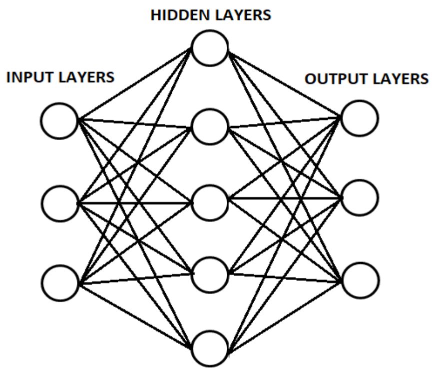

Fig. 5. Basic architechure of Neural Network

Single layer perceptron is the basic form of NN that consists of single neuron with adjustable weights and used to classify linearly separable classes. A Multilayer perceptron (MLP) is an efficient and robust algorithm used for modeling non-linear and complex problems. The determination of size of hidden layer is the complex task. The underestimation of number of hidden layers will results in poor approximation while overestimation may result in generalization error and overfitting [5, 11].

The experimental result based on accuracy of different machine learning approaches is shown in Table I.

TABLE I. COMPARISON OF ACCURACY FOR DIFFERENT MACHINE LEARNING ALGORITHMS [17]

S. No.	Algorithm	Accuracy	
		Test Accuracy	*Cross-validation*
1.	SVM RBF Kernel	0.627	0.819
2.	SVM Polynomial Kernel	0.761	0.821
3.	SVM Linear Kernel	0.766	0.785
4.	LDA	0.765	0.804
5.	QDA	0.765	0.802
6.	Logistic Regression	0.751	0.793
7.	CART	0.785	0.799
8.	C4.5	0.765	0.872
9.	Naïve Bayes	0.737	0.778
10.	Neural Network	0.746	0.811
11.	K-NN	0.632	0.819

III. DEEP LEARNING

Deep learning is recently become the trending investigation domain in the field of artificial intelligence. Selectivity-invariance problem is the major disadvantage of conventional machine learning because of which these approaches have limited capability of processing the data in their original form. Selectivity-invariance implies selection of those features that contain more information and ignore less information containing parameters i.e. selected features should be distinct to each other. This motivated researchers to advancement of machine learning approach i.e. deep learning. Deep learning sometimes also referred as representation learning having a set of algorithms that automatically discovers the classification or detection needed by allowing a machine to be input with original dataset [5]. Deep-learning approaches have multiple abstractions of levels by using non-linear model that transforms the original data into higher abstract levels for decision making. This simplifies finding the solution of complex and non-linear functions [18]. Deep learning is based on automatic learning of features that provides facility of modularity and transfer learning. As its name implies deep learning generally contain deep architecture of hidden layers as shown in Fig 6. Unlike the conventional machine learning, deep learning requires large amount of data to train a network.

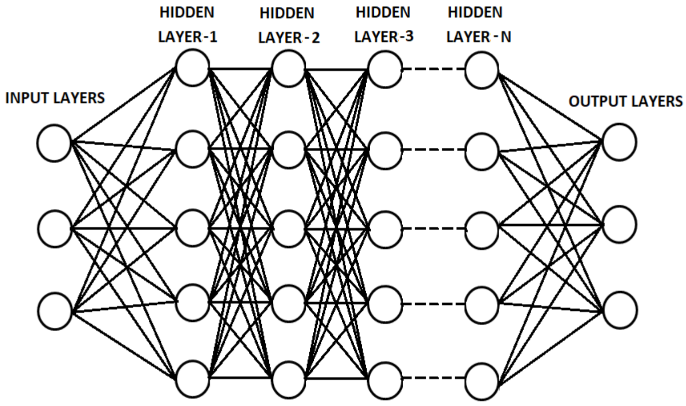

Fig. 6. Architechure of a Deep Network

Most commonly used deep networks are Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) which we reviewed in the upcoming sections, however our major focus on CNN.

A. *Convolutional Neural Network (CNN)*

CNN is the most popular deep approach that is based on the animal's visual cortex [19]. CNN is currently extensively used in object recognition, object tracking [20], text recognition and detection [21], visual detection [22], pose estimation [23], scene labeling [24] and in various other applications [25]. CNNs are almost similar to artificial neural networks that can be seen as acyclic graph with a collection of neurons in well-arranged form. In CNNs unlike the neural network, the hidden layers of neurons are only connected with previous layer containing the subset of neurons. This type of sparse connectivity makes the system able to implicitly learning of features.

Each section of CNN layer contains two or more dimensional filters that are convolved along with the input of layer. Deep convolutional network results in hierarchical extraction of feature [6]. The architecture of CNN consists of convolutional layers, pooling layers and fully connected layers as shown in Fig 7.

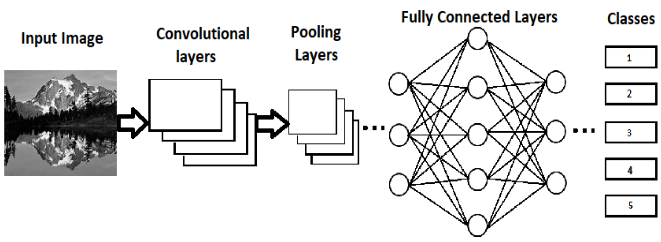

Fig. 7. Pooling operation in CNN

Convolutional Layer - This layer is the primary unit of CNN in which most of computations are performed. These layers consist of arrangement of neurons along with set of feature maps. These layers have filters (kernels) that are used to convolve with features for producing a separate 2-D activation map [6] as shown in Fig 8. Complexity of network is reduced by keeping the less number of features by sharing of weights by the neurons [26]. The CNNs are trained by Backpropagation that involves convolution operation with spatially flipped filters. These layers apply convolutional operation along with filtering and pass the result to the next layers.

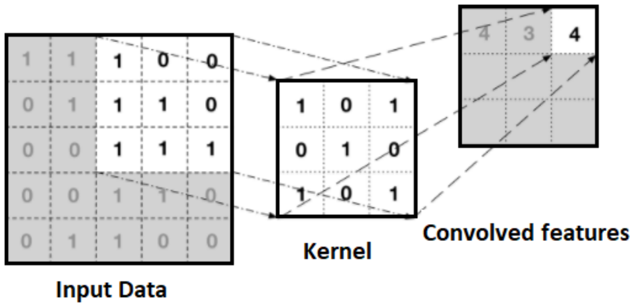

Fig. 8. Convolutional operation in CNN

Pooling layers - CNN have alternating convolutional and pooling layers. Pooling layers is used to reduce the spatial dimension of image activation maps and the number of features without any loss of information results in reduction in overall computational complexity. Pooling layers perfectly handles the problem cause due to overfitting. Average pooling, max pooling, stochastic pooling [27], spatial pyramid pooling [28], spectral pooling [29] and multiscale pooling [30] are some of commonly used pooling operations. Max pooling operation is shown in Fig 9.

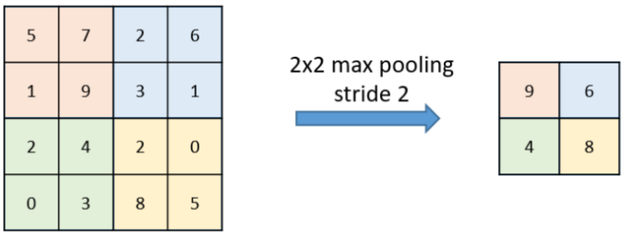

Fig. 9. Pooling operation in CNN

Fully Connected Layer – Fully connected layers connects all the neurons in of one layer to all neurons of the next layer like NN. It converts high dimensional feature map to one dimensional by high level reasoning that gives the probability of feature belong to a specific class. Some of the approaches are developed recently to replace fully connected layers such as "Network In Network"(NIN) architecture [31].

Some of the commonly used CNN architectures are LeNet [32], AlexNet [33], GoogleNet [34], Network in Network [31], etc. The comparative study of these CNNs network with some other deep neural network is given in Table II.

TABLE II. Comparative study of various Deep Neural Networks

Model	Type	Architecture	Learning
Deep Autoencoder / Stacked Autoencoder [35]	generative	multiple hidden layers, fully connected architecture with bidirectional connections	pretraining: unsupervised, fine-tuning: supervised
Deep Belief Network [36]	generative	multiple hidden layers, fully connected architecture with partial bi-directional connections	pretraining: unsupervised, fine-tuning: unsupervised
Restricted Boltzmann Machines (RBM) [37]	generative	multiple hidden layers, fully connected architecture with bidirectional connections	pretraining: unsupervised with large supply of unlabeled data, fine-tuning: supervised
LeNet-5 (Convolutional Neural Network) [32]	discriminative	multiple hidden layers, locally connected architecture	supervised
AlexNet (Convolutional Neural Network) [33]	discriminative	multiple hidden layers, locally connected architecture, dropout technique for fully connected layers	supervised
Network In Network (Convolutional Neural Network) [31]	discriminative	multiple hidden layers with small MLP networks between convolutional layers, locally connected architecture	supervised
GoogLeNet (Convolutional Neural Network) [34]	discriminative	multiple hidden layers, locally connected architecture with inception modules	supervised

B. *Recurrent neural networks*

RNNs are used for processing of tasks that involve sequential inputs for ex. text and speech. Initially backpropagation was used was for training of RNNs. RNNs process one element of an input sequence at a time with maintaining state vector in their hidden units in which implicitly within units contains information of all the past value of elements of that sequence. The general architecture for RNN is shown in Fig 10. The output at different discrete steps of time of hidden neurons that are the outputs given by all the neurons within a deep multilayer network, it becomes clear how we can apply backpropagation to train RNNs [5]. RNNs are quite powerful and dynamic systems, but the problem causes during the training procedure as in backpropagation algorithm gradients either would grow or shrink at every time step, hence they vanish after many time steps [38].

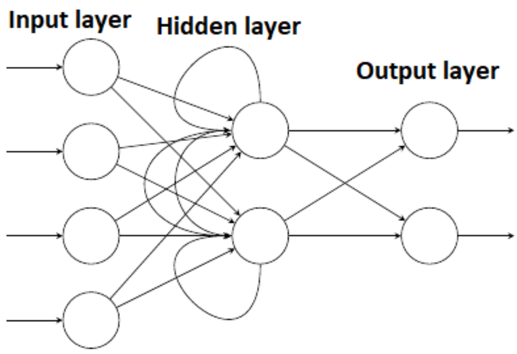

Fig. 10. Layered architechure of in RNN

RNNs when unfolded can be seen as deep feed-forward networks in which same weight is shared by all the layers. The problem in RNN is to store the past information for very long time i.e. long-term dependencies. To overcome this one approach given with explicit memory is long short-term memory (LSTM) method that uses special hidden nodes or units to remember the parameters in form of input for a long time [39]. LSTM gave much improved performance in speech recognition systems [40]. The performance comparison of LSTM with SVM and RNN is shown in Table III.

TABLE III. Error percentage for different models [39]

Methods	Error		
	1st observation point	*2ndobservation point*	*3rdobservation point*
SVM	12.3	10.6	8.2
RNN	6.3	6.3	6.7
LSTM	6.4	6	6.2

IV. CONCLUSION AND FUTURE WORK

In this paper we reviewed the difference between conventional machine learning and deep learning. Here we made a comparative study between different machine learning approaches such as SVM, PCA, LDA, decision trees, NN, etc. None of the machine learning approach holds good result for each application. Thus performance of these approaches differs based on the application. Deep learning, an extension machine learning required millions of data to train an efficient network, unlike machine leaning in which few data is also sufficient to train a network. Deep networks such as CNN and RNN, and the advancement of these approaches are reviewed here.

Deep neural networks achieved great success in recent time. However, effective and accessible parallel learning approaches are the domains where further investigation is required to speed up the learning process. Due to long tail distribution of object classes in several application domains, deep networks required to handle the class-imbalance problem appropriately.

REFERENCES

[1] Munoz, and Andres, “Machine Learning and Optimization,” Retrieved 2017-06-01 from https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf

[2] Hosch, and L. William, “Machine Learning,” Retrieved 2017-06-01 from http://www.britannica.com/EBchecked/topic/1116194/machinelearnin g

[3] Ron Kohavi, Foster Provost, "Glossary of terms: Machine Learning,” pp. 271–274, 1998.

[4] S. Setiowati, Zulfanahri, E. L. Franita and I. Ardiyanto, "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods," *9th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Phuket, 2017, pp. 1-6.

[5] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, May 2015.

[6] N. Aloysius, and M. Geetha, "A review on deep convolutional neural networks," *International Conference on Communication and Signal Processing (ICCSP),* Chennai, 2017, pp. 0588-0592.

[7] C. Cortes, and V. Vapnik, “Support-vector networks,” *Machine Learning*, pages 273–297, 1995.

[8] K. Crammer, and Y. Singer, “On the algorithmic implementation of multiclass kernel-based vector machines,” *Journal of Machine Learning Research*, pages 265–292, 2001.

[9] Y. Lee, Y. Lin, and G. Wahba, “Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data,” *Journal of the American Statistical Association*, 99(465):67–81, March 2004.

[10] S. W. Lin, K. C. Ying, S. C. Chen, and Z. J. Lee, “Particle swarm optimization for parameter determination and feature selection of support vector machines”, *Expert system with applications*, vol. 35, no. 4, pp. 1817-1824, 2008.

[11] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques”, *Informatica*, Vol. 31, No. 3, pp. 249-268, 2007.

[12] L. I. Kuncheva, “On the optimality of Naïve Bayes with dependent binary features”, *Pattern Recognition Letters,* vol. 27, pp. 830-837, 2006.

[13] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-Ahmed, “Investigating the Performance of Naive- Bayes Classifiers and K-Nearest Neighbor Classifiers,” *Journal on Convergence Information Technology,* vol. 5, no. 2, 2010.

[14] L. Rokach and O. Maimon, "Top-down induction of decision trees classifiers - a survey," *IEEE Transactions on Systems, Man, and Cybernetics*, Part C (Applications and Reviews), vol. 35, no. 4, pp. 476-487, Nov. 2005.

[15] K. Alsabati, S. Ranaka, V. Singh, “An efficient k-means clustering algorithm”, *Electrical Engineering and Computer Science*, 1997.

[16] M. Andrecut, “Parallel GPU Implementation of Iterative PCA Algorithms”, *Institute of Biocomplexity and Informatics*, University of Calgary, Canada, 2008.

[17] A. Singh, N. Thakur and A. Sharma, "A review of supervised machine learning algorithms," *3rd International Conference on Computing for Sustainable Global Development (INDIACom*), New Delhi, 2016, pp. 1310-1315.

[18] R. Ramachandran, D. C. Rajeev, S. G. Krishnan, and P Subathra, “Deep learning an overview,” *IJAER,* Volume 10, Issue 10, 2015, Pages 25433-25448.

[19] D. H. Hubel, and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *The Journal of physiology*, 1968.

[20] J. Fan, W. Xu, Y. Wu, and Y. Gong, “Human tracking using convolutional neural networks,” *Neural Networks*, IEEE Transactions, 2010.

[21] M. Jaderberg, A. Vedaldi, and A. Zisserman, “Deep features for text spotting”, *ECCV*, 2014.

[22] R. Zhao, W. Ouyang, H. Li, and X. Wang, “Saliency detection by multicontext deep learning,” *CVPR*, 2015.

[23] A. Toshev, and C. Szegedy, “Deep -pose: Human pose estimation via deepneural networks,” *CVPR*, 2014.

[24] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” *PAMI*, 2013.

[25] Nithin, D Kanishka, P Bagavathi, and Sivakumar, “Generic Feature Learning in Computer Vision,” *Elsevier*, Vol.58, Pages202-209, 2015.

[26] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, “Improving neural networks by preventing co-adaptation of feature detectors,” arXiv preprint arXiv:1207.0580, 2012.

[27] M. D. Zeiler, and R. Fergus, “Stochastic pooling for regularization of deep convolutional neural networks,” *CoRR*, 2013.

[28] A. Nguyen , J. Yosinski, and J. Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2015.

[29] O. Rippel, J. Snoek, and R. P. Adams, “Spectral representations for convolutional neural networks,” arXiv preprint arXiv:1506.03767, 2015.

[30] Y. Gong, L. Wang, R. Guo, and S. Lazebnik, “Multi-scale orderless pooling of deep convolutional activation features,” *ECCV,* 2014.

[31] M. Lin, Q. Chen, and S. Yan, “Network in network”, *Computing Research Repository (CoRR)*, vol. abs/1312.4400, 2013.

[32] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, *Intelligent Signal Processing (*S. Haykin and B. Kosko, eds.), vol. 86, pp. 306–351, IEEE Press, 2001.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks”, in *Advances in Neural Information Processing Systems*, 25-26th Annual Conference on Neural Information Processing Systems, (Lake Tahoe, Nevada, United States), pp. 1106– 1114, 2012.

[34] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, and S. Reed, “Going deeper with convolutions,” *CoRR*, 2014.

[35] G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks”, *Science*, vol. 313, pp. 504–507, July 2006.

[36] G. E. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets”, *Neural Computation*, vol. 18, pp. 1527–1554, July 2006.

[37] T. Tieleman, “Training of restricted botzmann machines using approximations to the likelihood gradient”, in *Machine Learning, Proceddings of the Twenty-Fifth Internaltional Conference (ICML)*, (Helsinki, Finland), pp. 1064–1071, June 5-9 2008.

[38] Y. Bengio, P. Simard, and P.Frasconi, " Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Networks 5*, 157–166 (1994).
[39] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural Comput. 9*, 1735–1780 (1997).
[40] Chan, W. et al., "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," in *Speech and Signal Processing (ICASSP)*,2016.