ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 12 , December 2019

An overview of Neural Networks and Deep Learning

Hanan Qassim Jaleel

Baghdad College of Medical Sciences

ABSTRACT: Deep learning is a set of learning methods attempting to model data with complex architectures combining different non-linear transformations. The elementary bricks of deep learning are the neural networks, that are combined to form the deep neural networks. These techniques have enabled significant progress in the fields of sound and image processing, including facial recognition, speech recognition, computer vision, automated language processing, text classification (for example spam recognition). Potential applications are very numerous. There exist several types of architectures for neural networks such as The multilayer perceptrons, that are the oldest and simplest ones , The Convolutional Neural Networks (CNN), particularly adapted for image processing and The recurrent neural networks, used for sequential data such as text or times series. They are based on deep cascade of layers. They need clever stochastic optimization algorithms, and initialization, and also a clever choice of the structure. They lead to very impressive results, although very few theoretical fondations are available till now.

This paper will explain an overview of neural networks anal back propagation , deep learning algorithms such as: convolutional neural networks and deep beliefe networks , advantgaes , disadvantages and applications

1. INTRODUCTION

Neural Networks (NN) are important data mining tool used for classification and clustering. It is an attempt to build machine that will mimic brain activities and be able to learn. NN usually learns by examples. If NN is supplied with enough examples, it should be able to perform classification and even discover new trends or patterns in data. Basic NN is composed of three layers, input, output and hidden layer. Each layer can have number of nodes and nodes from input layer are connected to the nodes from hidden layer. Nodes from hidden layer are connected to the nodes from output layer. Those connections represent weights between nodes[1].

II.ARCHITECTURE OF MULTIPLE LAYER NN

The following figure (1)represent architecture of an simple NN. It is made up from an input, output and one or more hidden layers. Each node from input layer is connected to a node from hidden layer and every node from hidden layer is connected to a node in output layer. There is usually some weight associated with every connection. Input layer represents an the raw information that is fed into the network. This part of network is never changing its values. Every single input to the network is duplicated and send down to the nodes in hidden layer. Hidden Layer accepts data from the input layer. It uses input values and modifies them using some weight value, this new value is than send to the output layer but it will also be modified by some weight from connection between hidden and output layer. Output layer process information received from the hidden layer and produces an output. This output is than processed by activation function[1,2].

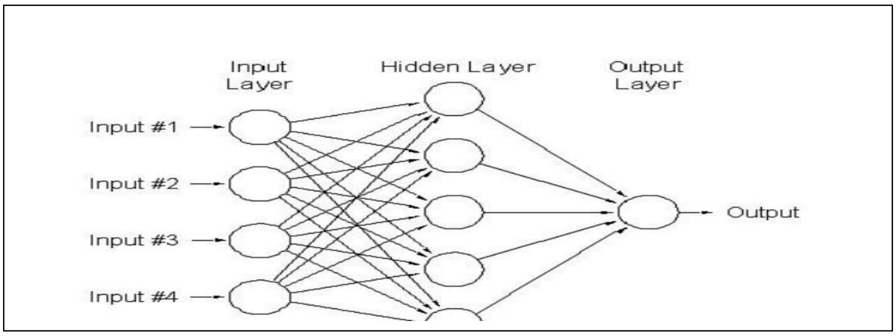


Figure (1) architecture of an Multi Layer NN

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 12 , December 2019

II.ANN Back Propagation (BP)

The most popular of ANN is **Back Propagation (BP)**. Idea behind BP algorithm is quite simple. The output of NN is evaluated against desired output. If results are not satisfactory, connection (**weights**) between layers are modified and process is repeated again and again until error is small enough. It is supervised where output values are known beforehand.

A. Supervised Learning

Supervised learning, is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing.
Training data for supervised learning includes a set of examples with paired input subjects and desired output (which is also referred to as the supervisory signal)

B . Unsupervised Learning

Unsupervised learning is the training (AI) algorithm using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.In unsupervised learning, an AI system may group unsorted information according to similarities and differences even though there are no categories provided[1,2,3].

III. DEEP LEANING ALGORITHMS

Also known as deep structured learning, hierarchical learning or deep machine learning) is a class of machine learning algorithms that use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised). There types are**deep belief networks (DBN)** and **convolutional neural networks (CNNs)[4].**

A . convolutional neural networks (CNN)

A typical CNN is composed of many layers of hierarchy with some layers for feature representations (or feature maps) and others as a type of conventional neural networks for classification. It often starts with two altering types of layers called convolutional and subsampling layers: convolutional layers perform convolution operations with several filter maps of equal size, while subsampling layers reduce the sizes of proceeding layers by averaging pixels within a small neighborhood (or by max-pooling.Facebook uses neural nets for their automatic tagging algorithms, Google for their photo search, Amazon for their product recommendations, and Instagram for their search infrastructure. Simple ConvNet is a sequence of layers: **Convolutional Layer**, **Pooling Layer,** and **Fully-Connected Layer[5,6].**

Figure shows a typical architecture of CNNs. The input is first convoluted with a set of filters (C layers in Fig). These 2D filtered data are called feature maps. After a nonlinear transformation, a subsampling is further performed to reduce the dimensionality (S layers in Figure 2). The sequence of convolution/subsampling can be repeated many times (predetermined by users)[7,8].

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Issue 12 , December 2019

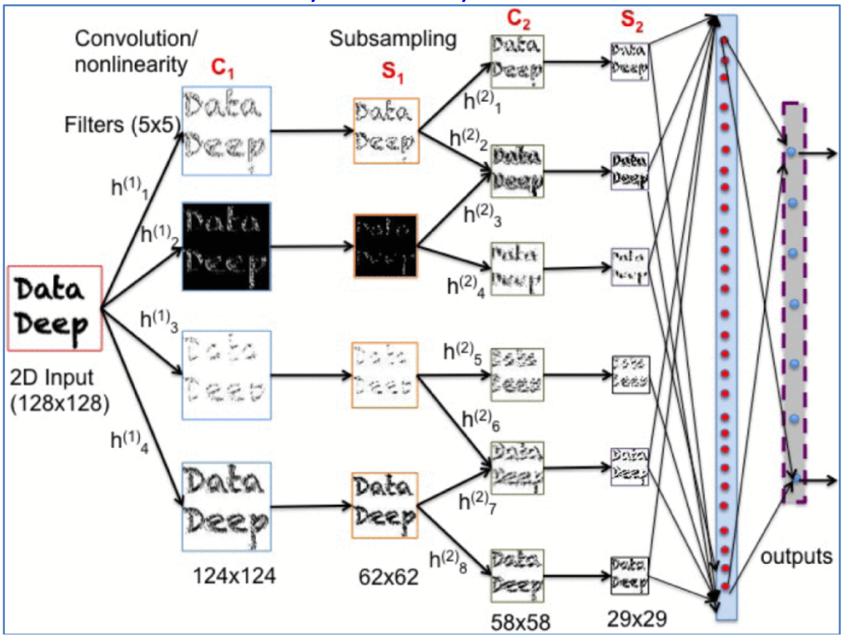


Figure (2) CNN

1-Convolutional Layer: Also referred to as Conv. layer, it forms the basis of the CNN and performs the core operations of training and consequently firing the neurons of the network. It performs the convolution operation over the input volume and consists of a 3-dimensional arrangement of neurons (a stack of 2-dimensional layers of neurons, one for each channel depth)[9,10]. That make :

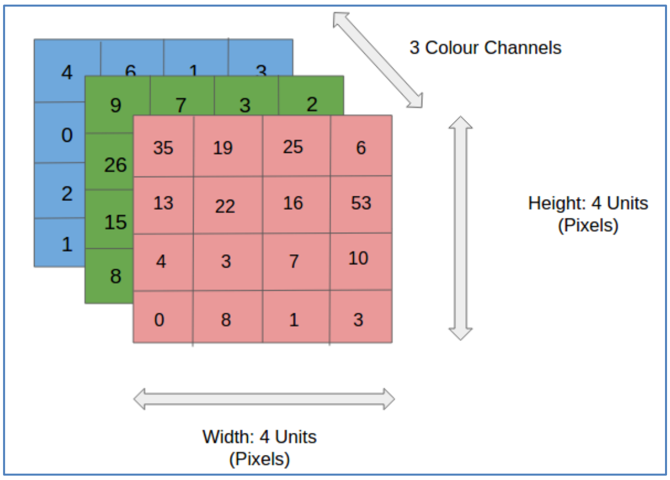
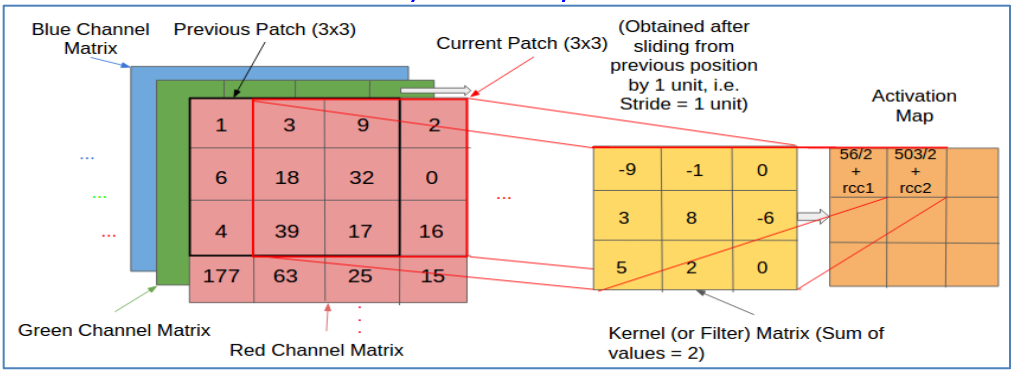


Figure (3) Filters (Convolution Kernels)

Filters (Convolution Kernels) :A filter (or kernel) is an integral component of the layered architecture. it refers to an operator applied to the entirety of the image such that it transforms the information encoded in the pixels. In practice, a kernel is a smaller-sized matrix in comparison to the input dimensions of the image. It's convolved with the input volume to obtain **Activation maps** indicate 'activated' regions, i.e. regions where features specific to the kernel have been detected in the inpu[11,12]t. This is depicted in figure(3)



Receptive Field: It is impractical to connect all neurons with all possible regions of the input volume. It would lead to too many weights to train ,we specify a 2 dimensional region called the 'receptive field' (say of size 5×5 units) extending to the entire depth of the input (5x5x3 for a 3 colour channel input), within which the encompassed pixels are fully connected to the neural network's input layer. It's over these small regions that the network layer cross-sections (each consisting of several neurons (called 'depth columns')) operate and produce the activation map[12].

Zero-padding:refers to the process of symmetrically adding zeroes to the input matrix. It's a commonly used modification that allows the size of the input to be adjusted to our requirement. It is mostly used in designing the CNN layers when the dimensions of the input volume need to be preserved in the output volume[13].

2-Pooling layer (subsampling layer): it is between successive Conv layers in ConvNet architecture. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control over fitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially, using the **MAX** operation. The most common form is a pooling layer with filters of size 2x2 applied with a stride of 2 down samples every depth slice in the input by 2 along both width and height, discarding 75% of the activations. Every **MAX** operation would in this case be taking a max over 4 numbers (little 2x2region in some depth slice). The depth dimension remains unchanged[12,14]. The poling layer is illustrated in figure(4)

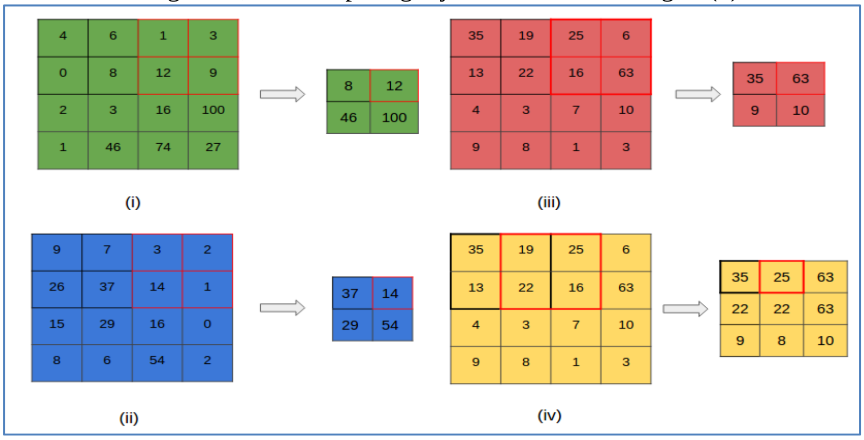


Figure (4) pooling layer(subsampling layer)

3-The Fully Connected layer: is configured exactly the way its name implies: it is fully connected with the output of the previous layer. Fully-connected layers are typically used in the last stages of the CNN to connect to the output layer and construct the desired number of outputs[15].

B. Deep belief networks (DBN)

Deep belief network (DBN) is a generative graphical model, or alternatively a type of deep neural network, composed of multiple layers of latent variables ("hidden units") with connections between the layers but not between units within each layer. When trained on a set of examples in an unsupervised way, a DBN can learn to probabilistically reconstruct its inputs. The layers then act as feature detectors on inputs. After this learning step, a DBN can be further trained in a supervised way to perform classification.

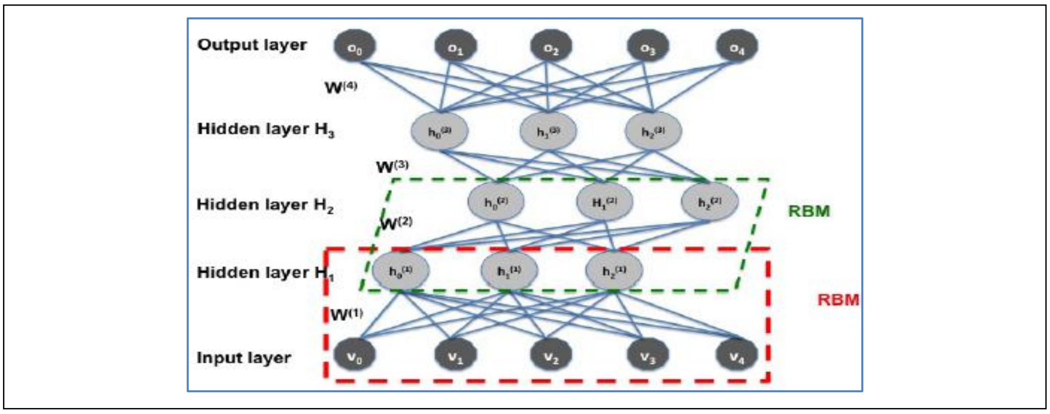
ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 12 , December 2019

the DBN network is composed of stacked modules, which can be trained in unsupervised manner, and can be complemented by final top stage for classification. In the case of DBNs, the individual modules take the form of Restricted Boltzmann Machines (RBMs), a two-layered neural networks with connections only between the layers.

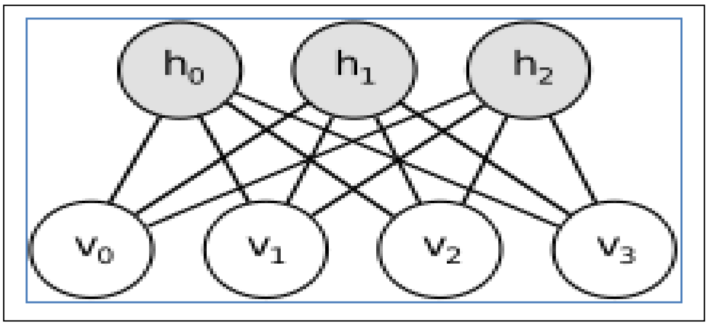
RBM is is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputswhere each sub-network's hidden layer serves as the visible layer for the next[10,16].

the DBN model is trained one module after another, proceeding from the lowest RBM and using its outputs (after its training is finished) as inputs for training the subsequent RBM module. After the training of all RBM modules is finished, the output of the final one can be used as representation of the inputs with reduced dimensionality (see [Hinton, 2006b]), or as input into classification layer. Alternatively, an auto-associative memory network can be trained as the last module, with label information added to the top layer, to add classification capability to the network. Also, the inner functionality of RBM modules allows their parameters to be used for initialization of deep perceptron network, which can then be fine-tuned by supervised criterion (e.g. back-propagation algorithm)[16].

An RBM network is composed of two layers of neurons. The lower one is called visible (or input) layer, and the output values of its neurons will be denoted as vi . The higher layer is called hidden, and outputs of its neurons will be denoted hj. The weight of connection between neurons vi and hj is denoted by wij , and all connections are symmetric, meaning that wij = wji. The visible and hidden neurons also have bias parameters, denoted as bi and cj , respectively[17]. Figure(5) shows the restricted Boltzmann machines (RBMs) and Figure(6) shows the structure of RBM



Figure(5) restricted Boltzmann machines (RBMs)



Figure(6) the structure of RBM

<u>The training algorithm for DBN</u>

Let *X* be a matrix of inputs, regarded as a set of feature vectors:

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 12 , December 2019

1-Train a restricted Boltzmann machine (RBM) on X to obtain its weight matrix, W. Use this as the weight matrix between the lower two layers of the network.

2- Transform X by the RBM to produce new data X', either by sampling or by computing the mean activation of the hidden units.

3- Repeat this procedure with $X \leftarrow X'$ for the next pair of layers, until the top two layers of the network are reached.

4- Fine-tune all the parameters of this deep architecture with respect to a supervised training criterion (after adding extra learning machinery to convert the learned representation into supervised predictions, e.g. a linear classifier).

For a simple RBM with Bernoulli distribution for both the visible and hidden layers, the sampling probabilities are as follows :

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_{i=1}^{I} v_i W_{ji}) \quad (1)$$

$$p(v_i = 1|\mathbf{h}) = \sigma(a_i + \sum_{j=1}^{J} h_j W_{ji}) \quad (2)$$

Where **v** and **h** represents a I×1 visible unit vector and a J×1 hidden unit vector, respectively; **W** is the matrix of weights (**wij**) connecting the visible and hidden layers, aj and bi are bias terms and σ(·) is a sigmoid function. For the case of real-valued visible units, the conditional probability distributions are slightly different[1,18].
Weights wij are updated based on an approximate method called **contrastive divergence** (CD) approximation . For example, the (t + 1)-th weight for wij can be updated as follows:

$$\Delta w_{ij}(t+1) = c\Delta w_{ij}(t) + \alpha\left(\langle v_i h_j\rangle_{data} - \langle v_i h_j\rangle_{model}\right) \quad (3)$$

Where **α** is the learning rate and **c** is the momentum factor; ⟨·⟩data and ⟨·⟩model are the expectations under the distributions defined by the data and the model, respectively.Other model parameters (e.g., the biases) can be updated similarly.
After pre-training, information about the input data is stored in the weights between every adjacent layer. The DBN then adds a final layer representing the desired outputs and the overall network is fine tuned using labeled data and back propagation strategies for better discrimination (in some implementations, on top of the stacked RBMs, there is another layer called associative memory determined by supervised learning methods)[2,19].

IV. DEEP LEARNING APPLICATIONS

1-Natural Language Processing
2-Optical Character Recognition
3-Speech Recognition
4-Object Classification and Detection in Photographs
5- Image classification

4-1Constrastive Divergenve Procedure CD

1-initialize the visible units to a training vector
2- update the hidden units in parallel given the visible

$$p(h_j = 1|\mathbf{v}) = \sigma(b_j + \sum_{i=1}^{I} v_i W_{ji})$$

3- update the visible units in parallel given the hidden units: This is called the "reconstruction" step.

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 6, Issue 12 , December 2019

$$p(v_i = 1|\mathbf{h}) = \sigma(a_i + \sum_{j=1}^{J} h_j W_{ji})$$

4-Re-update the hidden units in parallel given the reconstructed visible units using the same equation as in step 2.
5-Perform the weight update[7]

$$\Delta w_{ij}(t+1) = c\Delta w_{ij}(t) + \alpha\left(\langle v_i h_j\rangle_{data} - \langle v_i h_j\rangle_{model}\right)$$

V.ADVANTAGES

1-Reduces the need for feature engineering, one of the most time-consuming parts of machine learning practice
2-Is an architecture that can be adapted to new problems relatively easily (e.g. Vision, time series, language etc using techniques like convolutional neural networks, recurrent neural networks, long short-term memory etc
3-Has best-in-class performance on problems that significantly outperforms other solutions in multiple domains. This includes speech, language, vision, playing games etc[1,2].

VI. DISADVANTAGE

1-Requires a large amount of data.
2-Is extremely computationally expensive to train. The most complex models take weeks to train using hundreds of machines equipped with expensive GPUs.
3-What is learned is not easy to comprehend. Other classifiers (e.g. decision trees, logistic regression etc) make it much easier to understand what’s going on[18,19].

VII. CONCLUSION

This review presents a comprehensive review of deep learning , CNNs and deep belief networks for image classification tasks. It categorizes their progression into their early development, their contribution to the deep learning renaissance, and their rapid advancement over the past few years. In particular, it focuses on their advancement by deliberating and analyzing most of the notable advances in relation to their architectures, supervision components, regularization mechanisms, optimization techniques.
Despite successes in other domains, DCNNs have seen remarkable progression in image classification tasks, setting the state of the art on several challenging classification benchmarks and dominating numerous image-classification-related challenges and competitions.

REFERENCES

1-M. Kukaˇcka ,***“Overview of Deep Neural Networks”,*** WDS'12 Proceedings of Contributed Papers, Part I, 100–105, 2012..
2-Yichuan Tang & Abdel-rahmanMohamed , “***Multiresolution Deep Belief Networks”***, Department of Computer Science ,University of Toronto.
3-Nicolas Le Roux and YoshuaBengio***,” Representational Power of Restricted Boltzmann Machines and Deep Belief Networks”***, Dept. IRO, Universit´e de Montr´eal C.P. 6128, Montreal, Qc, H3C 3J7, Canada.
4-Florian Raudies1, Eric A. Zilli3 and Michael E. Hasselmo***,” Deep Belief Networks Learn Context Dependent Behavior”***, March 2014 | Volume 9 | Issue 3 .
5-Nicolas Le Roux1 and YoshuaBengio2 , ***‘Deep Belief Networks are compact universal approximators’***, 1Microsoft Research Cambridge, University of Montreal.
6-Philippe Hamel and Douglas Eck , ***‘LEARNING FEATURES FROM MUSIC AUDIO WITH DEEP BELIEF NETWORKS’***, we give the results of our features used in genre classifica339 11th International Society for Music Information Retrieval Conference (ISMIR 2010).
7-HussamHebbo & Jae Won Kim***,” Classification with Deep Belief Networks “***.
8-Geoffrey E. Hinton ,SimonOsindero and Yee-Whye The ***,” A fast learning algorithm for deep belief nets”***.
9-Tuomas Nikoskinen ,***“FROM NEURAL NETWORKS TO DEEP NEURAL NETWORKS”***, Mat-2.4108 Independent Research Project in Applied Mathematics , AALTO UNIVERSITY SCHOOL OF SCIENCE , Espoo 2015.
10-Grigorios Tzortzis and AristidisLikas , ***“Deep Belief Networks for Spam Filtering”,*** Department of Computer Science, University of Ioannina GR 45110, Ioannina, Greece.
11-Hanlin Goh , Nicolas Thome, Matthieu Cord and Joo-Hwee Lim , ***“Top-Down Regularization of Deep Belief Networks”*** .
12-Jake Bouvrie ,***“Notes on Convolutional Neural Networks”***, Center for Biological and Computational Learning Department of Brain and Cognitive Sciences Massachusetts Institute of Technology Cambridge, MA 02139.

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 6, Issue 12 , December 2019

13-Yoon Kim ,***"Convolutional Neural Networks for Sentence Classification "***,New York University , Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751,.

14-Quoc V. Le ***, "A Tutorial on Deep Learning Part 2: Autoencoders, Convolutional Neural Networks and Recurrent Neural Networks "***,October 20, 2015.

15-Samer Hijazi, Rishi Kumar, and Chris Rowen***," Using Convolutional Neural Networks for Image Recognition "***,IP Group, Cadence.

16-Andrej Karpathy , George Toderici and SankethShetty , ***"Large-scale Video Classification with Convolutional Neural Networks"*** , Computer Science Department, Stanford University.

17-Quoc V. Le ,***"A Tutorial on Deep Learning Part 1: Nonlinear Classifiers and The Backpropagation Algorithm"***, December 13, 2015.

18-Ian Goodfellow ,YoshuaBengio and Aaron Courville , ***"Deep Learning".***

19-Yann LeCun, YoshuaBengio& Geoffrey Hinton ,***"Deep learning"***, 2015 Macmillan Publishers .