IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-6, Issue-4 E-ISSN: 2347-2693

A Review of Optimization Methods in Deep Learning

Amita Khatana[1*], V.K Narang[2], Vikas Thada[3]

[1] Department of Computer Science, Amity University, Haryana, India
[2] Department of Computer Science, Amity University, Haryana, India
[3] Department of Computer Science, Amity University, Haryana, India

[*]*Corresponding Author: amitakhatana21@gmail.com, Tel.: 9911150040*

Available online at: www.ijcseonline.org

Received: 20/Mar/2018, Revised: 28/Mar/2018, Accepted: 19/Apr/2018, Published: 30/Apr/2018

Abstract— Deep learning technique is an emerging field of machine learning. In recent years, it has been successfully used in different fields, such as image classification, natural language processing, computer vision, speech reorganization, etc. When compared to the machine learning, deep learning has a high learning ability to extract features of large datasets. Deep learning came into existence in 1971 when Ivakhnenka used group method of data handling algorithm (GMDH) to train 8-layered neural network [1]. This paper focuses on the artificial neural network, learning techniques and optimization methods of deep learning like stochastic gradient descent, batch gradient descent, mini-batch gradient descent and ADAM.

Keywords— *Artificial Neural Network, Deep Learning CNN, RNN, Optimization Methods, Gradient Descent, ADAM, Framework,mageClassification.*

I. INTRODUCTION

Machine learning is a technique that enables the computer to learn from datasets. The datasets can be labeled or unlabeled. The machine is trained with the dataset to do work like a human being does. The Deep learning is a advanced field of machine learning that efficiently solves the problem of image classification. Image classification is a technique that classifies images into predefined categories. It is used in many fields like face recognition [7], brain tumor classification [8] etc. The accuracy of classifier is depends upon the training and testing algorithm used.

In image classification, different classifiers are used like arificial neural network (ANN) [19] and support vector machine (SVM) [20] and decisions tree algorithms etc. This paper discusses artificial neural network and different deep learning optimization algorithms for the image classification. The paper is organized into six sections as follows: Section 1 discusses about the image classification, Section 2 shows the working of artificial neural network, Section 3 introduce deep learning, its learning techniques and models, Section 4 describe deep learning optimization methods, and Section 5 shows deep learning frameworks and section 6 is about applications of deep learning.

II. ARTIFICIAL NEURAL NETWORK (ANN)

The artificial neural network is a computational model. The concept of artificial neural network is inspired from the human nervous system. The 'neuron' of artificial neural network is works like a 'neuron' of human nervous system. In biological neuron, dendrites take inputs from another neuron, nucleus performs computation on that input and axon transmits the result to the next neuron. Similarly, in ANN input passes through the activation function. The activation function performs computation on it and sends output to the other neuron.

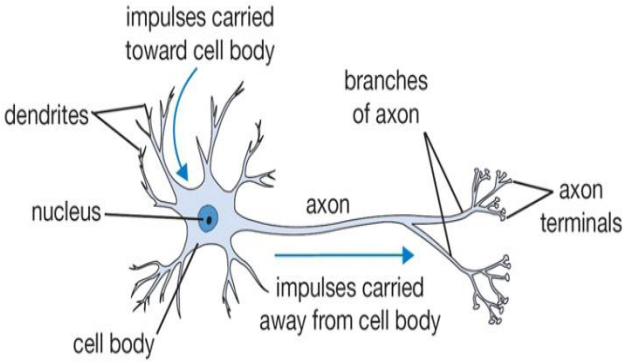

Fig 1: A model of biological neuron [31]

In artificial neural network the first layer neurons are connect with the second layer neurons; the neurons of the second layer are connected to the next layer and so on. The structure of neural network forms a multilayer structure. The artificial neural network mainly has three types of layers, i.e. input layer, an output layer and one or more hidden layer. Each layer consist number of neurons to perform computations. Input layer neurons takes set of inputs that produces outputs, using activation function on a weighted sum of inputs.

In figure 2 x1, x2…xn are the input values and w1, w2, w3…, wn are the weight values associated with them. The activation function performs computation on weights and inputs values, and generates output. Examples of activation functions are sigmoid, hyperbolic tangent, ReLU. Activation function is associated with each neuron of the layers [16].

The artificial neural network is used in a different variety of problems like face recognition, object detection, classification, etc. The ANN has been used efficiently in different areas like voice recognition, image recognition, robotics etc.

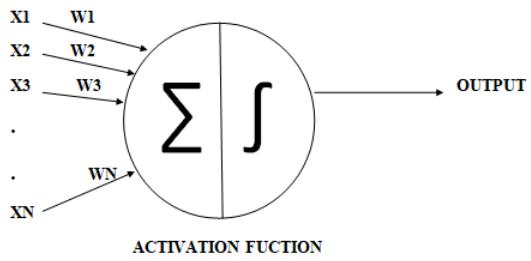

Fig 2: A model of simple artificial neuron

A network is called a fully connected network if its final layer neurons are fully connected. The last layer neurons indicates that model can distinguish or classifies that particular number of image classes in image classification problem, that's why this layer is called classification layer.

Walaa Hussein et al. [8] proposed a neural network to classify MRI brain images. It used the adaptive–dynamic back propagation technique to classify the images. His experiment achieved 96.33 % accuracy for image classification.

III. DEEP LEARNING

Deep learning is a machine learning technique that trains the computer with the help of datasets. In another word, we teach a computer with examples like humans learn through examples or experiences. On the basis of the learning process, machine learning algorithm can be classified into following categories [6].

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

(i) In Supervised learning all data is labeled data. It uses an algorithm that maps input values to corresponding output. In supervised learning a supervisor is present for supervising the learning process [6]. Types of supervised Learning are Classification and Regression.

(ii) In Unsupervised learning, data is unlabeled. In this type of learning process supervisor is absent. In unsupervised learning corresponding output values are not known [6]. Types of unsupervised learning are Clustering and Association.

(iii) In Reinforcement Learning works on mixture of data, some is labeled data and some are unlabeled data. In this type of learning, supervised and unsupervised learning algorithms are used [6].

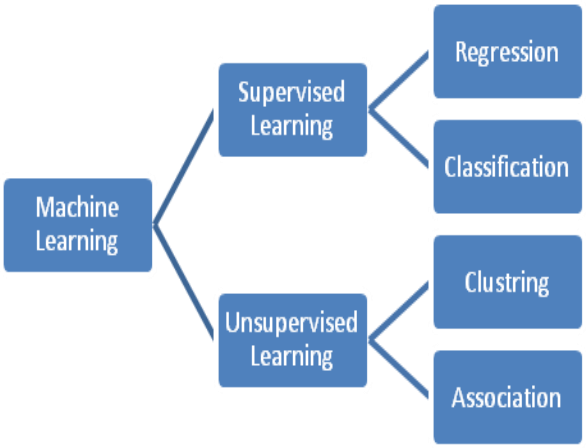

Fig 3: Different types of machine learning techniques.

Deep learning works on a neural network having an input layer, one or more number of hidden layers and an output layer. Input layer neurons take inputs and perform computations on it and passes results to output layer [10].

3.1 Different types of Deep Learning models

A. Deep Belief Network (DBN)

B. Convolutional Neural Network (CNN)

C. Auto encoder (AE)

D. Recurrent Neural Network (RNN)

A. Deep Belief Network (DBN)

The Deep Belief Network is a deep learning model. The Deep Belief Network is a network of the restricted Boltzmann machine (RBM). The restricted Boltzmann machine was developed for Boltzmann Machine. RBM consist of two layers, one is a visual layer and another is the hidden layer. The neurons of the visual layer are connected to neurons of the hidden layer. There is no connection between neurons of the same layer. If RBM having more

number of hidden layers than this machine is called Deep Boltzmann Machine. In this network at a time, only one layer is trained and output of the lower layer is used as input of higher layer. Back Propagation Algorithm is used for updating the network weights.

S. Kim et al. [9] presented a fingerprint liveness detection system that scanned the fingerprints and examines what it is live or fake. This model uses the deep belief network to identify the fingerprints. In this model, the last layer of deep belief network consists of only two output nodes to identify that fingerprint is live or fake. The authors concluded that this model efficiently detects the fingerprints and it helps to identify fake fingerprints.

Gang Liu et al. [10] performed image classification using the deep belief network. This paper proposed a new gradient descent algorithm to classify the images of MNIST dataset. This paper compared the performance of different classifiers and concluded that EGD-based software classifier gave higher classification accuracy.

B. Convolutional Neural Network (CNN)

The Convolutional Neural Network is a model of deep learning. CNN is using to process two dimensional data like images, audio, text, etc. CNN used to perform feature extraction in image classification.

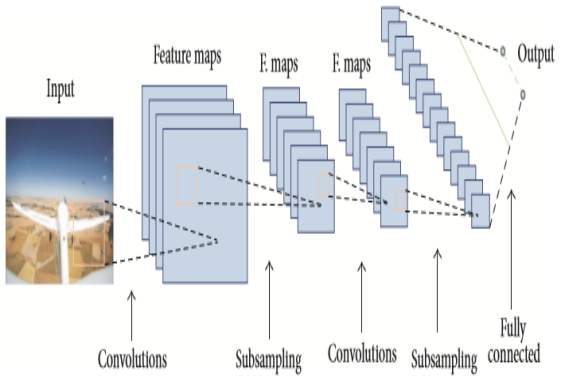

Fig 4: A generic example of a CNN model [6]

CNN model is trained using feed forward, back propagation algorithm with a large data set. Training with large dataset gives higher accurate results. The CNN network gives excellent performance in computer vision application.CNN has demonstrated good performance in face detection [17] and handwritten digit classification [11].

CNN has two types of layers, i.e., convolution layers and pooling layers. The convolution layer detects the co-occurrence of features from their previous layers. The pooling layer semantically merges all similar features into one [3].

Yann.Lecun et al.[11] proposed a CNN model for handwritten digits and alphabets classification. This work used gradient descent algorithm to optimize the result of classification. An MNIST dataset and ASCII dataset are used to train a network. This paper also compared stochastic gradient descent and batch gradient descent optimization algorithms.

Williams et al.[12] proposed a method to perform image classification on MNIST and CIFAR-10 datasets. The classification was performed with the help of wavelets and CNN model. This research work applies ADAM and stochastic optimization algorithm for training of MNIST dataset. In the training of the CIFAR-10 dataset, a stochastic gradient descent was used to train a network. The result showed more accuracy in classification than spatial CNN and SDA.

Abroyan et al.[13] proposed a model to classify real time data. A CNN model and LSTM model was used to classify real time data.

C. Auto encoders (AE)

An autoencoder (AE) is a deep learning model. Autoencoders is used to process high dimensional data. The main aim is to perform dimensionality reduction in data. This means that it changes the dimensions of input data and solves a classification problem efficiently. An improved structure of AE consists are:

- Denoising Auto encoder
- Sparse Autoencoder.

Denoising Auto encoder: Auto encoder takes original data without removing random noise to train a network is called denoising auto encoder. It gives efficient results in feature extraction.

Stacked autoencoder (SAE): SAE is a neural network having multiple layers of sparse auto-encoders. In this network, the output of the previous layer is the input of successive layer. It is used to solve classification problems.

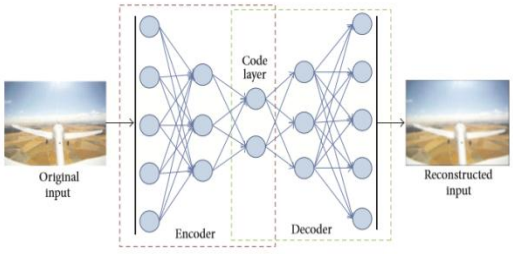

Fig5: Deep Auto-encoder [6]

Xiong et al.[21] proposed an autoencoder network to recognize anomalies in geochemical samples. Author used three steps to train a network, pre-training, unrolling, and fine-tuning using back propagation technique.

D. Recurrent Neural Network (RNN)

Recurrent Neural Network is a deep learning model. It is a feed-forward neural network with direct cycles between layers. These directed cycles circulate the information within the network. In RNN model output of the network is getting changed as information is updated periodically by directing cycles. The final output of the network is not depending upon the initial input. RNN is not effective for long-term memory tasks. For long-term memory tasks, an improved version of RNN is used e.i, Long–Short-Term-Memory (LSTM)

Shi et al.[22]proposed a model to predict a character after some character. This model used the standard RNN and LSTM model to estimate the next character. This paper compared the performance of standard RNN and LSTM model.
Taro et al.[23] proposed a model to determine the behavior of the person in the Tor network. A deep recurrent neural network (DRNN) identified the behavior of the network. This model was used to provide security to the system from unauthorized users.

IV. OPTIMIZATION METHODS

Deep learning has different models to solve different kind of problems. Deep learning works on a neural network, having the number of hidden layers to perform computations. Deep Learning works with large datasets. Training a network with large dataset takes more time to process. To increase the speed of the training process, optimization methods are used. Optimization Algorithm increases the efficiency of the neural network by updating the internal parameters of the network. In this paper, we discuss following optimization methods-

- Gradient Descent and its variants
- ADAM

4.1Gradient Descent

Gradient Descent is most commonly used optimization algorithm. It is used to find out the minimum value of the error function. It speeds up the training process of the deep neural network.
Gradient descent algorithm uses the concept of predictions. It makes a prediction on training data and the error function. It minimizes the error function by updating the values of internal parameters of the network. The cost function of Gradient algorithm is given by J (w, b).

$$J(w,b)=\frac{1}{m}\sum_{i=1}^{m} L\left(a^{(i)},y^{(i)}\right) \quad (1)$$

Where

J (w, b) = Cost function
w = weight value
b = bias value
L = loss value
a = activation function
y= output value
m = number of training set

w = w – alpha * dw (2)
b = b – alpha * db (3)

Where alpha = learning rate and dw is error.

The main goal of the algorithm is to find out the model's parameter like-model coefficients and weights that help to minimize the error on the training sets. On the basis of a number of training patterns used to find out the error and update the models, there are three variations of Gradient Descent.

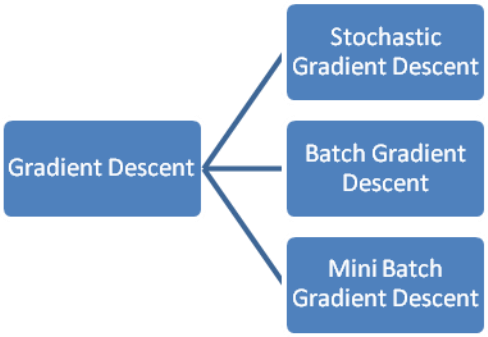

Fig: 6 Variants of Gradient Descent

Types of Gradient Descent-

- Batch Gradient Descent
- Stochastic Gradient Descent
- Mini-Batch Gradient Descent

4.1.1 Batch Gradient Descent

Batch Gradient Descent is one of the variations of the gradient descent algorithm. The batch gradient descent algorithm calculates the error for each example in the training data sets. After evaluating all the training examples it updates the neural model.
To implement batch gradient descent entire training set must be in memory, because the model is updated after completion of the training process. This algorithm becomes slow for large datasets.

4.1.2 Stochastic Gradient Descent

Stochastic Gradient Descent is another variation of gradient descent. The stochastic gradient descent algorithm examines all the examples of training datasets and calculates the error for each example. After calculating error it updates the model for each example of training data sets. Stochastic gradient descent updates weights using single learning rate, i.e. alpha

learning. The learning rate is not changed in the learning process of the network.
The frequent updates of the model to improve performance on one hand, but on the other hand, results in a noisy gradient signal. This noisy learning process creates a problem in minimization of an error function.

4.1.3 Mini–Batch Gradient Descent

Mini-Batch Gradient Descent is another variation of gradient descent. In this algorithm, the complete training dataset is divided into small baby sets called mini-batch training set. The error is calculated for each mini-batch training sets and model coefficient according to the requirements.
Mini-Batch gradient descent uses the robustness of stochastic gradient descent and efficiency of batch gradient descent. It combines the advantages of both algorithms.

4.2 ADAM

ADAM stands for Adaptive Moment Estimation. It is improved version of Stochastic Gradient Descent. ADAM used adaptive learning rate to iteratively update the weights in the network. ADAM includes the benefits of AdaGrad and RMS prop.

Adaptive Gradient Algorithm (AdaGrad): AdaGrad algorithm is used to solve problems in natural language processing and computer vision having the sparse gradient. This algorithm is based on per-parameter learning rate to optimize the performance.

Root Mean Square Propagation (RMSProp): RMSProp algorithm also operates on per-parameter learning rate, but it is different from the AdaGrad algorithm. The algorithm uses the moving average of the squared gradient. A rooted value is used to change the updated weights and the bias value, which helps to change the processing speed in the horizontal direction and vertical direction for getting gradient value. RMSProp works well with non-stationary settings. Here are the equations according to RMSProp.

$$w = w - \alpha \frac{dw}{\sqrt{sdw}} \qquad (4)$$

$$b = b - \alpha \frac{db}{\sqrt{sdb}} \qquad (5)$$

According to ADAM algorithm the value of w and b are as following:

$$w = w - \alpha \frac{Vdw}{\sqrt{sdw} + \epsilon} \qquad (6)$$

$$b = b - \alpha \frac{Vdb}{\sqrt{sdb} + \epsilon} \qquad (7)$$

β1 -This is used for decaying the running average of the gradient (default = 0.9).
β2 -This is used for decaying the running average of the square of gradient (default = 0.999)
α –learning rate
€ -It is to prevent the division from zero error.
Vdw – Corrected value of dw.
Vdb – Corrected value of db.

Dabrowski et al.[23] proposed a model that used ADAM and Stochastic Gradient Descent optimization algorithm to optimize the results of image classification. This mode compared the performance of ADAM and Stochastic Gradient Descent optimization algorithm.

V. APPLICATION OF DEEP LEARNING

Deep Learning technique is used to successfully solve the problem in different fields such as computer vision, image classification, object detection, speech reorganization, natural language processing. This section discusses different paper's research work in image classification field.
Dabrowski et al.[23] proposed a model for image classification using deep learning. The CNN model with transfer learning technique is used to classify real-life images with higher accuracy.
Ibrahim et al.[8]present a model of a neural network to classify images of MRI images. This model plays important role in medical science.
Wang et al.[25] present a model for classification of X-Ray scattering images. This work is obtained with the help of convolutional neural network (CNN) and convolutional autoencoder model.
Mohman et al [26] reviewed different methods used for classification of satellite images. This paper discussed artificial neural network using back prorogation and K-means algorithm for classification of satellite images.
Jyothi et al [27] proposed a model to recognize the paper currency. An artificial neural network model was used to reorganization the paper currency with minimal power consumption.

VI. DEEP LEARNING FRAMEWORK

Deep learning frameworks are a platform to implement models for different applications like image classification, natural language processing, computer vision, etc. Following are the most commonly used frameworks: Tensor Flow, Torch and Theano Caffe etc. These frameworks consist libraries, methods, tools that help to develop models to trained a network
Tensor Flow [28] is a machine learning platform having different algorithm to perform different machine learning

tasks. It describes computations in dataflow-like diagrams. It supports single-device, multi-device,and distribution execution. It is a python based platform. It is used for both R&D and development purpose.

Torch [29] is a platform that includes most popular algorithms and models to perform machine learning. It can be embedded into iOS, Android, etc. Following are the famous algorithms that are supported by Torch – Support Vector Machine (SVM), Bayes Classifiers, AdaBoost, CNN, Hidden Markov Models etc. It is an R &D framework for deep learning applications.

Theano [30] supports all three types of learning methods, i.e., supervised, unsupervised and semi-supervised learning methods and approaches, but in comparing to others its processing speed is slow. It is a python based framework and used only for R&D purpose.

Author and Publication	Models Used	Task	Reference
Xuedam, Conference Paper, 2016	Review paper Overview of deep learning methods and application	Discuss different papers and review them.	[2]
Gang Liu, IEEE 2017.	Deep belief Neural Networks.(MNIST dataset)	Used improved gradient descent for image classification	[10]
Henrik Petersson, IEEE 2016.	Review Deep Learning Models	A review on Hyper spectral Image analysis	[5]
Adrian Carrio, Hindawi Journal of Sensors 2017.	Review paper	Review learning method and applications for aerial vehicles	[6]
Walaa Hussein,ICCEEE, 2013.	Neural network	MRI image classification	[8]
S.Kim, Pattern Recog. Let., 2016.	Deep belief network	Finger print detection	[9]
Y. Lecun, IEEE, 1998.	Gradient descent optimization-algorithm	Document Reorganization	[11]
Travis Williams, IEEE 2016.	CNN (MNIST and CIFAR dataset)	Image classification	[12]
Narek Abroyan , INTECH 2017.	CNN and Recurrent Neural Network	Real time data classification	[13]
Marek Dabrowski,FedCSIS 2016.	CNN	Image Classification on real life photo	[16]
Zejian Shi, IEEE , 2017	Recurrent Neural Network and LSTM	Character Reorganization	[22]
Taro Ishitakl, Conference paper, 2017.	Deep Recurrent Neural Network	Security purpose. (Identify the user behavior in network)	[23]
R.Jyothi, ICEEOT, 2016.	ANN	Paper Currency Reorganization	[27]
B. Wang, IEEE, 2017.	CNN and Autoencoder	X- Ray images Classification	[25]

TABLE1: A short summary of research in deep learning and image classification.

VII. CONCLUSION

In recent years, machine learning technique is widely used for the image processing. Deep learning is a subfield of machine learning that contains different algorithms for computations. This paper discussed different deep learning models and optimization methods to improve the performance of network. Gradient descent, mini-batch gradient descent, stochastic gradient descent and ADAM, RMSProp etc optimization methods are explained in this research work. The paper also maintains various frameworks like theano, tensorflow etc and applications of deep learning in image processing.

Future research in the field of optimization algorithm can increase the efficiency and speed of the training process of the neural network. The review done in this paper on optimization algorithms will help us in implementing the same in future research work. The research work will use these algorithms for comparing the accuracy of different datasets, as applied to image classification using deep learning.

REFERENCES

[1] A.G. Ivakhnenko, "Polynomial theory of complex systems" IEEE Transaction on System, Man and Cybernetic vol. 1,no 4, pp. 364-378, 1971.

[2] Xuedam Du,Yinghao Cai, Wang, and Leijie Zhang " Overview of Deep Learning" 31st Youth Academic Annual Conference of Chinese Association of Automation Wuhan China November 11-13-2016.

[3] Siddhartha Sankar Nath, Janynyaseni Kar, Girish Mishra, Sayan Chakraborty , Nilanjan Dey " A Survey of Image Classification Methods and Techniques " ICCCICCT 2014.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". Neural Information Processing Systems, Nevada, 2012

[5] Henrik Petersson, David Gustafsson and David Bergstroom " Hyperspectral Image Analysis using Deep Learning- a Review" IEEE 2016.

[6] Adrian Carrio, Carlos Sampedro, Alejandro Rodriguez Ramos and Pascual Campoy "A Review of Deep Learning Methods and Applications for unmanned Aerial Vehicles " Hindawi Journal of Sensors 2017.

[7] H. Kamitomo and C. Lu, "3-d face recognition method based on optimum 3-d image measurement technology," Artificial Life and Robotics, vol. 16, no. 4, pp. 551–554, 2012.

[8] Walaa Hussein Ibrrahim, Ahmed AbdelRhman Ahmed Osman, Yusra Ibrahim Mohamad" MRI Image Classification Using Neural Network" ICCEEE, 2013.

[9] S.Kim, B.Park, B.S Song, and S.Yang, " Deep belief network based statistical feature learning for fingerprint liveness detection, " Pattern Recog. Lett., vol 77, ,pp. 58-65,2016.

[10] Gang Liu, Liang Xiao, Caiquan Xiong " Image Classification with deep belief network and improved gradient descent" IEEE 2017.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradientbased learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] Travis Williams, Robert Li " Advanced Image Classification using Wavelets and Convolution Neural Network" IEEE 2016.

[13] Narek Abroyan , " Convolutional and Recurrent Neural Network for real time data classification" The Seventh International Conference on Innovative Computing Technology (INTECH 2017).

[14] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, "Deep Learning", Book in preparation for MIT Press, 2016, on-line version available at http://www.deeplearningbook.org.

[15] Michael A.Nielsen, "Neural Networks and Deep Learning", Determination Press, 2015.

[16] Marek Dabrowski, Justyna Gromada, Tomasz Michalik Orange Centrum "A Practical study of neural network –based image classification model trained with transfer learning method" FedCSIS 2016.

[17] C. Lu and X. Tang, "Surpassing human-level face verification performance on lfw with gaussianface," arXiv preprint arXiv:1404.3840, 2014.

[18] L. Deng and D. Yu, "Deep learning: methods and applications," Foundations and Trends in Signal Processing, vol. 7, no. 3-4, pp. 197–387, 2013.

[19] J. A. Hertz," Introduction to the theory of neural computation." Boulder, USA: Westview Press, 1991.

[20] J. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," Neural Processing Letters, vol. 9, no. 3, pp. 293–300, 1999.

[21] Y. Xiong and R.Zuo, " Recognization of geochemical anomalies using a deep autoencoder network,"Computer Geosci-UK, vol.86, pp. 75-82, 2016.

[22] Zejian Shi, Minyong Shi, Chunfang Li," The prediction of character based on recurrent neural network," IEEE computer society, Wuhan China, 2017

[23] Taro Ishitakl, Ryolchlro Obukata, Tetsuya Oda, Leonard Baroll, "Application of deep recurrent neural network for prediction of user behavoiur in Tor Network," 31st International Conference on Advanced Information Networking and Application Workshops, 2017.

[24]Marek Daabrowski, J. Gromada, T. Michalik," P practical study of neural network-based image classification model trained with transfer learning method," Federated Conference on Computer Science and Information Systems pp. 49-56, 2016.

[25] B. Wang, K. Yager, D.Yu, Minh Hoai," X- ray Scattering image classification using deep learning," IEEE Winter Conference on Application of Computer Science,2017

[26] Nur Anis Mohmon and Norsuzila ya acob," A review on classification of satellite image using artificial neural network (ANN)," IEEE 5th Control and System Graduate Research Colloquium,2014.

[27] R.Jyothi, Y.K. SundaraKrishna, V. Srinivasa Rao," Paper Currency recognition for color images based on Artificial Neural Network," International Conference on Electrical , Electronics and Optimization Techniques (ICEEOT), 2016.

[28]M.Abadi, A. Agarwal, p. Barham, E. Brevdo, Z.f. Chen, C. C itro,et at.," Tensorflow : Large-scale machine learning on heterogeneous distributed systems," arXiv preprint arXiv: 1603.04467,2016.

[29]R. Collobert, S.Bengio, and J.Mariethoz, " Torch: a modular machine learning software library," Idiap, No.EPFL-REPORT-82802, 2002.

[30]R. AI-Rfou, G.Alain, A.Almahairi el at.," Theano ; a python framework for fast computation of mathematics expression," arXiv preprint arXiv : 1605.02688,2016.

[31]http://www.wpclipart.com/medical/anatomy/cells/neuron/neuron.png.html

