A Review on Deep Learning Architectures and Applications: Image Recognition

Yawar Rasool Mir
Department of Computer Science & Engineering
Desh Bhagat University
Punjab, India

Navneet Kaur Sandhu
Department of Computer Science & Engineering
Desh Bhagat University
Punjab, India

Abstract : Deep learning (DL) today is being used for various real world problems such as cancer diagnosis, precision medicine, self-driving cars, predictive forecasting and speech recognition. DL can overcome the limitations of earlier shallow networks that prevented efficient training and abstractions of hierarchical representations of multi-dimensional training data The review covers different types of deep architectures, such as deep convolution networks, deep residual networks, recurrent neural networks and others. The review also covers one important application of deep learning for Image Recognition.

***Indexterms*— Deep Learning, Convolution Neural Network, Recurrent neural networks.**

I. INTRODUCTION

Deep Learning is the subpart of Machine Learning and it is a technique for learning a deep neural network which is composed of many layers. The important reasons for the popularity of deep learning are the drastic increase in chip processing abilities, increasing dataset sizes, and the recent advances in machine learning research[1,2].Simple machine learning algorithms work well with structured data. But when it comes to unstructured data, their performance makes to take quite a dip. This is where neural networks have proven to be effective and useful. They perform exceptionally well on unstructured data. Most of the ground-breaking research these days has neural networks at it's core. As the amount of data increases, the performance of traditional learning algorithms, like SVM and logistic regression, does not improve by a whole lot. In fact, it tends to plateau after a certain point. In the case of neural networks, the performance of a model increases with increase in the data you feed to the model.

There are basically three scales which drive a typical deep learning process:

1. Data 2. Computation Time 3. Algorithms

II. DEEP NEURAL NETWORK ARCHITECTURES

Deep neural network consists of several layers of nodes. Different architectures have been developed to solve problems in different domains or use-cases. E.g., CNN is used most of the time in computer vision and image recognition, and RNN is commonly used in time series problems/forecasting. On the other hand, there is no clear winner for general problems like classification as the choice of architecture could depend on multiple factors[3]. Below are three of the most common architectures of deep neural networks.

1. Convolution Neural Network.
2. Autoencoder.
3. Restricted Boltzmann Machine (RBM).
4. Long Short-Term Memory (LSTM).

A. CONVOLUTION NEURAL NETWORK

CNN is based on the human visual cortex and is the neural network of choice for computer vision (image recognition) and video recognition. LeCun et al[4,5] introduced CNN for handwritten digits classification. CNN consists of two core structures: a convolutional layer and a pooling layer as shown in Fig 1. Units in a convolutional layer are organized in feature maps, within each unit is connected with local patches in the feature maps of the previous layer through a set of weights called filter bank. The result of this local weighted sum is then passed through a non-linearity such as a ReLU. The role of pooling layer is to merge semantically, similar features into one. Because the relative positions of features forming a motif can vary somewhat, reliably detecting the motif can be done by coarse-graining a position of each feature[5]. The convolutional layer shares more weights, and the pooling layer sub-samples the output of a convolutional layer and reduces the data rate from the layer below.

Here are the well-known variation and implementation of the CNN architecture.

1. AlexNet:

CNN developed to run on Nvidia parallel computing platform to support GPUs.

2. Inception:

Deep CNN developed by Google.

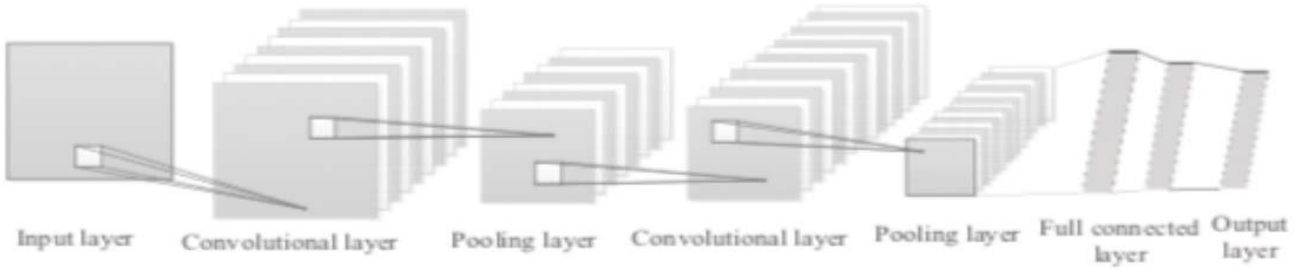

Fig. 1 : Convolution Neural Network architecture

3. ResNet:
Very deep Residual network developed by Microsoft. It won 1st place in the ILSVRC 2015 competition on ImageNet dataset.
4. VGG:
Very deep CNN developed for large scale image recognition.
5. DCGAN:
Deep convolutional generative adversarial networks proposed by [6]. It is used in unsupervised learning of hierarchy of feature representations in input objects.

B. AUTOENCODER

Autoencoder is the neural network that uses unsupervised algorithm and learns the representation in the input data set for dimensionality reduction and to recreate the original data set. The learning algorithm is based on the implementation of the backpropagation.

Autoencoders extend the idea of principal component analysis (PCA). Autoencoders use encoder and decoder blocks of non-linear hidden layers to generalize PCA to perform dimensionality reduction and eventual reconstruction of the original data. It uses greedy layer by layer unsupervised pretraining and n-tuning with backpropagation [10]. Figure 2 illustrates a simplified representation of how autoencoders can reduce the dimension of the input data and learn to recreate it in the output layer.

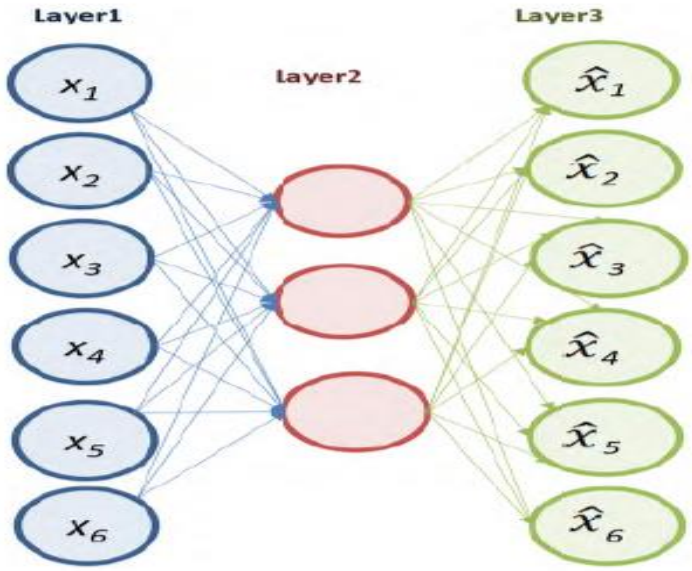

Figure 2. Autoencoder nodes.

C. RESTRICTED BOLTZMANN MACHINE (RBM)

Restricted Boltzmann Machine is an artificial neural network where we can apply unsupervised learning algorithm to build non-linear generative models from unlabeled data [7]. The goal is to train the network to increase a function (e.g., product or log) of the probability of vector in the visible units so it can probabilistically reconstruct the input. It learns the probability distribution over its inputs. As shown in Figure 3, RBM is made of two-layer network called the visible layer and the hidden layer. Each unit in the visible layer is connected to all units in the hidden layer and there are no connections between the units in the same layer.

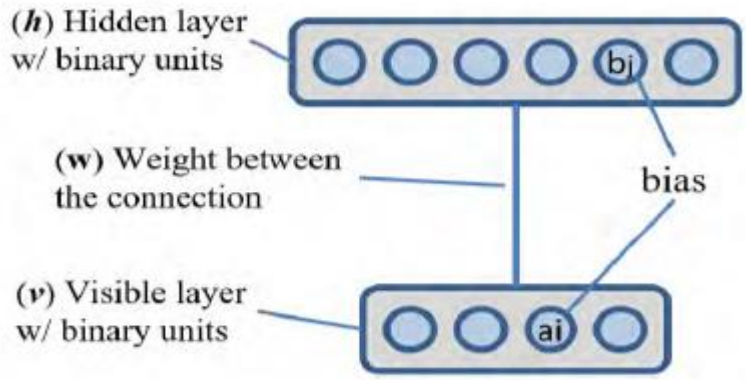

Figure 8. Restricted Boltzmann machine.

D. LONG SHORT-TERM MEMORY (LSTM)

LSTM is an implementation of the Recurrent Neural Network and was first proposed by Hochreiter et al. In 1997 [8]. LSTM can retain knowledge of earlier states and can be trained for work that requires memory or state awareness. LSTM partly addresses a major limitation of RNN, i.e., the problem of vanishing gradients by letting gradients to pass unaltered. As shown in the illustration in Figure 4, LSTM consists of blocks of memory cell state through which signal flows while being regulated by input, forget and output gates. These gates control what is stored, read and written on the cell. LSTM is used by Google, Apple and Amazon in their voice recognition platforms [9].

In figure 4, C, x, h represent cell, input and output values. Subscript t denotes time step value, i.e., t − 1 is from previous LSTM block (or from time t − 1) and t denotes current block values. The symbol σ is the sigmoid function and tanh is the hyperbolic tangent function. Operator + is the element-wise summation and x is the element-wise multiplication.

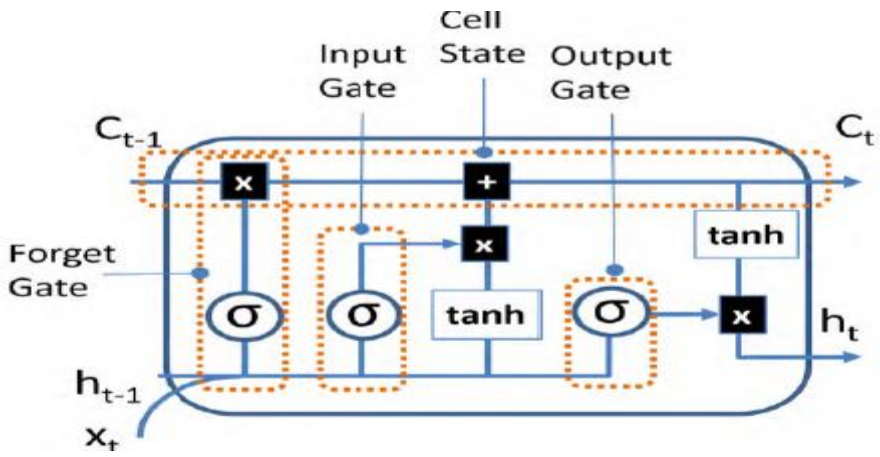

Figure 4. LSTM block with memory cell and gates.

E. COMPARISON OF DNN NETWORKS

Table 1 provides a compact summary and comparison of the different DNN architectures. The examples of implementations, applications, datasets and DL software frameworks presented in the table are not implied to be exhaustive.

III. DEEP LEARNING APPLICATIONS

Few of newer and recent application developments of deep learning are elaborated in examples below:

A. One example of an application of deep learning in Big Data is Microsoft speech recognition (MAVIS). Using deep learning enables searching of audios and video files through human voices and speeches [11].

B. Deep learning on Big Data environment is used by Google for image search service. They used deep learning for understanding images so that it can be used for image annotation and tagging that is further useful in image search engines and image retrieval as well as image indexing [11].

C. In 2016, Google's AlphaGo program defeated Lee Sedol in Go competition, which showed that deep learning had a strong learning ability.

Table 1: DNN network comparison table.

Network Type	Architecture	Network Model	Training Type	Training Algorithm	Implementation Sample	Common Application	Popular Dataset Sample	DL Framework (sample)
Feedforward Neural Network	CNN	Discriminative	Supervised	Gradient Descent based Backpropagation	Siamese Network, Deep CNN	Image recognition/classification	MNIST	TensorFlow, Caffe, Theano, Torch, Deeplearning4j, Microsoft Cognitive Toolkit, Keras, MXNet, PyTorch
	Residual Network	Discriminative	Supervised	Gradient Descent based Backpropagation	Deep ResNet; HighwayNet; DenseNet	Image recognition	ImageNet	TensorFlow, PyTorch, Keras
	Autoencoder	Generative	Unsupervised	Backpropagation	Sparse Autoencoders, Variational Autoencoders	Dimensionality Reduction; Encoding	MNIST	TensorFlow, Deeplearning4j, Keras
	Adversarial Networks	Generative & Discriminative	Unsupervised	Backpropagation	Generative Adversarial Network	Generate realistic fake data; Reconstruction of 3D models; Image improvement	CIFAR10	TensorFlow, Keras
	RBM	Generative with Discriminative finetuning	Unsupervised	Gradient Descent based Contrastive divergence	Deep Belief Network; Deep Boltzmann Machine	Dimensionality Reduction; Feature learning; Topic modeling	MNIST	TensorFlow, Deeplearning4j, Keras, MXNet, Theano, Torch
Recurrent Neural Network	LSTM	Discriminative	Supervised	Gradient Descent & Backpropagation through Time	Deep RNN, Gated Recurrent Unit (GRU), Neural Machine Translation (NMT)	Natural Language Processing; Language Translation	MNIST Stroke Sequence	TensorFlow, Caffe, Theano, Torch, Deeplearning4j, Microsoft Cognitive Toolkit, Keras, MXNet, PyTorch
Radial Basis Function NN	RBF Network	Discriminative	Supervised and Unsupervised	K-means Clustering; Least Square Function	Radial Basis Function NN	Function approximation; Time series prediction	Fisher's Iris data set	TensorFlow
Kohonen Self Organizing NN	Nodes arranged in hexagonal or rectangular grid	Generative	Unsupervised	Competitive Learning	Kohonen Self Organizing NN	Dimensionality Reduction; Optimization problems; Clustering analysis	SPAMbase	TensorFlow

D. Google's Deep Dream is software which can not only classify images but also generate strange and artificial paintings based on its own knowledge.

E. Facebook announced a new artificial intelligence system named Deep Text. It is a deep learning-based text understanding engine which can classify massive amounts of data, provide corresponding services for identifying users chatting messages and clean up spam messages.

- Computer vision, prediction, semantic analysis, natural language processing, information retrieval and customer relationship management are the application domains for deep learning.

- Computer vision has object recognition, object detection and processing as sub-domains. It includes automatic speech recognition, image recognition, speech and audio processing and visual art processing which newer applications are being explored using deep learning.
- Prediction: The different sub-domains here are classification, analysis, and recommendation. Drug discovery and toxicology, Bioinformatics, mobile advertising are newer applications being developed using deep learning.
- Semantic Analysis and Natural Language Processing and Information Retrieval: These are three areas in which machine learning techniques have been explored in the past but additionally now researchers are applying deep learning techniques.
- Customer Relationship Management: Given customer's history, data analysis is done which is used to enhance business relationships. In this case, deep learning can be useful. These above mentioned deep learning applications have been presented in Figure 5 below.

Also, we would like to describe different deep learning approaches which have been used in the field of image processing and pattern recognition.

A. AlexNet

It contains mainly 5 convolution layers, maxpooling layers, dropout layers and 3 fully connected layers. In AlexNet Rectified linear unit(Relu) is used, instead of tanh or sigmoidal function which is basically used as the

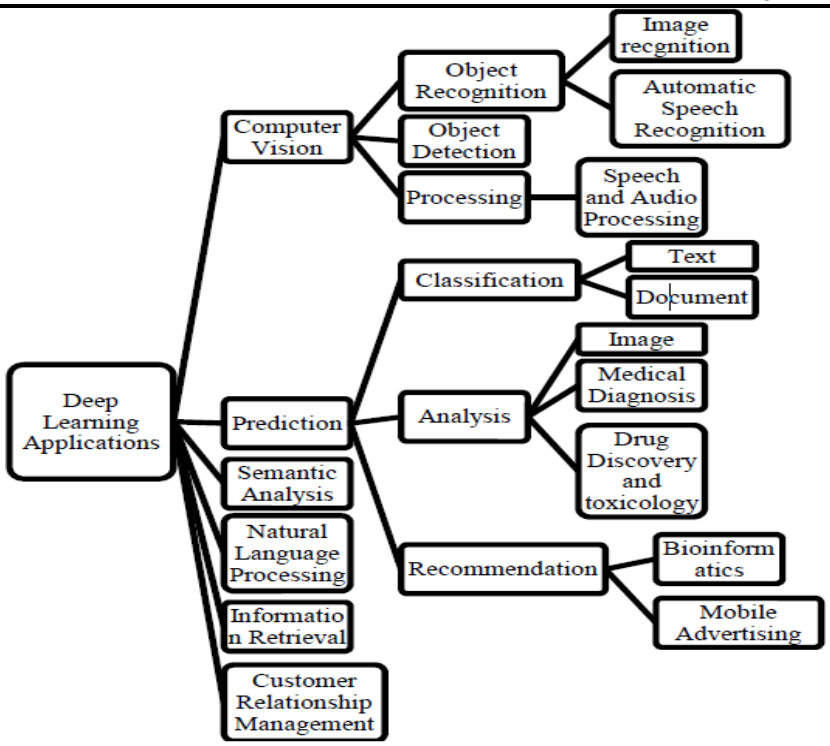

Figure 5. Deep learning applications

traditional activation function. AlexNet is mainly used with the large number ImageNet dataset, which has 15 million annotated images from the total 22000 categories. It used for the Data augmentation techniques.

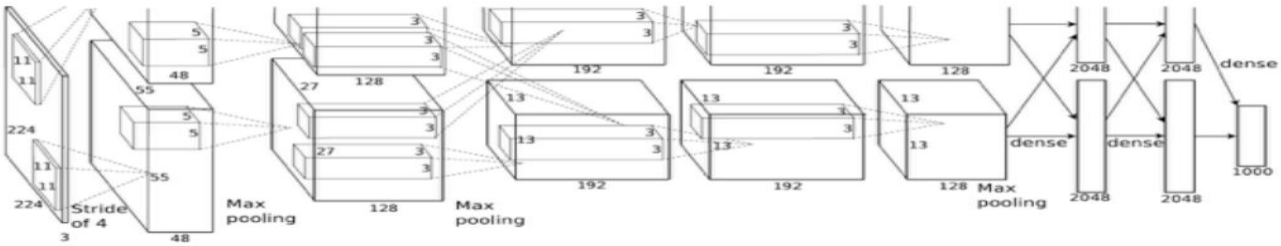

Figure 6. AlexNet Architecture

B. VggNet

Model created in 2014 with the best utilization and lowest error rate. Karen Simonyan and Andrew Zisserman from the Oxford University, created the model with 19 layer of CNN that uses 3x3 filters with stride and pad of 1, along with the 2x2 maxpooling layer with 2 strides[13].

It has 3x3 filters which is quite differ from the Alexnet's 11x11 filter in the first layer.Main thing about the VggNet is that it strengthened the idea that convolution neural networks have to have a deep network of layers in order for the hierarchical representation of the visual data, which keeps it simple and deep.

C. Inception/GoogleNet

"GoogleNet" which is basically a Inception architecture in inception modules are used[14]. It has more depth and large Inception network to improvised the result marginally. therefore it is the most efficient framework of deep learning network. There is inadequate connection between the activations, which means that all the 512 output channels does not have connection with all those 512 channels provide as input channels.

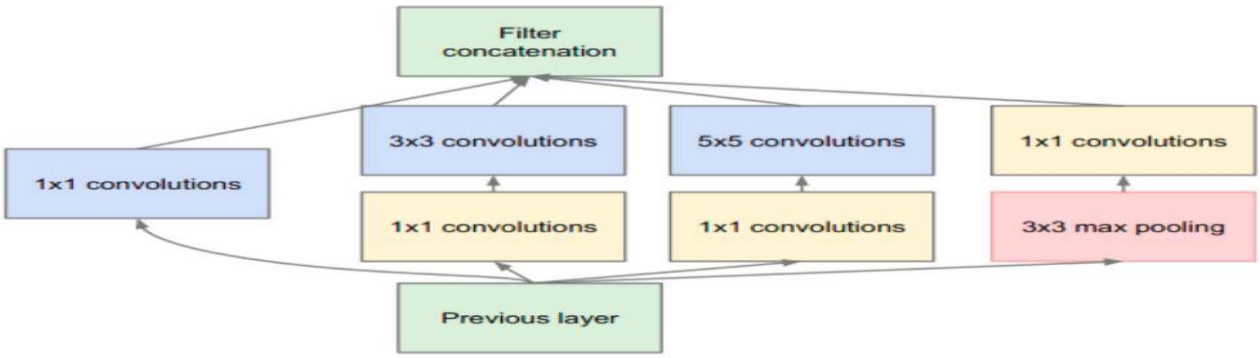

Figure 7 : Inception module in GoogleNet

D. SegNet (For Segmentation)

SegNet, which is basically performed pixel-wise segmentation for the input image for example, it performs labelling each pixel of an image to belong some class as shown in the following figure 8. it can generate the labelling for the different class such as car, tree, road from given input image. In SegNet, mainly two components are used : encoder and decoder[15][16].

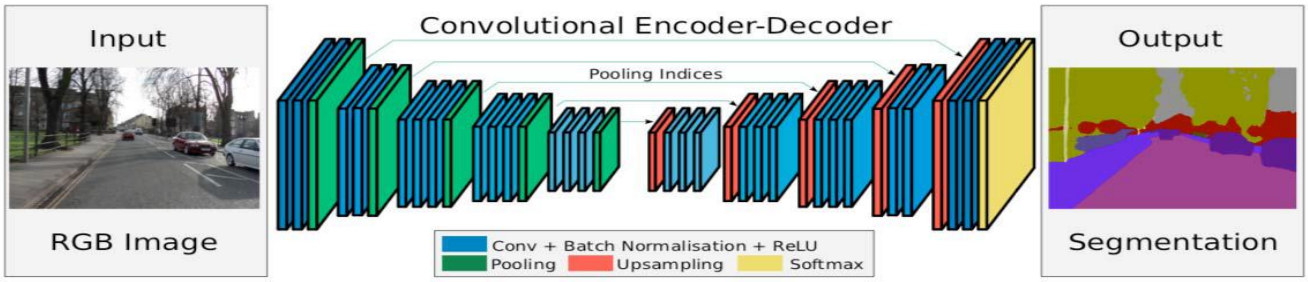

Figure 8 : SegNet model architecture.

SegNet provides better accuracy in classification but it reduces the feature map size, which leads it to detrimental image representation with dazzled boundaries. Due to this reason decoder is used to upsampling and store the information[17].

IV. CONCLUSION

The primary reason why deep learning is suited for newer areas of applications is data dependencies, GPU hardware, and feature engineering. Data dependencies term is to refer to deep learning algorithms work well on a huge amount of data. GPU is a high-end machine which stands for Graphics Processing Unit. The distinctive part of deep learning when compared to machine learning is the ability to learn high-level features from data called as feature engineering. Therefore, there could be many more areas of applications of deep learning which will be seen in forthcoming years.

With this review of applications, now we can explore any one of the newer areas of application of deep learning which will yield better results and will add on to the ongoing research in this field. There is even scope for evolving new architectures for deep learning as research is still going on in the early stage. Apart from this enhancements can be done in analysis and prediction sub-domain.

V. REFERENCES

[1]I. Goodfellow, Y. Bengio, and A. Courville, Deep learning, The MIT Press, Cambridge, Massachusetts, 2016.

[2] L. Deng, and D. Yu, Deep learning:methods and applications, NOW PUBLISHERS, 2014

[3] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, ‘‘Do we need hundreds of classifiers to solve real world classification problems?’’J. Mach. Learn. Res., vol. 15, no. 1, pp. 3133–3181, 2014

[4] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, and L.D. Jackel, “Backprogagation applied to handwritten zip code recognition,” Neural Computation, vol. 1, pp. 541-551, 1989.

[5]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, pp. 2278-2324, 1998.

[6] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, pp. 436-444, 2015.

[6]http://www.image-net.org/challenges/LSVRC/

[7] Y. W. Teh and G. E. Hinton, ‘‘Rate-coded restricted Boltzmann machines for face recognition,’’ in Proc. Adv. Neural Inf. Process. Syst., 2001,pp. 908–914

[8] S. Hochreiter and J. Schmidhuber, ‘‘Long Short-term Memory,’’NeuralComput., vol. 9, no. 8, pp. 1735–1780, 1997.

[9] C. Metz, ‘‘Apple is bringing the AI revolution to your phone, in wired,’’Tech. Rep., 2016.

[10] K. Noda, ``Multimodal integration learning of object manipulation behaviors using deep neural networks," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nov. 2013, pp. 17281733. [35] K. Noda, ``Multimodal integration learning of object manipulation behaviors using deep neural networks," in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst., Nov. 2013, pp. 17281733.

[11] M. Gheisari, G. Wang, and M. Z. A. Bhuiyan, “A Survey on Deep Learning in Big Data,” in 22017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017.

[12] N. M. Elaraby, M. Elmogy, and S. Barakat, “Deep Learning: Effective Tool for Big Data Analytics”, International Journal of Computer Science Engineering (IJCSE), Sep 2016.

[13] Karen Simonyan and Andrew Zisserman,"Very Deep Convolutional Networks for Large-Scale Image Recognition",arxiv eprinted : 1409.1556, 2014

[14] Christian szegedy, Wei liu, Yangging Jia, Pierre Sermanet and Scott E. Reed etc., "Going Deeper with Convolutions",arxiv eprinted: 1409.4842, 2014

[15] Vijay Badrinarayanan, Alex Kendall and Roberto Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image segmentation",arxiv eprinted: 1511.00561, 2015

[16] Alex Kendall, Vijay Badrinarayanan and Roberto Cipolla "Bayesian SegNet: Model Uncertainty in Deep Convolutional Encoder-Decoder Architectures for Scene Understanding."

[17] Vijay Badrinarayanan, Ankur Handa and Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling."arXiv preprint arXiv:1505.07293, 2015.