Deep Learning based Computer Vision: A Review

Ashwini Patil[1], Prof. Amit Zore[2]
[1]*PG Student, Computer Engineering Dept., Dhole Patil College of Engineering, Pune*
[2]*Head of Dept., Computer Engineering Dept., Dhole Patil College of Engineering, Pune*

***Abstract*- Computer vision tasks include methods for acquiring, processing, analyzing and understanding digital images, and extraction of high-dimensional data from the real world in order to produce numerical or symbolic information. It may be in the forms of decisions. The feature extraction is strongly carried out by the deep learning with promising benefits, it has been broadly utilized as a part of the field of computer vision and among others, and step by step supplanted conventional machine learning algorithms. This work first presents state of the art of deep learning in connection with computer vision. Later it introduces deep learning concept and methods of deep learning. It then focuses some of the computer vision applications including face recognition, object recognition and activity recognition.**

Index Terms- Computer Vision, Convolutional neural network, Deep learning.

I. INTRODUCTION

The computer vision field is greatly influenced by deep learning. This is because of high-speed computing computer hardware and the availability of labeled image datasets [3].

The computer vision basically enables the vision properties on a computer. The examples of a computer may be in the form of CCTV, drones, smartphones etc. The output of sensor is in the form of a digital form and computers interpret this form.

The deep learning methods solves the different problems occurred in computer vision [12]. For digit recognition, Fukushima Neocognitron [7] and LeCun's neural networks [8] were presented.

Deep learning entered in in computer vision research field after development of AlexNet model [9]. Convolutional Neural Networks (CNNs), Restricted Boltzmann Machines (RBMs), Deep Belief Nets (DBNs), and Autoencoders (AEs) outperformed in computer vision applications like surveillance, remote sensing etc [3].

In the last few years, artificial intelligence and deep learning attracted people's attention in the fields of biology, physics, engineering, and manufacturing. The tools such as convolutional neural networks (CNN), full convolutional networks (FCN), Generative Adversarial Networks (GAN), Google's Dropout became very popular in AI and deep learning community [1].

Deep learning (DL) is a subfield of machine learning. The various DL algorithms proposed to solve traditional AI problems. By adopting hierarchical architectures the DL algorithms learn high-level abstractions. DL is a promising filed and now has been widely used in traditional AI domains such as natural language processing, computer vision etc.

Basic reasons of DL becoming very popular today are, i] enhanced chip processing capability with GPU development ii] decreased cost of computing hardware and iii] noticeable development in the machine learning algorithms [6].

In recent years, Artificial Intelligence (AI), Deep Learning (DL) and Machine Learning (ML) are highly researched topics in this cutting edge. Basically AI deals with the theory and development of computers or machines to perform tasks like human and such tasks require human intelligence. The tasks may be language interpretation, speech recognition, visualization and decision making etc [5].

There are two crucial ideas which motivated deep learning to be used for computer vision as: convolutional neural networks (CNN) and backpropagation. For timeseries deep learning applications, the Long Short-Term Memory (LSTM) algorithm was developed [2].

II. DEEP LEARNING REALM

A. Basics of deep learning

The AI can be achieved through machine learning and deep learning [14]. Fig. 1 depicts a Venn diagram for deep learning.

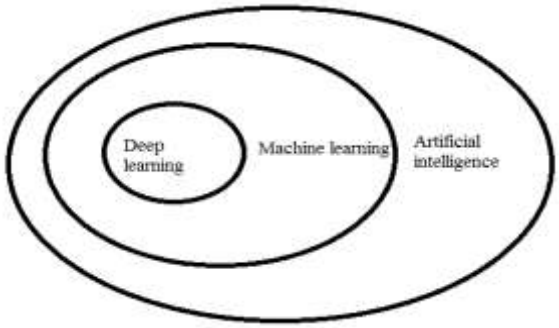

Fig. 1 A Venn diagram for deep learning [14]

Artificial neuron

An artificial neuron is also called as a perceptron. It takes several inputs and performs a weighted summation to produce an output. During training process the weight of the neuron or perceptron is determined and it is based on the training data [12]. Fig. 2 indicates a perceptron/artificial neuron.

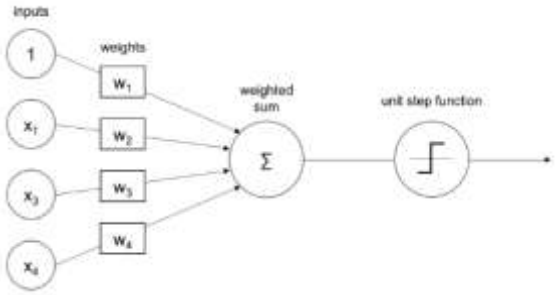

Fig. 2 A Perceptron [12]

Activation functions

The neural nets can be made nonlinear using the activation functions. It decides whether a neuron or perceptron should fire or not [12]. The gradients can be adjusted by functions. Fig. 3 shows different activation functions.

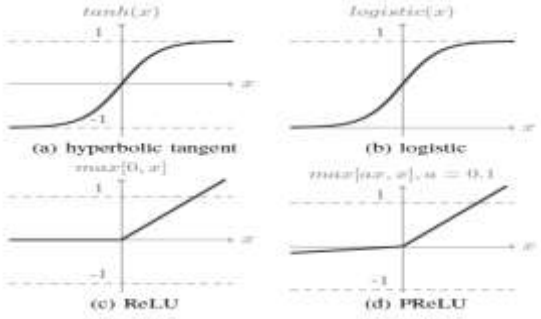

Fig. 3 Activation functions [3]

Artificial neural network (ANN)

Artificial neural network consists of neuron or perceptrons and activation functions. The hidden layers or units can be formed by connecting the perceptrons. The hidden units form the nonlinear basis. It maps the input layers to output layers in a lower-dimensional space and that is what can be termed as artificial neural networks. ANN is basically a map from input to output. The map can be obtained by weighted addition of the inputs with biases. A model can be formed by the values of weight and bias values along with the architecture [12].

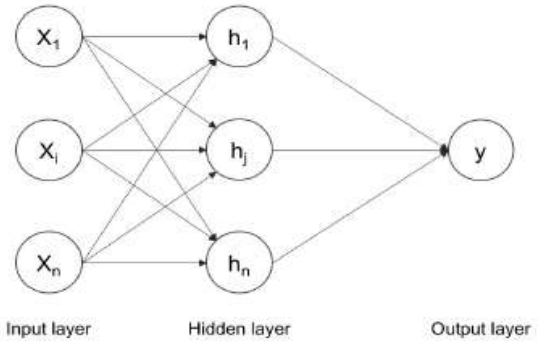

Fig. 4 A Multilayer perceptron or ANN [12]

B. Methods of deep learning

This section focuses on deep learning methods.

I. Convolutional Neural Networks (CNN)

In this CNN model, multiple layers are trained in an end-to-end manner. It outperformed effectively in computer vision applications. It comprised of three layers namely, convolutional layers, pooling layers, and fully connected layers. Each layers functions separately [6]. Fig. 4 shows a general CNN architecture for image classification.

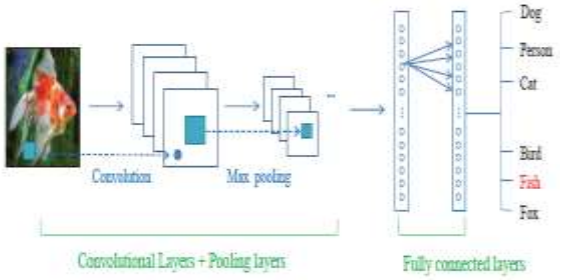

Fig. 5 A general CNN architecture [6]

Training of the network can be divided in two stages: a forward stage and a backward stage. The forward stage represents the input image with the current parameters in each layer. Here the currents parameters are weights and bias of each layer. The

loss cost with the ground truth labels can be computed by the prediction output. The backward stage computes the gradients of each parameter with chain rules based on the loss cost. For the next forward computation, all the parameters are prepared by updating the parameters based on the gradients. The network learning can be stopped after sufficient iterations of the forward and backward stages [6].

II. RESTRICTED BOLTZMANN MACHINES (RBMS)

The Restricted Boltzmann Machine (RBM) proposed by Minton et al. [11]. It is a generative random neural network technique. Basically the RBM architecture is a double barreled graph, the hidden layer units H & the visible layer units V1 display conditional independence. Hence,

P(HV1)=P(H1V1)P(H2V1)...P(Hn V1) (1)

Where, V1 as input, 'H' can be derived using P(HV1). By altering the parameters, we can reduce the difference between V1 & V2 by altering the parameters. And thus resulting 'H' will output a fine lineament of V1. Deep Belief Networks (DBNs), Deep Boltzmann Machines (DBMs) and Deep Energy Models (DEMs) are the RBM based methods [10].

III AUTOENCODER

The auto encoder is a special type of artificial neural network. It is used for learning efficient encodings. It is trained to reconstruct its own inputs X. And hence, the output vectors have the same dimensionality as the input vector. Fig 6 shows the general process of an auto encoder.

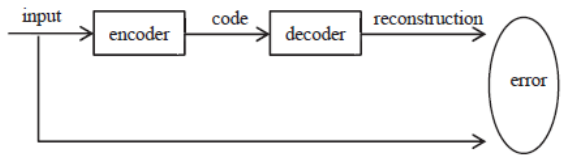

Fig. 6 Auto encoder [6]

The auto encoder can be optimized by reducing the reconstruction error and the corresponding code is the learned feature. We cannot get the discriminative and representative features of raw data using a single layer. Basically the deep autoencoder forwards the code learnt from the previous autoencoder to the next, to accomplish their task. Autoencoder has different variants: sparse autoencoder, contractive autoencoder, and denoising autoencoder [6].

IV. COMPUTER VISION APPLICATIONS

This section focuses on different computer vision applications.

A. Face recognition

For extracting high-level visual features, ConvNets has claimed with good results. The learned feature of CNN can directly conduct face verification [1].

The DeepID project uses a 4-layers convolutional neural network. This CNN is without the input, output and max pooling layers. The DeepFace uses a 5-layers convolutional neural network structure. This CNN is not including the input layer and the output layer, where the last three layers do not use weight sharing. The DeepID2 consisted of the four convolutional layers. There is local weight sharing in the third and fourth convolutional layers, and the output layer. The third and fourth layers are fully connected. This technique outperformed against all deep learning algorithms. There is significant improvement in the recognition rate [1].

B. Object detection

Detection of object is the process of detecting instances of semantic objects of a certain class in digital images and Video. A class may be humans, birds, airplanes, or plants etc [4].

Object detection includes the creation of a large set of candidate windows. Such windows are in the sequel classified using CNN features. The paper [13] employed selective search to derive object proposals. For each it extracts CNN features for each proposal. For deciding whether the windows include the object or not, it feeds the features to an SVM classifier. As proposed in [13], a large number of works is based on the concept of Regions with CNN feature.

C. Activity recognition

The human action includes the labeling of people's actions in the video. In the security applications, the monitoring system marked the criminal actions of terrorists and thieves. The camera can be used to identify the elderly or children with dangerous behaviors. Activity recognition comprises of two stages namely: feature extraction and behavioral

understanding [1]. Fig. 7 indicates a block diagram of human activity recognition.

Fig. 7 Block diagram of human activity recognition [1]

V. CONCLUSION

In this information age, Artificial intelligence is attracting research community's attention. In this AI research filed, machine learning and deep learning are playing a crucial role. This paper introduced deep learning concept and methods of deep learning. It also focused some computer vision application including face recognition, object recognition and activity recognition.

The future scope of this work is to implement deep learning based computer vision application including activity recognition or object recognition.

VI. ACKNOWLEDGMENT

I would like to thank Dhole Patil College of Engineering, Pune for a great support. I also would like to thank to Head of Dept. Prof. Amit Zore for guiding me and sharing his knowledge and experience in connection with this work. I also would like to thank the authors who contributed directly and indirectly in the field of Deep Learning and computer vision and due to which I wrote this survey paper.

REFERENCES

[1] Qing Wu.et al., "The application of deep learning in computer vision", pp. 6522-6527, 2017.

[2] Francois Chollet, "Deep learning with Python", by Manning Publications, 2018.

[3] Moacir A. Ponti, Leonardo S. F. Ribeiro, Tiago S. Nazare, Tu Bui, John Collomosse, "Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask", 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2017.

[4] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E., "Deep Learning for Computer Vision: A Brief Review", pp. 1-13. https://doi.org/10.1155/2018/7068349, 2018.

[5] Ashwini Patil, Vandana Navale, "A Review of Machine Learning Algorithms", IJIRT, Vol. 4 Iss. 12, 2018.

[6] Yanming Guo, "Deep Learning for Visual Understanding", Ridderprint, 2017.

[7] K. Fukushima, "Neocognitron: A hierarchical neural network capable of visual pattern recognition," Neural networks, vol. 1, no. 2, pp. 119–130, 1988.

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

[9] 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., 2012, pp. 1106–1114.

[10] Rajat Kumar Sinha, Ruchi Pandey, Rohan Pattnaik, "Deep Learning For Computer Vision Tasks: A Review", International Conference on Intelligent Computing and Control (I2C2), 2017.

[11] Boureau Y L, Ponce J, LeCun Y. A theoretical analysis of feature pooling in visual recognition, in: ICML, 2010.

[12] Rajalingappaa Shanmugamani, "Deep Learning for Computer Vision", Packt publication, 2017.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '14), pp. 580–587, Columbus, Ohio, USA, June 2014.

[14] Ian Goodfellow, Yoshua Bengio, Aaron Courville, "Deep Learning", MIT Press, 2017.