**Advances and Applications in Mathematical Sciences**
Volume 21, Issue 12, October 2022, Pages 6919-6939


# DEEP LEARNING AND COMPUTER VISION: A REVIEW

**MAMTA, ANURADHA PILLAI and DEEPIKA PUNJ**

[1]PhD Scholar
[2]Associate Professor
[3]Assistant Professor
Department of Computer Engineering
JC Bose University of Science
and Technology, YMCA
Faridabad 121006, India
E-mail: mamtamahiya@gmail.com
anuangra@yahoo.com
deepikapunj@gmail.com

## Abstract

Deep learning is among the subfields of machine learning that deals with the algorithms which endeavor to achieve human intelligence. With the arrival of deep learning, various applications related to computer vision techniques have been ushered and are now experiencing prominent development these days. The field of Computer vision has been revolutionized by deep learning. Computer vision in itself is not unraveling, but deep learning work as a backbone to address extensive variety of traditional applications. Deep learning has surmounted various machine learning techniques in many areas e.g. image processing, forgery detection, and medical science, among many others. This paper outlines some of the most crucial deep learning techniques such as; Convolution Neural Network, Autoencoders, Deep Belief Network and Restricted Boltzmann machine. It then presents features of various CNN architectures, which have evolved in the last few years. Some of the applications of deep learning in various computer vision tasks including image classification, object detection, object tracking, semantic image segmentation and instance segmentation, among many others have been discussed. Finally, the conclusions that have been emerged from the comparison of all existing techniques are summarized at the end.

**Abbreviation.** CNN, convolution neural network, DBN, deep belief network, RBM, restricted boltzmann machine, ReLU, rectified linear unit.

2020 Mathematics Subject Classification: 68T01, 68T07, 68T45.
Keywords: Deep learning, Computer Vision, Convolution Neural Network, Machine Learning, Image classification.
Received September 20, 2021; Accepted November 8, 2021

## 1. Introduction

Amongst various subfields of machine learning deep learning is gaining importance because of its vast applications. It allows computational models to understand and mimic how the brain perceives and understand information. Computer vision is a prominent subfield in the era of artificial intelligence that aims to provide computers and machines with an understanding of the contents of digital data (i.e., images or video). With the advancement of deep convolution neural networks, deep learning has gained astounding performance in various fields. In particular, it has brought an uprising to the computer vision area with the introduction of progressive and systematic solutions to many problems that are not resolved from a long time. The main focus is to achieve an artificially intelligible system that can challenge human brain. Deep learning is originated from conventional neural network but it surmounts the capability of its antecedents. With deep learning solutions many promising applications of computer vision techniques are untangled including object detection [1] and classification, image segmentation, object tracking, image reconstruction, video processing, autonomous driving and robot localization as well as medical image related applications.

Various deep learning methods are achieving extraordinary results on several problems such as CNN (convolution neural network) provides nearly human accuracy in many computer vision problems including classification, detection, segmentation and speech recognition. CNN is a multilayered architecture; the first layer extracts the lower level features and the last layer extracts the higher level features. Inspired by the human brain it can automatically process information and extracts the required features from a given input. Another model based on deep learning that is used for representational learning is an Autoencoder. It is an unsupervised model. It aims to learn encoding for a given dataset by training the network; it reconstructs its own input by reducing the redundant information from the given input data.one of the application of an autoencoder in the field of computer vision is image denoising. Further Autoencoder has been proven effective in variety of computer vision tasks like image cleaning, image compression and dimension reduction. Dimensionality reduction makes it easier to classify, visualize and storage of high dimension data [2].

Recently, DBN (Deep Belief Network) and RBN (Restricted Boltzmann Machine) have attracted considerable interest and are used to model high dimension data such as video and speech classification data. By combining a variety of neural networks DBN forms a new neural network model. DBM is a stacked RBM [2], which is further trained in a greedy approach. DBN uses RBM as a learning module, so in this review, the explanation of RBM is given with all the mathematical equations, and then DBN is explained with its architecture.

The rest of the paper is organized as follows. Firstly we have provided a brief overview of computer vision and its applications based on the background reviews. Then we have explained various deep learning methods in section 2, and these methods include CNN [3], Autoencoder [4], DBN and RBN [5]. The basic features of the existing models based on CNN are summarized in a diagrammatic way. This will help the researchers to select an appropriate architecture taking into account of their research work. Section 3 deals with some of the applications of deep learning like image classification, object detection [1], [6], object tracking [7], [8] and semantic image segmentation [9]. An overview of them is presented, and their further applications have also been mentioned. The last section of the paper is the conclusion that compares all the techniques.

**1.1 Computer Vision.**

In its simplest definition, Computer vision is a science that gives machines the ability to recognize. It enables the machines to understand the world through the processing of signals. In the last few decades, Computer vision has outperformed in assessing images and identifying the human movements. One of the famous applications of compute vision is image classification. Image classification has various applications in lots of technologies [10]. Its applications varies from large scale generalization such as understanding the context of an image to smaller and detailed result such as looking for medical MRI images to detect some diseases such as Autism[11]. There exists various deep learning based architectures which help in classifying an image and the most popular one is CNN based architecture. CNN's performance is majorly dependent on the fact that it can learn the most important middle level image features rather than the manual low level representations which are used in particular applications of image classification [12].

Object detection is another application in the field of computer vision. An object can be identified on the basis of its main properties such as color, size and texture which can be extracted from an image using CNN. On the basis of color information in an image an object can be identified from its background. Then the objects can be cut off from an image which further helps in various researches such as face mask detector [13] that gives a solution to prevent the spread of current pandemic Corona Virus disease.

Real time image Reconstruction and processing is a new concept in the area of deep learning. The process of image reconstruction involves reconstructing a well defined, high resolution image from a very noisy image. It has various applications in medical science such as X-Ray tomography, ultrasound and MRI (Medical resonance imaging). In [14] author provides a comprehensive survey on advancement in deep learning for image reconstruction and processing.

The world of computer vision has taken great stride in solving localization problems. Estimation of a mobile robot pose with respect to known location in an environment is known as Robot localization problem. When the problem is solved with use of sensor system installed in the robot itself then it is called as self localization. Using computer vision approaches the position of robot can be estimated from image captured by its visual system. In [15] the author explains how deep learning helps to achieve better performance while using existing state of the art algorithms in localization of robot.

## 2. Deep Learning Methods

Deep learning is an engrossing area that can solve various tasks of computer vision. It is a subset of machine learning that uses multiple layers to extract main features from the given input. Deep learning came into existence in 1943 when a computer based on human brain neural network was discovered by Warren McCulloch and Walter Pitz but due to the lack of hard and software availability those days, their model was not given much importance. Now deep learning has become a hot topic for research since 2006[16]. There are various applications of deep learning including image classification, text and speech recognition, video processing and many more. Considering each application various algorithms and models have been

developed over past few years. Some of the most important models that have contributed in improving performance are discussed here:

### 2.1 Convolution neural network

Deep learning, making use of machine learning algorithm, simplifies the process of feature extraction and object detection through a Convolution Neural Network (CNN). CNN divides the task of classifying an image into three layers: Convolution layer, Pooling Layer and Fully Connected layer. The detailed description of all the layers is as follows.

**Convolution Layer.**

This layer extracts the low-level features from an input image. It takes an image as input and sees it as an array of pixels. The image is represented in H*W*D where $H$ is height, $W$ is width and $D$ represents the color resolution (RGB). All input images will pass through a sequence of convolution layers and filters of size $(f_H * f_W * D)$ and gives an output volume of dimension $(H - f_W + D) * (W - f_W + 1) * 1$ [17].

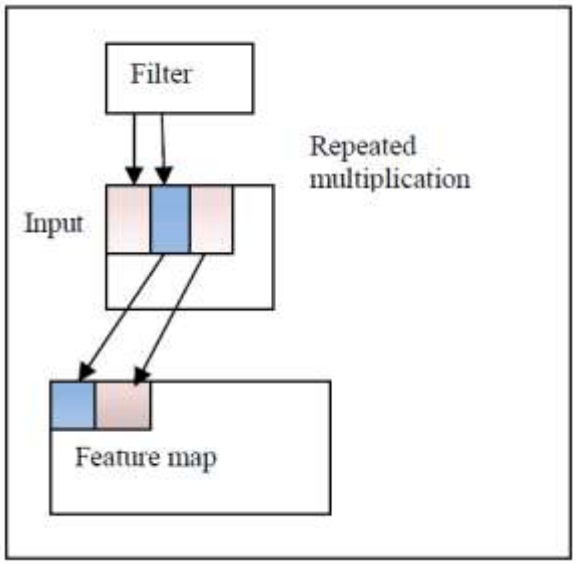


**Figure 1.** Filter multiplied to input matrix to get a Feature map.

It preserves the relationship among the pixels of the image by taking a small square of input data from the image matrix. Consider a 5*5 matrix

from the input matrix and multiply it with a filter matrix of size 3*3. A convolved feature matrix of size 3*3 is obtained. The feature matrix formed is called a "feature map". This operation helps in extracting the features of an image such as edge detection, the sharpness of image etc. The pictorial representation of how the filter is multiplied with the input matrix is shown in figure 1.

After getting a feature map, each of its value will pass through a nonlinearity function, ReLU. ReLU stands for a Rectified linear unit.

The ReLu function is defined in the equation below:

$$F(z) = \max(0, z) \tag{1}$$

ReLU is an activation function which helps in learning the neuron and decide whether to fire or not on the basis of input.

**Pooling layer**

This layer helps in the reduction of the number of parameters in the image and thus retains only the relevant information. There are three types of pooling: max pooling, average pooling and sum pooling.

Maximum value among all elements in a rectified feature map is taken in case of maximum pooling, the average value in average pooling and sum of all elements in sum pooling.

**Fully-connected layer**

After passing through the above two layers, the model can understand the features of the input image. Now, the final output will be flattened, and this output will be fed to the neural network for classification.

This layer computes the score of each class on the basis of input taken from the previous layer and gives an output in the form of a 1-D array which depicts all the classes that exist in the input image. The class to which the image belongs will have the highest score. The score is calculated using the softmax classification function.

In the first layer of CNN, each stride of input image matrix multiplied with the filter matrix creates a feature map, and on that feature map, ReLU activation function is applied, and then the image is converted into a suitable form for Multilevel Perceptron as shown in figure 2.

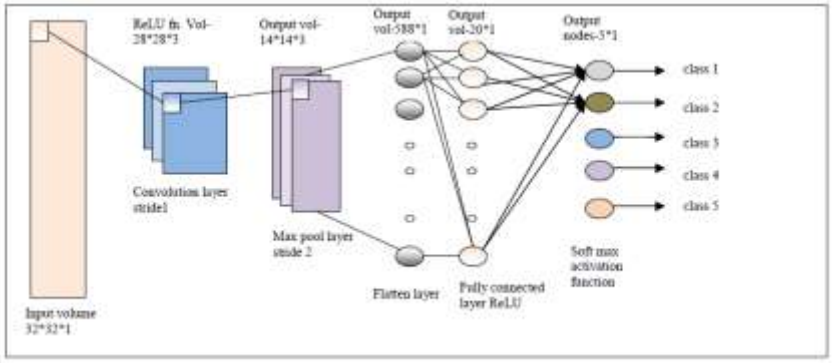


**Figure 2.** Classification in Fully Connected Layer.

Now the image is flattened into a column vector, and the output obtained will be given as input to the feed-forward neural network for training. For backward propagation of errors and learning the input, back-propagation [2] is applied in each iteration of training. Over a period of time, the CNN model understands the lower-level features of the image and can classify the image into various classes with the help of Softmax Classification Function [17].

### 2.1.1 CNN Models

CNN's are the most popular models of neural network and are able to solve variant of problems such as image recognition, image forgery detection, classification and many others. There are various models that have emerged in last two decades, each have different parameters and hyper parameters, including number of layers, learning rate, weight, filter size, stride and activation functions. Several standard architectures have been reviewed [18] and tested for their performance in different tasks.


**Figure 3.** Brief Overview of CNN Models.

During the last decade, various CNN models have been introduced [19]. Starting with LeNet which came in 1998 but due to unavailability of hard and software resources it became obsolete at that time. Then in 2012 again work started on CNN models which were effective to solve many computer vision tasks. From then, every year a new architecture was introduced. Each architecture perform some modification such as parameter optimization, structural changes and is able to upgrade the performance of CNN which is further utilized for various applications. Figure 3 summarized some of the key features of these architectures which will help researchers in choosing the suitable architecture for their target tasks.

### 2.2 Autoencoder

An autoencoder is a neural network derived from multilayer perceptron that transforms its input to its output with the least possible amount of distortion. Autoencoder learns to reconstruct the input and then extracts those attributes that help to predict the input accurately.

An autoencoder is a special case of feed-forward neural network and is trained using stochastic gradient descent algorithm that follows the gradients computed by the back-propagation algorithm.

Suppose there is a set of training attributes $\{y(1), y(2), \ldots,\}$ and $y\{i\} \in R_n$. An autoencoder is an unsupervised machine learning algorithm which applies back-propagation, setting the target value equal to the input value. i.e. $r(i) = y(i)$ [4]. An Autoencoder is also called an autoassociator or diabolo network [20].

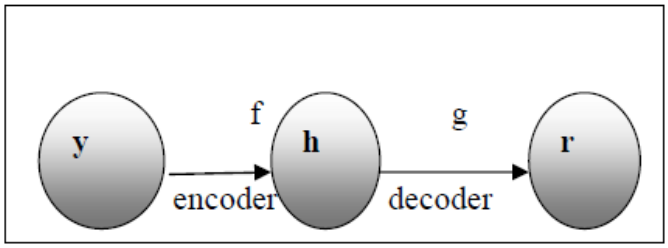


**Figure 4.** Autoencoder Structure Representing Mapping from Input to Output.

As shown in figure 4, an autoencoder has an input layer, a hidden layer and an output layer. An input layer receives the input; hidden layer describes the code used to represent the output. It maps an input $y$ to output $r$ (reconstruction) using the internal representation (code) $h$. The encoder function (f) maps $y$ to $h$ and decoder function (g) maps $h$ to $r$.

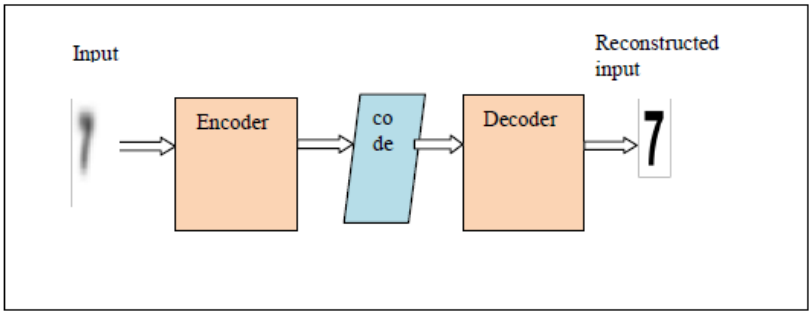


**Figure 5.** Autoencoder for image denoising.

Structure: suppose the input $y \in [0, 1]^d$ then it maps to hidden representation $h \in [0, 1]^d$ using an encoder function through a deterministic mapping:

$$h = s(Wy + b) \tag{2}$$

where $s$ can be any of the nonlinear function such as sigmoid function $s(1) = 1/+ e - t$, $b$ is the bias value, and $w$ is the weight assigned to input then the representation $h$ is mapped back to reconstruction $r$ of the same input as $y$. the mapping will be like:

$$r = s(W'h + b) \tag{3}$$

So $r$ predicts the input $y$ given the code $h$. [21].

### 2.2.1 Applications

The primary application of an autoencoder is data denoising, i.e. to reduce the noise from an image (as shown in figure 5) and dimension reduction that includes image compression. It can be used as a generative model for image generation, and can also be used to extract the features since an encoder helps in learning the essential hidden features from the input data. Some other applications of Autoencoder such as sequence to sequence prediction where the input sequence can be a series of random numbers, the output sequence can be the reverse of a subset of the numbers given in the input sequence. It also helps in a recommendation system that recommends relevant items to the user based on the understanding of the user's preferences.

## 2.3 Deep Belief Network and Restricted Boltzmann Machine

Deep Belief Network (DBN) and Restricted Boltzmann Machine (RBM) are deep learning models, and both belong to "Boltzmann Family". DBM makes use of RBM as a learning module. A brief introduction of RBM and DBN is found in the next subsection, and the graphical depiction of DBN can be found in figure 6.

### 2.3.1 Restricted Boltzmann Machine

Restricted boltzmann machine is a generative machine learning model

that may contain variety of input data such as labeled/unlabelled images, speech, videos and sentences, among many other [2]. They can be used to model high dimension data like videos or speech, and most importantly, they are used in deep belief networks [7].

Consider a binary vector taken from a binary image as a training set. To model the training set with the help of RBM we use two types of units to correspond the pixels; visible unit and hidden unit. Visible units need to be observed and hidden units help in feature detection. To calculate the energy we use the Hopfield energy function as mentioned in equation 4:

$$En(v, h) = -\sum_{p \in vis} x_p v_q - \sum_{q \in hi} y_p h_q - \sum_{p,q} v_p h_q w_{pq} \tag{4}$$

$w_{pq}$ is weight and $x_p$, $y_q$ are the biases between visible and hidden unit. Then, to calculate probability between them via this energy function equation 5 is used:

$$\text{Prob}(v, h) = (1/G)e^{-E(v, h)} \tag{5}$$

Where $G$ is a partition function, obtained by adding all pairs of visible and hidden vectors:

$$G = \sum_{v,h} e^{-E(v, h)} \tag{6}$$

The network assigns probability to a visible vector, $v$, is obtained by adding all of the hidden vectors:

$$\text{Prob}(v, h) = \frac{1}{G} \sum_{v,h} e^{-E(v, h)} \tag{7}$$

The log probability derivative for training vector in correspondence to the weight is mentioned by:

$$\partial \log p(v)/\partial w_{pq} = [v_p h_q]data - [v_p h_q]\text{model} \tag{8}$$

To perform stochastic steepest ascent in the log probability of the training data, a learning rule is given in equation 9:

$$\Delta w_{pq} = \in ([v_p h_q]data - [v_p h_q]\text{model}) \tag{9}$$

Here $\in$ denotes learning rate.

Given visible unit $v$, the $h_q$ of hidden unit, $q$ is set to 1 with probability:

$$\text{Prob}(h_q = 1 \mid v) = \sigma(d_q + \sum_i v_p w_{pq}) \tag{10}$$

Here $\sigma(x)$ denotes logistic sigmoidal function that is represented by $1/(1 + \exp(-x))$. Provided with a hidden vector $v_p$ of visible unit p is 1 denoted by probability function as in equation 11:

$$\text{Prob}(v_p = 1 \mid h) = \sigma(c_p + \sum_i h_q w_{pq}) \tag{11}$$

We perform Gibbs sampling in an alternate manner and form an unbiased sample of $[v_p h_q]$ model. An iteration of Gibbs sampling will update all the hidden units using equation 7 and of the visible unit using equation 8 in parallel.

As per Hinton [2], after calculating the binary states of the hidden unit using equation 10, set each $v_i$ to 1 using probability calculated using equation 11 and produce a reconstruction".

Then the change in weight is evaluated in equation 12.

$$\Delta w_{pq} = \in ([v_p h_q]data - [v_p h_q]recon) \tag{12}$$

Here recon is reconstruction. This learning helps in achieving many significant applications and one of them is DBN.

### 2.3.2 Deep belief network

DBN is a generative machine learning model that has many hidden layers of Restricted Boltzmann machine, and its last layer works as a classifier. The hidden layers are connected to each other but not the hidden units. The hidden units express the similarity present in the data units and represent features based on similarity. The connection among all the lower level layers is directed whereas between the top two layers is undirected. The bottom layer having the visible unit accepts the input data.

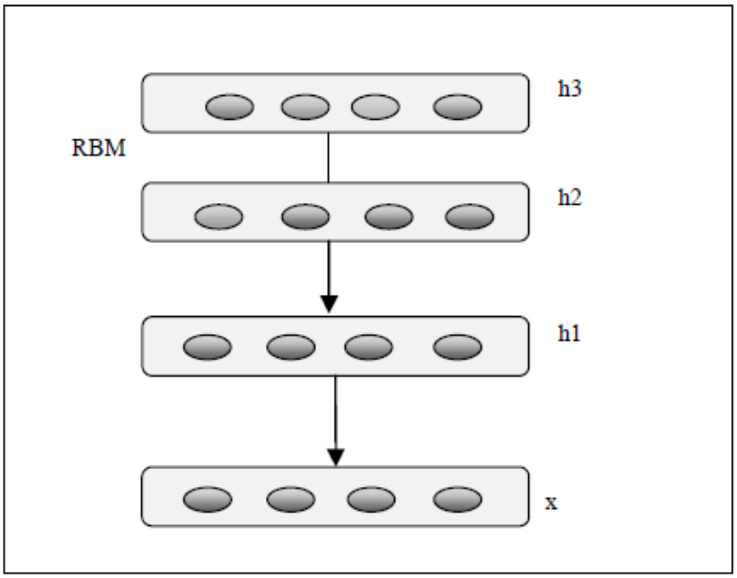


**Figure 6.** Architecture of Deep Belief Network.

Hinton [2] conveys that many RBM's are stacked and trained using the greedy approach to form a Deep Belief Network. It provides the joint probability distribution between observed vector $x$ and the hidden layers $h_j$ as follows:

$$\text{Prob}(x, h^1, h^2, \ldots, h^k) = (\prod_{j=0}^{k-2} \text{Prob}(h^j \mid h^{j+1}))\text{Prob}(h^{k-1}, h^1) \tag{13}$$

Where $x = h^0$, $\text{Prob}(h^{j-1}, h^j)$ denotes the conditional probability of the visible units conditioned on the hidden units of the RBM at level $j$, and $\text{Prob}(h^{k-1}, h^1)$ is the visible-hidden joint probability distribution in the upper-level RBM [22] as can be seen in figure 6.

To train the DBN with RBM, an unsupervised training is done in a greedy approach [2], [23]. The first layer, i.e. visible layer, generates the raw input $x = h^0$, and this input is used as data for the next layer, and then Gibbs sampling [32] is applied. It generates sample $\text{Prob}(h^1, h^0)$. The following layer is trained as an RBM and receives the sample as a training example and this process is repeated while propagating the collected data

upwards. So, the maximum values of weights are obtained in between the layers and these weights are used in the reversed direction using the fine-tuning.

## 3. Applications

Deep learning is a recent swing in the field of artificial neural network. There are many potential applications of it in our social life. These application include social network analysis, healthcare, Natural language processing and audio, video processing. Figure 7 shows a pictorial representation of various applications. Table 1 provides reference to some of the application discussed in research articles. Some of the most notable applications related to the field of image processing are presented in the following subsections:

### 3.1 Image classification

Image classification is a supervised learning problem of computer vision that helps in process of classifying an image on the basis of its visual content. It assigns labels to the different objects present in an image on the basis some specific rules. For example, given a medical image it can detect whether a particular disease is shown in the image or not. This task is trivial for human beings but still it is challenging in computer vision applications.

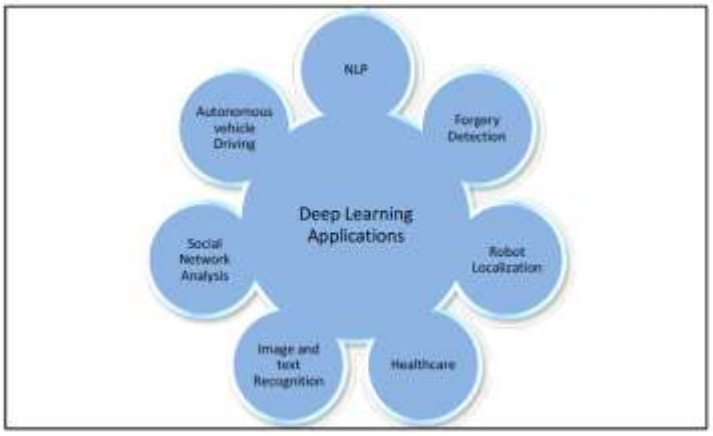


**Figure 7.** Example of applications of Deep Learning.

Classification is labeling group of pixels and classifying them based on

their features in an image. CNN performs a vital role in classifying the images and provides better efficiency than humans on some datasets. Motivation by [17], the basics of image classification using deep learning is explained in section 2 with the help of CNN technique.

### 3.2 Object detection

Object detection also termed as object recognition [6] aims at determining and locating the object from an image and bounding the object with a box by its coordinates height and width and also determining the type and class of located object in the image. Object detection was the foremost step in visual recognition activity. It can be classified further into two forms, single class object detection and multiclass object detection. In single class object detection, we take a single instance of the class from an image whereas in multiclass object detection, the classes belonging to all the objects in an image are taken. There are many challenges that need to be handled in this task such as face recognition, determining a distorted pattern in a picture etc. Deep CNN is a widely used algorithm for object identification after the deformable part model [5]. With deep learning based architecture support vector machine, we can split up the image into many classes and then the presence of an object can be found. For fast object identification, we can use Recurrent Neural Network. The concept of the region with CNN is proposed in [31]. Detailed explanation of object detection using CNN is given in [24].

### 3.3 Object tracking

Object tracking is a crucial task in computer vision that involves the process of tracking a moving object which could be a human being, a vehicle or a ball across a series of frames. Multi Object Tracking (MOT) [24] aims at analyzing the videos to locate and track objects related to various categories say vehicle, humans or animal without providing any previous information about the appearance and number of objects present in the input video.MOT algorithms allot a target id to each object, if an object moves away from the frame the id is dropped and if a new object appears a unique identifier is allotted to it. Still, there are many challenges to this process as the object looking similar cause the model to switch the id's or an object may disappear when it gets hidden behind other object or may appear in later frames [8]. Object Tracking has a vital part in resolving several computer vision tasks

including autonomous cars, vehicular pattern recognition, crowd behavior analysis and many more.

### 3.4 Semantic image segmentation

Semantic segmentation also referred to as pixel-level classification of the image. In earlier techniques, a rounded box on the object of the image was used to locate the object so the accurate idea of the object shape cannot be determined. In semantic image segmentation, every pixel of the image is allotted with a class, and that part of image is clustered, which belongs to the same class of object. The process of image segmentation is done in two parts: classification and detection. Classification treats each image as it belongs to the same category and detection localize and recognize the object. For a detailed explanation, refer to the literature [9]. Semantic image segmentation covers a broad range of applications, such as land use classification and land cover classification [26], facial segmentation, for autonomous driving, precision agriculture.

**Table 1.** Applications of deep learning discussed in various research article.

| **Technique** | **Application** | **Reference** |
|---|---|---|
| CNN+DBN+AE | Image Classification | [11][12] |
| CNN, AE | Object Detection | [6][24][33] |
| CNN, RBM | Object tracking | [24][8][34] |
| CNN | Image segmentation | [26][9] |

**Table 2.** Showing the comparison of various Deep Learning methods.

| Techniques | Details | Pros | Cons |
|---|---|---|---|
| Convolution Neural Network | CNN performs better with 2D data. Due to its filter in the convolution layer it transforms the 2D data into 3D [28]. | Uniform to transformation. CNN Provides more better and accurate result than other machine learning | Heavily depends upon labeled data. Requires more computing power. |

| | | | |
|---|---|---|---|
| | | techniques. | |
| Deep Belief Network | This network can learn in unsupervised manner. Hidden layer of each sub layer work as visible layer for the next layer [27]. | More robust in classification of input image in terms of size position and color etc. [29]. Model can be pre-trained in an unsupervised manner. | More computation complexity for training the network Using the greedy approach. Features are learned layer by layer. Does not re-adjusts its lower-level parameters. |
| Restricted Boltzmann Machine | Generative model that is used to model unknown distribution of data (image, text etc.) | Can be stacked up to create DBN and hence provide better computational power. Better than autoencoder in terms of ignoring random noise in training data [29]. | Estimating the partition function efficiently is hard. They are tricky to train well. |
| Autoencoder | It uses unsupervised learning and designed mainly for reducing the dimension of the input. | Easy to train in real time i.e. unsupervised learning | Pre-training is required and while training it may lose some data [28]. |

Table 2 compares all the Deep learning methods mentioned in the paper.

This table provides the basic detail of all the methods, their pros and cons. It can help in choosing one of the methods while applying them on one of the above mentioned applications.

## 4. Conclusion and Future Scope

The concept of deep learning has grown infinitely primarily due to its application in the area of computer vision. Various techniques that are discussed in this paper, namely, CNN, DBN/RBM and Autoencoder have achieved excellent performance in several computer vision tasks that includes object detection, Image Classification, object tracking and semantic image segmentation.

As shown in table 2 there are various categories based on which the methods mentioned in this paper can be compared, each technique is useful in some and lacks in other categories, such as CNN is uniform to transformation, but it heavily depends upon labelled data and requires more computing power. Feature learning is supervised in CNN, whereas DBN and Autoencoder can learn automatically based on given input data, i.e. in an unsupervised manner. Talking about the training efficiency, the Autoencoder can be trained in real-time easily whereas CNN and DBN need more computations for training purpose. But autoencoder needs the pre-training. While comparing RBM and autoencoder, RBM performs better in ignoring the random noise in the training data. However, all of the methods are good at generalization.

This paper introduces many deep learning methods, their application and concludes with the comparison of various methods. Many researches in deep learning have been reported until now, and there is considerable scope for further advancement. There are potential applications in social life like in healthcare sector, medical imaging system, forensic analysis, navigation, remote sensing and many more.

This paper gives an idea to new researchers to make use of existing methods of deep learning and explore more in this field.

## References


[1] Ajeet Ram Pathak, Manjusha Pandey and Siddharth Rautaray, Application of Deep Learning for Object Detection, Procedia Computer Science 132 (2018), 1706-1717.

[2] G. E. Hinton and R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, Science 313(5786) (2006), 504-507.

[3] Athanasios Voulodimos, Nikolas Doulamis, Anastasios Doulamis and Eftychios Protopapadakis, Deep Learning for Computer Vision: A Brief Review, Computational Intelligence and Neuroscience 5(4) (2018), 115-133.

[4] J. Zhai, S. Zhang, J. Chen and Q. He, Autoencoder and Its Various Variants, IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, 415-419.

[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, Object Detection with Discriminatively Trained Part-Based Models, in IEEE Transactions on Pattern Analysis and Machine Intelligence 32(9) (2010), 1627-1645.

[6] L. Liu, W. Ouyang, X. Wang, et al. Deep Learning for Generic Object Detection: A Survey, Int. J. Comput. Vis 128 (2020), 261-318.

[7] Geoffrey Hinton, Department of Computer Science, University of Toronto, A Practical Guide to Training Restricted Boltzmann Machines, Version 1, (August 2, 2010).

[8] Dina Chahyati, Mohamad Ivan Fanany and Aniati Murni Arymurthy, Tracking People by Detection Using CNN Features, Procedia Computer Science 124 (2017), 167-172.

[9] Li Y, Qi H, J. Dai, X. Ji and Y. Wei, Fully convolutional instance-aware semantic segmentation, In: Computer vision and pattern recognition (CVPR), IEEE (2017), 4438-4446.

[10] O. Alzakholi, H. Shukur, R. Zebari, S. Abas and M. Sadeeq, Comparison among cloud technologies and cloud performance, Journal of Applied Science and Technology Trends 1(2) (2020), 40-47.

[11] X. Yang, S. Sarraf and N. Zhang, Deep learning-based framework for Autism functional MRI image classification, J. Ark. Acad. Sci. 72(1) (2018), 47-52.

[12] Y. Li, H. Zhang, X. Xue, Y. Jiang and Q. Shen, Deep learning for remote sensing image classification: A survey, Wiley Interdiscip, Rev. Data Min. Knowl. Discov. 8(6) (2018), 1254-1264.

[13] S. Singh, U. Ahuja, M. Kumar, et al., Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment, Multimed Tools Appl. 80 (2021), 19753-19768.

[14] P. Shamsolmoali, M. E. Celebi and R. Wang, Advances in deep learning for real-time image and video reconstruction and processing, J. Real-Time Image Proc. 17 (2020), 1883-1884.

[15] Y. Jia, X. Yan and Y. Xu, A Survey of simultaneous localization and mapping for robot, IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (2019), 857-861.

[16] Q. Zhang, L. T. Yang, Z. Chen and P. Li, A survey on deep learning for big data, Inf. Fusion 42 (2018), 146-157.

[17] Ali Fadhil Yaseen, A Survey on the Layers of Convolutional Neural Networks, International Journal of Computer Science and Mobile Computing 7(12) (2018), 191-196.

[18] A. Dhillon and G. K. Verma, Convolutional neural network: a review of models, methodologies and applications to object detection, Prog. Artif. Intell. 9(2) (2020), 85-112.

[19] L. Alzubaidi, J. Zhang, A. J. Humaidi, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8(53) (2021).

[20] Acedemia Article, Andrea de Giorgio, https://www.academia.edu/39917960/A_study_on_the_similarities_ofDeep_Belief_Networks_and_Stacked_ Auto encoders (accessed 20 march 2021).

[21] Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning, MIT Press, (2016) https://www.deeplearningbook.org/contents/autoencoders.html.

[22] Yuming Hua, Junhai Guo and Hua Zhao, Deep Belief Networks and deep learning, Proceedings of International Conference on Intelligent Computing and Internet of Things (2015), 1-4.

[23] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, Greedy layer-wise training of deep networks, in Advances in Neural Information Processing Systems, MIT Press, (NIPS'06) 19 (2007), 153-160.

[24] K. L. Masita, A. N. Hasan and T. Shongwe, Deep Learning in Object Detection: a Review, International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), (2020), 1-11.

[25] Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano and Roberto Tagliaferri, Francisco Herrera, Deep learning in video multi-object tracking: A survey, Neurocomputing 381 (2020), 61-88.

[26] X. Yao, H. Yang, Y. Wu, P. Wu, B. Wang, X. Zhou, S. Wang, Land use classification of the deep convolutional neural network method reducing the loss of spatial features, Sensors, 19(12) (2019), 2792.

[27] Neelam Agarwalla et al., Deep Learning using Restricted Boltzmann Machines, (IJCSIT) International Journal of Computer Science and Information Technologies 7(3) (2016), 1552-1556.

[28] Razzak Muhammad Imran, Naz Saeeda and Zaib Ahmad, Deep learning for medical image processing: overview, challenges and the future, In Dey Nilanjan, Ashour Amira S. and Borra Surekha (ed), Classification in BioApps: Lecture Notes in Computational Vision and Biomechanics, Springer, Cham, Switzerland 26 (2018), 323-350.

[29] N. Lopes, B. Ribeiro and J. Gonçalves, Restricted Boltzmann Machines and Deep Belief Networks on multi-core processors, The 2012 International Joint Conference on Neural Networks (IJCNN), (2012), 1-7.

[30] M. D. Zeiler and R. Fergus, Visualizing and understanding convolutional networks, in European Conference on Computer Vision, (2014), 818-833.

[31] 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) (2014), 580-587.

[32] Keyvanrad Mohammad Ali and Mohammad Mehdi Homayounpour, A brief survey on deep belief networks and introducing a new object oriented MATLAB toolbox (DeeBNet), (2014). ArXiv abs/1408.3264

[33] J. Nourmohammadi-Khiarak, S. Mazaheri, R. Moosavi-Tayebi and H. Noorbakhsh-Devlagh, Object detection utilizing modified auto encoder and convolutional neural networks, Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA) (2018), 43-49.

[34] Rahul Roy, Susmita Ghosh and Ashish Ghosh, Salient object Detection based on Bayesian Surprise of Restricted Boltzmann Machine, ICVGIP 2018: Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing, (2018), 1-8.