IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-7, July 2019 E-ISSN: 2347-2693

Deep Learning Algorithms and Applications in Computer Vision

Savita K Shetty[1*], Ayesha Siddiqa[2]

[1,2]Information Science and Engineering, Ramaiah Institute of Technology, Bangalore, India

Corresponding Author: savita_ks1@msrit.edu, ayeshas95@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i7.195201 | Available online at: www.ijcseonline.org

Accepted: 16/Jul/2019, Published: 31/Jul/2019

Abstract—Deep Learning is a system powered by huge amounts of data. With the generation of massive amounts of data, the data analysing keeps getting complex. Deep learning solves the problem of Traditional **ML** algorithms that fail to perform well when the amount of data is enormous. Deep learning can be applied to any type of data such as text, image and so on. Deep learning algorithms generally used and best suited for image data are DBN and CNN. Analysing Computer vision using CNN brings a lot of use cases such as detection, recognition from the images, which can be useful in many fields such as medical images to detect a tumour and recognize its type, or help a robot navigate by identifying obstacles. In this paper we discuss what is Artificial Intellignece(AI), Machine Learning(ML) and Deep Learning and explore some of the Deep learning algorithms. We also understand how CNN can be applied in different applications of Computer vision and study the three major applications of Computer vision which are Image captioning, Medical image analysis and Robots Navigation.

Keywords **- AI, ML, Computer Vision, DBN, CNN, RNN**

I. INTRODUCTION

Deep Learning belongs to the family of Artificial Intelligence methods. It is inspired by the structure and ability of the cell neuron. It takes an input, analyses it and gives an output hence the name, Artificial Neural Networks. Deep Learning is based on ANN.

Artificial Intelligence is the development of intelligent systems, usually computers that are enabled to make independent decisions. These systems can make human like decisions without explicitly being informed. Any AI system is built upon the idea of learning, reasoning and self-correction. Where Learning is acquiring information(data), reasoning is using this information in making decisions and self-correction is confirming the correctness and remembering the choice and its credibility.

AI is growing popular because of the extensive amount of data generated each minute with the digital transformation. Most businesses and individuals are using technology to reduce their dependency on humans. To support a digital transformation there is also cheaper technology, cheaper storage space, and the convenience, which urges organizations and individuals to use it more. This data can be used in many ways to upgrade business and automate many mundane tasks.

Machine Learning is a section of AI that is associated with acquiring knowledge or skill by analysing, understanding and recognizing certain patterns from the data. Machine Learning is the study of algorithms that allow computer programs to improve through experience as defined by Tom Mitchell. [1]

In machine learning most of the features considered in analysis need to be chosen manually by an expert to make patterns more easily visible. Deep learning algorithms learn from high levels features incrementally.

Machine Learning algorithms are suitable for problems with moderate high amount of data. It takes up to few hours to train the algorithm. Deep Learning algorithms are more suitable for problems with enormous amounts of data so it takes much longer to train the algorithm. But at test time, Deep learning algorithms take less time to work.

These machine learning algorithms are further sorted into Supervised and Unsupervised. Supervised learning is when learning a function and training an algorithm that maps an input to an output based on example input-output pairs. Unsupervised learning is a (self-organized) learning that finds previously undiscovered patterns in data set without labels.

Further Deep Learning is a section of machine Learning as shown in fig 1. Deep Learning is inspired by structure and ability of a human neuron called Artificial Neural Network. ANN have a superiority over most other ML algorithms

because of its ability to employ **supervised, semi supervised and unsupervised learning** on diverse types of data.

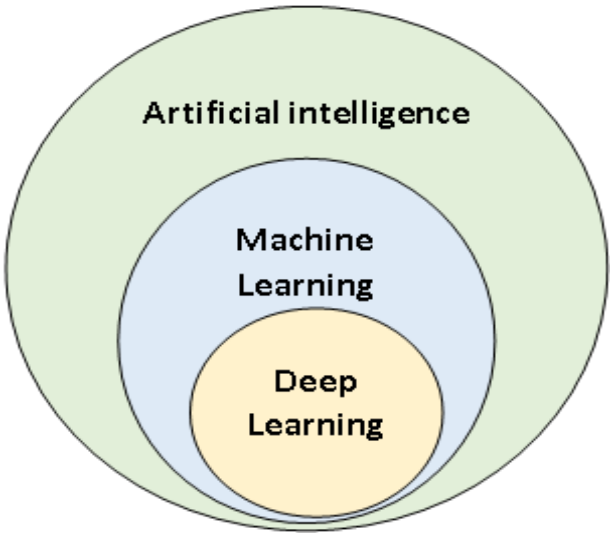

Fig 1 Deep Learning, Machine Learning and Artificial Intelligence.

Deep learning is applying deep neural networks with multiple layers and a lot more data than traditional ML algorithms and hence, it needs bigger models and more computation. It is also helpful as performance of traditional machine learning algorithms cannot be enhanced after a point even if the amount of data is increased but the performance of deep learning algorithms is directly proportional to amount and variety of data. As shown in fig 3. Artificial Neural Networks are systems that learn to take actions based on examples, without an explicitly specific program.

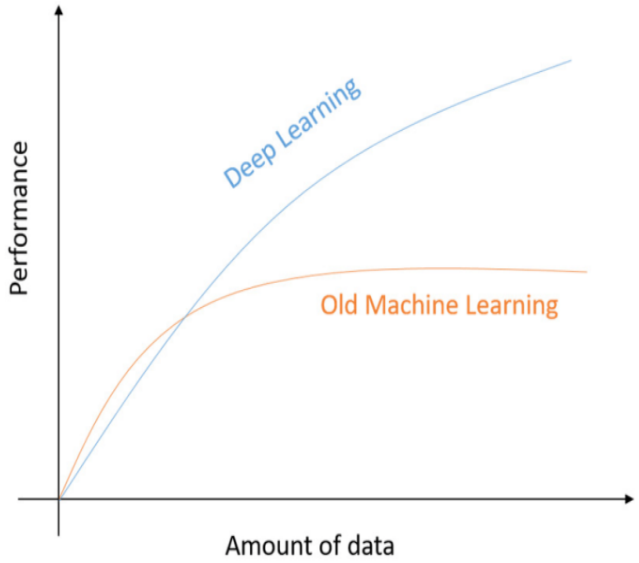

Fig 2: Performance comparison based on amount of data for AI and ML Algorithms

ANN architecture is made of three layers, namely, input, output and one hidden layer. Deep neural networks are ANNs with multiple layers between input and output layers i.e. multiple hidden layers.

The major Deep Learning algorithms are Deep Neural Network, Deep belief Network, Recurrent Neural Network and Convolutional Neural Network. These algorithms are applied for different applications based on the requirement and performance with different types of data.

MNIST, COCO, ImageNet, VisualQA are some of the open image datasets. IMDB reviews and Sentiment140 are some of the NLP datasets. Similarly, there are Voice datasets. These are labelled and pretrained datasets. So it can be used easily.

Organization of the rest of the paper is Section 2 contains the Literature review on CNN applied to Different computer vision applications, section 3 contains the basic Deep Learning Algorithms RBM, Autoencoder, DBN, CNN, RNN, and in section 4 we see how CNN works and its applications in Computer vision. Finally we conclude the research.

II. RELATED WORK

Computer vision is how computers can see and understand from images or videos. Computer vision is an attempt to replicate the human visual system to process to analyze and understand image data to make decisions. Intelligent systems capable of making decisions are built by using different Deep learning technique like DBN and CNNs. In this section we discuss the applications of CNN in computer vision i.e. Image captioning, Medical image analysis and Robot navigation.

2.1 Medical Image Analysis: this application of CNN can be used for many different requirements like research and diagnosis of diseases. There are several types of medical images that can be used like X Rays, MRIs and Ultrasounds. X Rays produce a 2 D image while MRI produce a 3D representation of the organ and ultrasound is a live video using conditions such as pneumonia, TB and cardiomegaly using X Ray images. By using a simple classification technique of CNN in the X-rays diagnosis of Pneumonia and TB can be done. The diagnosis of TB using Chest X rays in [2] considers two different approaches. First is feature extraction using *local and global filters* or feature descriptors and the second approach where feature extraction is done using *pre-trained CNN* networks. In both the approaches, the classification of the features is done using *SVM algorithm*. To use the pre-trained data many times there is a need to down sample the existing data, in this process, data might be lost out on useful information and that's why the first approach proposes the use of local and global filters. In the

second idea the pre-trained networks are tested for accuracy using ImageNet dataset. The generalization or trade-off between these two methods is suggested to be the best way to diagnose TB from a chest x ray by the authors.

In computer vision, segmentation of images is the process of partitioning it to better analyse or simplify the image. Semantic segmentation is when each pixel of the image is classified. Fully convolutional network is used in semantic segmentation, it is like a regular CNN but the fully connected layers that are usually present towards the end of the network are absent. [3] suggests the use of FCN for lesion segmentation in an MRI image of the Brain. FCNs are fast to segment and robust in learning the shape of the object from the 3D image.

However, it has two major drawbacks. One, the FCN fails to *identify the boundaries of the lesion*. To overcome this drawback, segmentation is done through *patch wise sampling* so that the prediction is based on local samples. But an FCN is likely to misclassify objects that are of extreme sizes, hence the second problem. FCN is *not sensitive to classify data with unbalanced class ratio*. [3] suggests methodology to overcome both cons by first, sampling of original images is done in 3D patches and train the patches so that the class ratio is not very different.

Diagnosis of other diseases such as prostate cancer is also possible through CNN. [5] defines a CAD: computer aided diagnosis system for detection of Prostate cancer. Firstly, images are obtained from a DWI volume, which is MRI result image dataset. The next step is prostate segmentation or delineation. The segmentation process is based on three features, *appearance, shape prior and spatial relationship*, for more efficiency. The third step is *feature extraction* differentiating features are identified from the images to classify them into malignant or benign. The final step is classification into benign or malignant. This is done using a CNN and the features extracted to differentiation in earlier steps.

2.2 Navigation for Robots or autonomous vehicles is implemented as a combination of GPS system and image analysis. This can be done using Image processing techniques such as edge detection for identifying the lane/ or the free path to move in or by using better Deep learning algorithms since they don't need explicit training with all types of data, they perform better with inputs not familiar with.

YOLO is an object detection technique used in real time. YOLO uses a single CNN network to detect and localize the object in the image. YOLO is used in real time as it is believed to be very fast. [8]

2.3 Image captioning: General idea is to use the CNN to extract features from the image and a Recurrent Neural Network (RNN) is used to generate the words.

A similar solution for image captioning is offered by [11] where CNN and RNN are used for feature extraction and sentence generation respectively to get a natural language description of the image. The use of VGG is made which is a form of CNN and is widely used for Image recognition. LSTM is used for caption generation with the idea of encoder decoder. The variable sequence of words is compared to an encoder while the sequence is formed using a decoder in order to obtain the natural and grammatically right sentence. CSMN context sequence memory network architecture which consists of recurrent model and context memory is an approach to image captioning Suggested by [6] Context memory types are specified like for different types of context information such as image memory, used context memory and word memory. Context memory is where author of the image query is identified earlier, and the words used by them most frequently are in memory. Word memory is words generated by analysing the image and previous use of words. Finally, the memories are concatenated to obtain the final sentence with words in the right order. Prediction of each word is done by analysing the previous use of words.

Further extending the use of the image captioning application of CNN [7] suggests visual question answering using similar technique.

A specialized RNN known as LSTM Long Short-Term Memory can also be used for generating the textual sentence.

III. DEEP LEARNING ALGORITHMS

Deep neural networks are not easy to train with back propagation due to the problem of vanishing gradient which impacts the time taken for training and reduces accuracy. Artificial Neural Networks calculate cost function based on the net difference between the Neural Network's predicted output and actual output in the training data. Based on the cost, weights and biases are altered after each process. Till the cost is as little as possible. Gradient is the rate at which cost will change based on weights and biases.

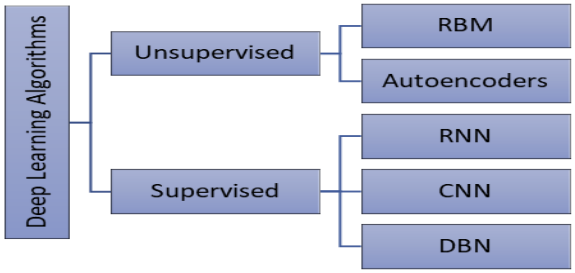

Fig 3: Classification of deep learning algorithms.

Table 1: Summary of Different applications of Deep learning in computer vision.

Problem	Method (Description)	References
Object detection	• YOLO is a technique based on CNN used for real time object detection which can be used in Robot navigation systems.	[8,9]
Object recognition	• Face recognition systems • Traffic signal recognition for self-driving cars.	[6] [10]
Medical image analysis	• xRays can be classified using CNN to detect TB and pneumonia. • Segmenting Lesion using an MRI Scan • Detecting Prostate cancer and predicting if its Benign or malignant.	[10] [11] [5]
Image captioning	• CNN and LSTM are used in feature extraction and sentence generation. • Extension of this application by visual question answering.	[11] [7]

This is the reason for late bloom of Deep nets. The problem of vanishing gradient can be avoided by using Deep Learning techniques. That is why Deep learning algorithms perform best with problems with huge data set.

3.1 Restricted Boltzmann Machine (RBM)

Restricted Boltzmann Machine is a shallow 2- layer Neural Network with each neuron in one group connected to each neuron in the other group without having any connections within a group. The two groups are visible (input) and hidden layers of neural network.

RBMs are trained to reconstruct the input data. Training procedure of RBM is forward pass, backward pass, compare result to input. In forward pass every input is combined with its individual weight and a single bias. In the backward pass, each neuron is combined with a weight and an overall bias and the result is passed to visible layer for reconstruction. The model in visible layer is compared with the original input. This process is repeated with various weights and biases till input and reconstruction are very similar.

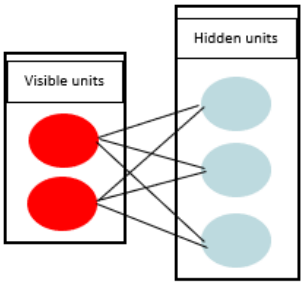

Fig 4: Architecture of RBM

3.2 Autoencoders

Autoencoder is a specialized Artificial Neural Network which learns a representation (encoding) for a set of data, by training the network to ignore signal noise. It also tries to re-generate the initial input from the reduced encoding a representation. The process of re generation of the input helps with dimensionality reduction as the system learns to ignore noise. An autoencoder may have any number of hidden layers.

Both RBMs and Autoencoders support unsupervised learning and are used in generative models because both techniques attempt to recreate the input.

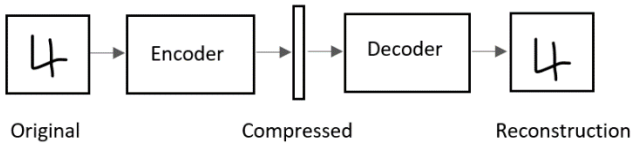

Fig 5: Autoencoder

3.3 Deep Belief Networks (DBN)

DBNs may be defined as a simple combination of unsupervised learning algorithms suck as RBMs and autoencoders.

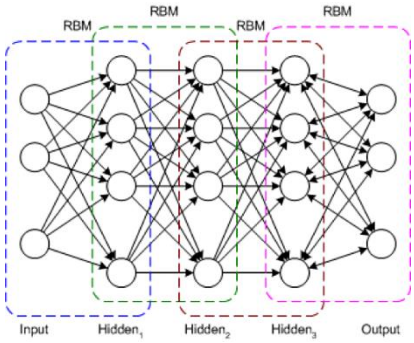

Fig 6: DBN Architecture

The structure of a DBM might look identical to an MLP but the training is like that of a stack of RBMs which is useful in reducing the vanishing gradient problem. Simple architectures of autoencoder or RBM are sequential placed for example, Hidden layer for the first RBM becomes the Visible layer for the next RBM.
General applications of DBNs are in Image/Video/ Face recognition.

3.4 Recurrent Neural Network

RNN is used when the output needs to be sequential like in image captioning and language translation. In a regular MLP each layer has its own weights and biases and hence cannot be combined. To combine these layers, use of same weights and biases is made (Recurrent layer). This ensures that the

neuron remembers the existing state and based on this state the next output is generated.

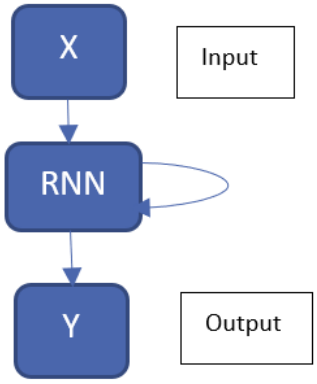

Fig 7: RNN architecture

3.5 Convolutional Neural Network

Convolutional neural networks consist of three major layers, Convolution, Pooling and fully connected layers.

Convolution function is defined as the integral of the product of the two functions after one is reversed and shifted.

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(\tau) \cdot g(x-\tau)\, d\tau$$

'*' Denotes Convolution and '.' denotes multiplication

Convolution layer is the first layer and it extracts the features from an input image. It uses an array known as Filter or Feature Extractor to identify features. In the convolutional Layer an activation function is applied which helps in getting non-linearity and to get an output based on the input. The activation functions may be ReLU, Sigmoid, tanh and so on. The most commonly used in CNN is ReLU.

Sigmoid function:

$$S(x) = \frac{1}{1+e^{-x}} = \frac{e^{x}}{e^{x}+1}.$$

Tanh function:

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}.$$

ReLU function:

$$RELU(x) = \begin{cases} 0 \; if \; x < 0 \\ x \; if \; x >= 0 \end{cases}$$

Fig 8: Activation functions sigmoid, tanh and ReLU

Pooling layer is the second layer and it helps in reducing the number of parameters since the images may be large. There are different types of pooling techniques based on requirement but most commonly used is Max Pooling which only considers the highest concentrated element of the obtained feature map.

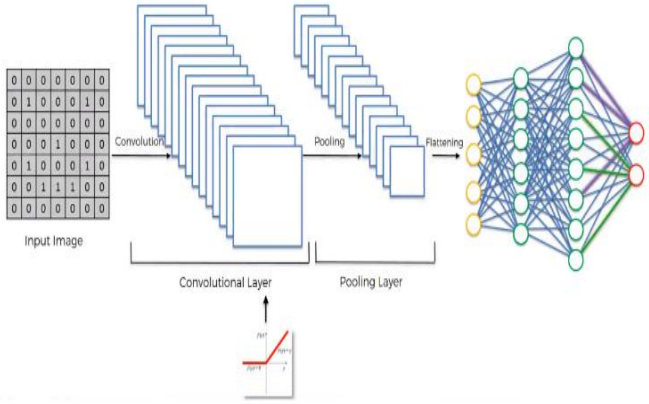

Fig 9: Architecture of a Convolutional Neural Network.

After flattening of the derived feature map, 1D array if fed into the Fully connected network. This functions just as any other neural network and after some back propagation most appropriate classification is presented.

Other than image analysis CNN is also used in Optical Character recognition to convert hand written document into digital text, which is a part of Natural Language Processing.

IV. APPLICATIONS OF DEEP LEARNING IN COMPUTER VISION

Convolutional Neural Networks are specialized fully connected neural networks with a shared weight architecture. It uses to small simple patterns to identify and analyse larger complex patterns thus reducing the risk of overfitting that may be caused by full connection.

A human eye tries to identify certain features in an image to analyse, classify and recognize the image. Similarly, CNNs use features through feature map vectors to identify the feature that can be used to classify images. Input to a CNN is image or video stream and the output is class the image belongs to.

Table 2 : COMPARISON OF THE DL ALGORITHMS

Parameter	RBM	Autoencoder	CNN	RNN	DBN
Type of learning	Unsupervised	Unsupervised	Supervised	Supervised	Supervised
Generative / Discriminative	Generative	Generative	Discriminative	Discriminative / Generative	Generative
Input data	Any type of data	Any type of data	3-D structured data, like Voice, Images	Mainly, Textual data,	Text, image
Output	Reconstructed input	Reconstructed input	Classified, predicted output	Sequence Prediction	Classified, predicted output
Application	Dimensionality Reduction/ Classification	Dimensionality Reduction	Image and voice analysis, classification, detection, recognition	NLP, Speech recognition	NLP, dimensionality reduction [6]

Some of the other major applications of CNN are in (a) Image recognition system like face recognition in smartphones use CNN and analysis of medical images to find tumors and classify it. (b) Video analysis since videos are like images with a temporal dimension it can be applied on videos. (c) NLP: CNNs are used for sentence retrieval, classification, prediction and other NLP tasks. (d) Drug Discovery AtomNet is a specialized CNN architecture used to discover chemical features like bonds between elements.

4.1 Medical Image analysis: is one of the most used applications of CNN. The main idea is to use medical images to extract information that may be missed because of human error or just automate the feature extraction to further help in more effective clinical diagnosis. Image analysis is useful for different purposes like segmentation, abnormality detection, disease classification, or computer aided diagnosis where images from medical imaging techniques such as X-ray or MRI are interpreted by systems.

4.2 Robot Navigation: Navigation of an autonomous Robot may happen through sensors, GPS or vision. Vision based systems to navigate through the way of a robot's motion can be implemented using Convolutional Neural Network. This implementation requires high computation and dataset with perfect labels. This application requires object detection and lane detection techniques.

4.3 Image Captioning: is a more specific application of CNN. Basic architecture of an image captioning architecture may be made of a combination of CNN and an RNN for generation of text sentences. To identify what is in an image and automatically generate a textual description of the image using artificial intelligence is image captioning.

The output generated must be grammatically right and convey more than just one-word classification or recognition answers. It can be the properties, actions and the description of what is in the image.

This problem has two aspects. One image aspect to identify what is in the image and another language aspect to describe the image in English. Different recommendation of using CNN in this problem are summarized in the Review section.

V. CONCLUSION AND FUTURE SCOPE

Deep Learning is a part of artificial intelligence that is based on artificial neural networks. Deep Learning algorithms more suitable for problems with huge datasets, other problems with smaller datasets may be solved simply by using Machine Learning. We compare the different models used in different problems such as object detection, object recognition, captioning and so on.

Some of the major deep learning algorithms are briefly studied such as RBM and Autoencoder that use unsupervised learning and CNN, DBN and RNN that use supervised learning. The algorithms are compared based on their inputs, outputs and basic working. We compare these algorithms based on parameters such as inputs data, output data and applications.

Using CNN can reduce a lot of computation because it doesn't need to visit the image pixel by pixel instead CNN uses filters. We discuss some of the CNN applications in computer vision such as in Medical image analysis of digital medical images such as EEG, ECG, X-ray and MRI scan reports to find any anomalies or unusual growths. CNN in robot navigation to help robots or other autonomous systems to move without any human intervention. CNN is also used in combination with RNN for image caption generation.

Based on this study, it can be concluded that CNN can accomplish the desired result in deep learning problems with image inputs. However, CNN's require high computational costs since they require a GPU and in absence of GPU they are very slow to train since they need lot of training data. In many cases, this drawback can be overcome by using pre trained models by fine tuning based on requirements.

REFERENCES

[1] T. M. Mitchell, "Machine Learning", McGraw Hill Education; First edition, New York, USA

[2] S. Rajaraman, S. Candemir, Z. Xue, P. O. Alderson, M. Kohli, J. Abuya, G. R. Thoma, and S. Antani, "A novel stacked generalization of models for improved TB detection in chest radiographs"
[3] B. Xu, Y. Chai, C. M. Galarza, C. Q. Vu, B. Tamrazi, B. Gaonkar, L. Macyszyn, T. D. Coates, N. Lepore, and J. C. Wood, "Orchestral Fully Convolutional Networks forsmall lesion segmentation in Brain MRI"
[4] S. Shabir, S. Y. Arafat, "An image conveys a message: A brief survey on image description generation"
[5]I. Reda, B. O. Ayinde, M. Elmogy, A. Shalaby, M. El-Melegy, M. A. El-Ghar, A. A. El-fetouh, M. Ghazal, A. El-Baz, "A new CNN based system for early diagnosis of Prostate Cancer"
[6] C. C. Park, B. Kim, and G. Kim, "Towards Personalized Image Captioning via Multimodal Memory Networks"
[7] Q. Wu , C. Shen , P. Wang, A. Dick, and A. v. d. Hengel, "Captioning and Visual Question Answering Based on Attributes and External Knowledge"
[8] M. Vicky, G. Aziz, H. Hindersah, "Implementation of Vehicle Detection Algorithm for Self-Driving Car on Toll Road Cipularang using Python Language"
[9] C. Lin, J. Lu, J. Zhou, "Multi-Grained deep feature Learning for Pedestrian detection"
[10] S. Hussain, M. Abualkibash, S. Tout, "A of Traffic Sign Recognition Systems Based on Convolutional Neural Networks"
[11] C. Amritkar, V. Jabade, "Image Caption Generation using Deep Learning Technique"
[12] S. Shabir, S. Y. Arafat, "An image conveys a message: A brief survey on image description generation"
[13] Boya Akhila, Burgubai Jyothi, "Face Identification through Learned Image High Feature Video Frame Works"
[14] N.S. Lele, "Image Classification Using Convolutional Neural Network"

Authors Profile

Savita K.Shetty received her B.E. (1996) in Computer Science and Engineering from Karnataka University ,Dharwad, and M.Tech (2004) in Computer Science and Engineering from Visvesvaraya Technological University Belagavi. She is currently working on her Ph.D. in Computer Science and Engineering, Visvesvaraya Technological University, Belagavi, Karnataka. Her research interests include Data Analytics, Data Mining and Machine learning.

Ayesha Siddiqa completed her B.E in computer science from Nittte Meenakshi Institute of technology affiliated to Visvesvaraya Technological University Belagavi, Karnataka in 2013. She is currently pursuing her MTech in Software Engineering at Ramaiah Institute of Technology, Bengaluru, Karnataka. Her areas of interest include, Machine learning, Deep learning and Data analytics.

