*2019 5th International Conference for Convergence in Technology (I2CT)*
*Pune, India. Mar 29-31, 2019*

# Techniques of Deep Learning for Image Recognition

Ganesh G. Patil
*Department of Computer Science and Engineering*
*SVERI's College of Engineering, Pandharpur, India*
ganeshpatil0222@gmail.com

Dr. R. K. Banyal
*Department of Computer Science and Engineering*
*Rajasthan Technical University, Kota, India*
rkbayal@gmail.com

***Abstract*— Image recognition is the reference of machine vision, is the ability of software to identify objects, peoples, places, actions and writing in images. Deep learning is the core field of machine learning and the most prominent algorithm of deep learning is convolutional neural network (CNN). Convolutional neural networks with more convolution hidden layers have more complex structure and more powerful feature extraction abilities than existing machine learning techniques. Nowadays, CNN has got tremendous success in image recognition. In this paper review the different methods of deep learning which has been utilized in the area for the recognition of image such as localisation and classification.**

***Keywords*— Convolutional Neural Networks, Deep Learning, Image Recognition.***

## I. INTRODUCTION

In current years, the deep learning got great success in image recognition. A machine learning has included deep learning method and is a strategy for learning a different deep neural networks included many layers such as hidden layers, input layer, and output layer[1]. As compared to the machine learning, deep learning provide the excellent performance facility on a large amount of data. There are three learning models used by deep learning: Unsupervised learning, Semi supervised learning, and Supervised learning. For running these models, it required multiple hidden layers and training a large data, for the improving the accuracy of recognition[1]. Deep learning has proved accuracy in many software applications including natural language processing, computer vision, speech processing, online advertising, robotics, video games, bioinformatics, search engines, audio processing, and etc[1]. Nowadays, the quick growth of innovation in the big data area creates deep learning to get a extreme research area. A vital field of deep learning, a CNN have generally used in image localisation, image classifications, image recognition, and have made extraordinary progress[1].

The technique of image processing is to transmute an image into digital form and apply some operations on that image, in instruction to become a quality image or to be extraction some useful feature from that image. The purpose of image processing has separated into five steps. These steps are as follows:

1. Visualization- Look the objects from an input image those are not noticeable.
2. Image sharpening and restoration- From an input image to make a quality image.
3. Image recover- To look the interest of image from an input image.
4. Measurement/Enumeration of pattern- From an input image measures the different objects.
5. Image Recognition- From an input image distinguish the different objects.

Image recognition is the reference of machine vision, is the ability of software to identify objects, peoples, places, actions and writing in images[2]. Nowadays, image recognition used in many areas e.g. Handwritten character recognition, Monument recognition, Human diseases detection, Crop diseases detection, Vehicle logo recognition, Vehicle number plate detection, and etc. To performing image processing, the regular steps are extraction of feature, which have produced essential features from an input images. To create a training model depends on the feature extraction from the input image dataset, which will be compared with each of the testing image. The above process is represented in fig. 1.

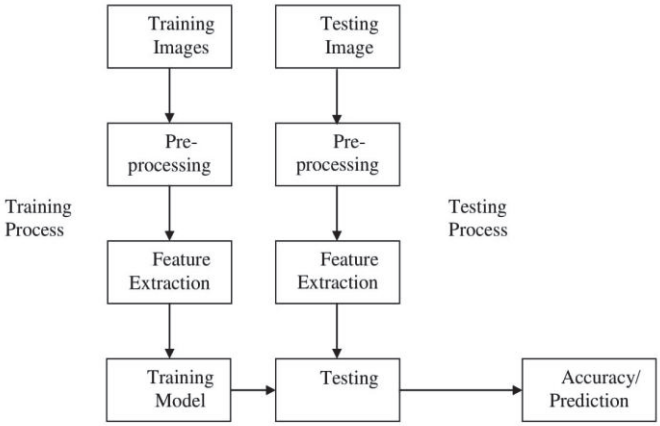


Fig. 1. General procedure of training and testing model[24]

## II. RELATED WORK

Nowadays, image recognition is the most intensively studied technologies in computer vision. To improve many applications by using the technologies of image recognition.

To be used deep learning on handwritten character (MNIST and handwritten character database) and javanese character recognition[3,4,5]. The good feature learning ability is available in the deep learning and it do not required to physically extraction of features. The deep learning learned maximum features automatically. Deep belief network (DBN) has a place with unsupervised technique for learning, it is a

**978-1-5386-8075-9/19/$31.00 ©2019 IEEE**

generate a deep model, while Convolutional neural networks has a place with supervised technique for learning, it is a separation a deep model[3]. It has conducted the experiment valuation of CNN and DBN on these database. The classification accuracy has obtained by CNN was 99.28% and DBN was 98.12% on the MNIST database, and CNN was 92.91% and DBN was 91.66 on the handwritten character database[3]. DBN is normally appropriate for one dimensional information demonstrating, for example talk, while CNN is more reasonable for 2-D information displaying, for example images[3]. In javanese character recognition, it used images from javanese characters. The classification has performed by CNN and DNN. The precision obtained by CNN was 70.22% and DNN was 64.65%[4]. The convolutional layer goes about as the extraction of feature process by using the fully connected layers. The other way, there is no feature extraction process is available in DNN[4].

Applied deep learning on medical images for detection/prediction of diseases[6,7,8]. Breast cancer is the universal cancer disease problem in women. The automatic image distribution of breast cancer histopathological have a challenging task in medical application. In current investigation technique is difficult to tending to the breast cancer grouping issue on the grounds that the element inconspicuous contrasts of between category histopathological picture and the order precision still difficult to met the medical application prerequisite[6]. The challenging this problem, authors proposed the breast cancer histopathological classification of image strategy depends on DCNN, called as BiCNN method, this method apply the breast cancer distribution on a pathological images using two classes, has got accuracy up to 97%, it provides capability tools for clinical diagnosis to detect the breast cancer[6]. Authors proposed an automatic Fetal Facial Standard Plane (FFSP) used for image recognition method depends on the DCNN model with powerful feature extraction capability[7].

Applied deep learning on plant/seed images[9,10]. In plant disease recognition, authors introduced a CNN method, but CNN required a maximum number of sample images for training. It is unfeasible or expensive to getting the training dataset required for this model[10]. The augmentation technique has generated number of training sample images because CNN required a maximum number of training dataset[10]. The augmented technique has generated training datasets utilizing activities for mirroring, shifting, contrast adjustment, extra gaussian noise, rotation, and adjustment of brightness[10]. Using deep learning, the seed sorting system of optical achieved purity and throughput by exploiting an inference strategy used by deep learning. Batch inference was a method to allow the use of a convolutional neural network classification, it has got maximum precision but it was not easily method to applied in the real time applications because it have huge amount of latency inference[9].

Applied deep learning on Vehicle license plate, Vehicle logo and traffic sign recognition[11,12,13,14]. License plate characteristics that vary from one country to another in terms of numbering system, language of characters, fonts, colors and size[11]. The system have basically three steps procedure: detection of license plate, character separation, and character recognition. In various plate, variations problem has occurred such as number scale, direction, lighting, and background[11]. Authors[11] was used pre-processing and CNN classifier steps for detection of license plate. In the procedure of recognition, character segmentations are very essential stage, it extract the numbers from the license plate image. The last step is character recognition, a number of procedures utilized for the character and number recognition, for example the statistical, the syntactic and neural networks[11]. Deep learning architecture represented by a CNN model has used for detection of license plate and recognition. A CNN got accuracy in detection of license plate was 93%, separation of character was 98% and recognition of character was 98%. The time for execution was 0.11 sec, it will be necessary for implementation of real time application[11,12]. For the driver, the recognition of traffic sign is an essential thing. In the recognition of traffic sign of Malaysia database, the performance of CNN has assessed and achieved the accuracy was 99%[13]. A manufacturer company of vehicle required the clearest indicator of vehicle logo, it has used a CNN for recognition of vehicle logo and obtained 99.07% accuracy[14].

Deep convolutional network got tremendous success in the vision of computer including the face recognition problem. This algorithm applied on AT&T database, poisson noise, and gaussian noise apply on sample images with pose variation. Authors[15] was used augmented technique for the creating a dataset for training model. For the gaussian noise dataset, the synthetically augmented techniques created a training sample dataset. In limited dataset of training, the recognition rate was achieved 95.21% with four samples of training and 99.92% with five samples of training on the AT&T database of face[15]. The dataset of training has increased, then automatically accuracy rate has been increased[15]. This technique will be applied for the another face analysis problem in future. A convolutional neural network has achieved great success in face recognition[23]. Authors introduced a new layer to embed the patch technique in convolutional neural network architecture. To partition a face image into patches using patch strategy technique, the process of cropped images and learned parameters of each patch can be conveyed out simultaneously in a convolutional neural network structure[23]. The patch that includes progressively and flawless face landmarks performs best because of more powerful data. Additionally, the patches including nose and eyes to get better accuracy than those including mouth and nose, which demonstrates that the data from eyes may be more helpful for face recognition than other landmarks[23].

A multi-object image recognition is challenging task in the image processing[21]. The classification of multi-object is to coordinate the objects of predicted labels given by the DCNN. It is indicated that the prediction of image happened in the classification of multi-object[21]. A DCNN was used

the BING algorithm to forecast target objects from input image and deep convolutional neural network for vectorization used for extraction of features and recognition of that objects[21]. The human body have patterns for verification in biometric recognition. Authors[22] proposed a fingerprint recognition in the deep learning method which inclose a pre-processing steps for extracting features of fingerprint, previously to applied the CNN. A feature extraction of fingerprint has carried out with equalization of histogram, fingerprint thinning, and gabor filter. The K-NN algorithm has replaced with a CNN because it enhanced fingerprint accuracy from 77% to 98.21%[22]. The deep learning model has effectively achieved better accuracy in image classification for fingerprint technique[22].

The recognition of monument has a challenging task in the image classification area due to the maximum variations in the monument structure[25]. The main instinct backward the classification of monument has the utilization of essential hand crafted highlights like LBP(Local Binary Patterns), GIST, and HOG(Histogram of Oriented Gradients) to see if the order of monuments used for extracting the features with a superior accuracy[25]. Authors[25] introduced a DCNN for extracting the feature of monument images. It has achieved 92.7% accuracy as compared to the features of hand crafted. It has proved that performance of this model increased with convolutional neural network[25].

## III. METHODOLOGY

Deep learning has basically three important algorithms:

1. Deep Neural Network
2. Convolutional Neural Network
3. Recurrent Neural Network

A deep neural network has utilized for forecasting and image classification. The convolutional neural network has used for feature extraction and classification of images. The Recurrent neural network has used for sequence of events, language models, time series, etc. In current years, a CNN is generally used in image localization, classification of image, image recognition, and have made extraordinary progress[1]. In Table 1 show that other deep learning models and its sample IoT applications[26].

TABLE 1. THE PERFORMANCE OF DEEP LEARNING MODELS ON SAMPLE IOT APPLICATIONS[26]

| Model | Learning mode | Data(Input) | Sample Applications(IoT) |
|---|---|---|---|
| AE | Unsupervised | Various | Emotion recognition, Machinery fault diagnosis |
| RNN | Supervised | Serial, time-series | Behaviour detection, Identify movement pattern |
| RBM | Unsupervised, Supervised | Various | Energy consumption prediction, Indoor localization |
| DBN | Unsupervised, Supervised | Various | Security threat identification, Fault detection classification |
| LSTM | Supervised | Time-series, long time dependent data, Serial | Mobility prediction, Human activity recognition |
| CNN | Supervised | Sound, Image, etc.( 2-D) | Traffic sign detection, Plant disease detection |
| VAE | Semi-supervised | Various | Failure detection, Intrusion detection |
| GAN | Semi-supervised | Various | Image to text, Localization and way finding |
| Ladder Net | Semi-supervised | Various | Authentication, Face recognition |

**Convolutional Neural Network:**

In deep learning, a convolutional neural network has a class of deep learning, artificial neural network for feed forward, most ordinarily applied to analysing visible imagery[16]. A CNN has a artificial neural network for multi layer especially intended to handle two dimensional input information. It utilize a variety of multilayer recognitions intended to require negligible pre-processing. Convolutional neural systems were inspired by biological procedures[16]. It utilize generally little pre-preparing contrasted with other image classification algorithm. A CNN has a best method for recognition of image and face but it required a large dataset for training. It used supervised learning model for feature extraction. A CNN includes multiple hidden layers, input layer, and output layer. The hidden layers typically includes convolution layer, pooling layer, normalization layer, and fully connected layer[16]. The above process has represented in fig. 2.

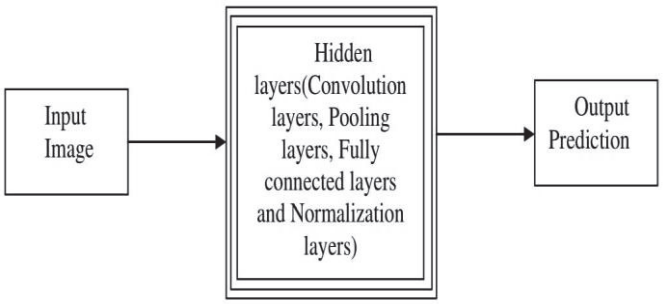


Fig. 2. Procedure of CNN

A convolutional layer applied to the input data using a convolution task and passing output to the next layer. A convolution imitates the reaction of an individual neuron to visual improvements[16]. Each convolutional neuron forms

information just for its receptive field. Convolutional networks may incorporate neighborhood or worldwide pooling layers, which consolidate the yields a neuron bunches at a one layer into a solitary next layer of neuron[16]. For instance, max pooling utilizes the most extreme incentive from every one of a bunch of neurons at the earlier layer. Another precedent is normal pooling, which utilizes the normal incentive from every one of a group of neurons at the earlier layer[16]. Each layer of neuron is connected to the every another layer of neuron by using fully connected layers. It is on a basic level the equivalent as the customary multi-layer perception neural system[16]. Another primary element of CNN is weight sharing. The system have a large training dataset and more hidden layers, they provide more accuracy, and it have less training dataset and more hidden layers, they provide less accuracy.

A CNN have four architecture models shown in Table 2. These models was applied on ILSVRC dataset (ImageNet Large Scale Visual Recognition Challenge). Table 2 summarized that, a large number of layers provide more accuracy (decreased error rate) in image recognition. AlexNet contended in the ILSVRC in 2012. The system accomplished a top 5 error of 16.4%, in excess of 10.8 rate focuses lower than that of the sprinter up[17]. AlexNet contained eight layers, the initial five are convolutional layers, the last three are completely fully layer. It utilized the non-immersing ReLU actuation work, which indicated enhanced preparing execution over tanh and sigmoid. AlexNet was initially composed with CUDA to keep running with GPU bolster[17]. VGG contended in the ILSVRC in 2014. VGG included total 16 layers (Convolution, output layers, and fully connected)[18]. GoogLeNet contended in the ILSVRC in 2014. GoogLeNet included total 22 layers[19]. ResNet contended in the ILSVRC in 2015[20]. On the ImageNet dataset assess ResNet with a profundity of up to 152 layers—8x more profound than VGG nets yet at the same time having lower multifaceted nature[20].

TABLE II.THE EXECUTION OF CNN MODELS ON ILSVRC DATASET

| Method | Error | Hidden layers | Proposed |
|---|---|---|---|
| AlexNet[17] | 16.4% | 8 | 2012 |
| VGG[18] | 7.3% | 16 | 2014 |
| GoogLeNet[19] | 6.7% | 22 | 2014 |
| ResNet[20] | 3.6% | 152 | 2015 |

An ensemble of these residual nets accomplishes 3.6% mistake on the ImageNet test set. This outcome won on the ILSVRC 2015 order undertaking [20].

Python is a open source language used for implementation of deep learning algorithms. Many authors used python for implementation of convolutional neural network models such as VGG[18], AlexNet[17], ResNet[20], and GoogLeNet[19]. It also support several commonly used deep learning models for developing, such as AE, RNN, and LSTM, on both GPU and CPU. Fig.3 show the development framework[5] which includes OS layer, Programming layer, and Toolkit layer. In first layer, perform the experiments on a Ubuntu operating system. In the programming layer[5], authors used Python language because Python is a standout amongst the most prevalent programming language in logical network and the other reason is that Python works exceptionally quick in runtime. In the toolkit layer, it construct the image recognition encompassing Tensorflow library which is a deep learning library[5].

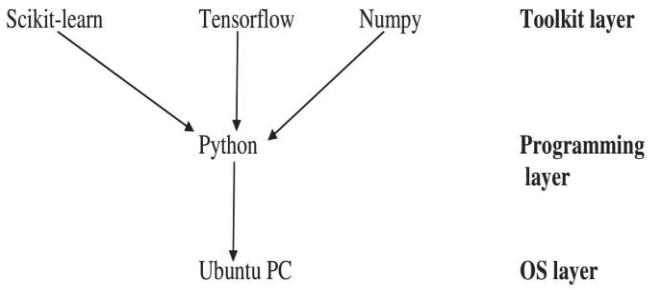


Fig. 3. Development Framework[5]

## IV. CHALLENGES AND OPPORTUNITIES IN IMAGE RECOGNITION

Although there are various challenges given by various authors. But few of them are important in respect of image recognition.

The first challenge in the image recognition are blur image, environment effected images (light, noise), occlusion, bad quality input image, distortion, facial expression, and etc. Various authors used various techniques to minimize these challenges as much as possible, but still these challenges are existing in image recognition. In future developed a new techniques for overcome these challenges.

The second challenge is the image recognition in less training dataset. If training dataset is less then it provide less accuracy in image recognition. The image augmentation technique is used for the extending the smaller dataset which lead to achieving higher accuracy on smaller dataset of the original images[9], because many algorithms are required a large dataset for training model. In future developed a new techniques will work on smaller training dataset and provide a higher accuracy.

## V. CONCLUSION

Deep learning is currently a very popular research area, it has got tremendous success because of that it uses broad information, strong computer system, and methods. A CNN is a standout amongst the essential algorithm of deep learning, its won the challenge of ILSVRC in current years[20]. It gained excellent results in recognition of image and face. In this paper, reviewed the successful and popular deep learning techniques which has been utilized in the area of recognition of image and pattern, e.g. handwritten character recognition, medical images recognition, plant/seed images recognition, monument recognition, fingerprint image recognition, license

plate of vehicle recognition, logo of vehicle recognition, recognition of traffic sign, face recognition, and etc. In these all recognition systems, a convolutional neural network achieved high accuracy. In future convolutional neural network will apply on complex images such as blur images, environment effected images (light, noise), gestures, facial expressions, and etc.

LIST OF ACRONYMS

| | |
|---|---|
| AE | Auto-encoder |
| CNN | Convolutional Neural Network |
| CPU | Central Processing Unit |
| DCNN | Deep Convolutional Neural Network |
| DNN | Deep Neural Network |
| DBN | Deep Belief Network |
| GPU | Graphics Processing Unit |
| GAN | Generative Adversarial Network |
| IoT | Internet of Things |
| LSTM | Long Short-Term Memory |
| RBM | Restricted Boltzmann Machine |
| ResNet | Residual Networks |
| RNN | Recurrent Neural Network |
| ReLU | Rectified Linear Units |
| VAE | Variational Auto-encoder |
| VGG | Visual Geometry Group |

## REFERENCES


[1] L. Hui-bin, W. Fei, C. Qiang, and P. Yong, "Recognition of individual object in focus people group based on deep learning," *2016 International Conference on Audio, Language and Image Processing (ICALIP)*, Shanghai, 2016, pp. 615- 619.

[2] https://searchenterpriseai.techtarget.com/definition/image-recognition.

[3] Meiyin Wu, and Li Chen, "Image recognition based on deep learning," *2015 Chinese Autom. Congr.*, pp. 542–546, 2015.

[4] Rismiyati, Khadijah, and A. Nurhadiyatna, "Deep learning for handwritten Javanese character recognition," *Proc. - 2017 1st Int. Conf. Informatics Comput. Sci. ICICoS 2017*, vol. 2018–Janua, pp. 59–63, 2018.

[5] N. H. Phong, B. Ribeiro, and P. De Marrocos, "Offline and Online Deep Learning for Image Recognition," pp. 171–175, 2017.

[6] X. Jin, T. Zhang, L. Li, H. Wu, and B. Sun, "Lesion recognition method of liver CT images based on random forest," *Proc. - 2016 8th Int. Conf. Inf. Technol. Med. Educ. ITME 2016*, pp. 227–230, 2017.

[7] Z. Yu *et al.*, "A deep convolutional neural network-based framework for automatic fetal facial standard plane recognition," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 3, pp. 874–885, 2018.

[8] Benzheng Wei, Zhongyi Han, Xueying He, and Yilong Yin, "Deep learning model based breast cancer histopathological image classification," *2017 IEEE 2nd Int. Conf. Cloud Comput. Big Data Anal.*, pp. 348–353, 2017.

[9] Y. J. Heo, S. J. Kim, D. Kim, K. Lee, and W. K. Chung, "Super-High-Purity Seed Sorter Using Low-Latency Image-Recognition Based on Deep Learning," *IEEE Robot. Autom. Lett.*, vol. PP, no. c, pp. 3035–3042, 2018.

[10] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosyst. Eng.*, vol. 172, pp. 84–91, 2018.

[11] Z. Selmi, M. Ben Halima, and A. M. Alimi, "Deep Learning System for Automatic License Plate Detection and Recognition," *2017 14th IAPR Int. Conf. Doc. Anal. Recognit.*, pp. 1132–1138, 2017.

[12] A. C. Nath, P. Dhar, K. Deb, and M. S. Hossain, "License Plate Recognition System Based On Contour Properties and Deep Learning Model," pp. 21–23, 2017.

[13] M. M. Lau, K. H. Lim, and A. A. Gopalai, "Malaysia traffic sign recognition with convolutional neural network," 2015 IEEE Int. Conf. Digit. Signal Process., pp. 1006–1010, 2015.

[14] Y. Huang, R. Wu, Y. Sun, W. Wang, and X. Ding, "Vehicle Logo Recognition System Based on Convolutional Neural Networks With a Pretraining Strategy," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 1951–1960, 2015.

[15] U. Aiman, and V. P. Vishwakarma, "Face Recognition Using Modified Deep Learning Neural Network," Iccnt, pp. 3–7, 2017.

[16] https://en.wikipedia.org/wiki/Convolutional_neural_network

[17] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, pp. 1097-1105, 2012.

[18] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[19] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-9, 2015.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[21] H. Shi, X. Mu, and S. Wang, "BVCNN: A Multi-object Image Recognition Method Based on the Convolutional Neural Networks," *2015 Int. Conf. Virtual Real. Vis.*, pp. 81–84, 2015.

[22] B. Pandya, G. Cosma, and A. A. Alani, "Fingerprint Classification using a Deep Convolutional Neural Network," *2018 4th Int. Conf. Inf. Manag.*, pp. 86–91, 2018.

[23] Y. Zhang, K. Shang, J. Wang, N. Li, and M. M. Y. Zhang, "Patch strategy for deep face recognition," *IET Image Process.*, vol. 12, no. 5, pp. 819–825, 2018.

[24] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., vol. 14, no. 3. Prentice Hall, 2008.

[25] A. Saini, T. Gupta, R. Kumar, A. K. Gupta, M. Panwar, and A. Mittal, "Image based Indian monument recognition using convoluted neural networks," *2017 Int. Conf. Big Data, IoT Data Sci. BID 2017*, vol. 2018–Janua, pp. 138–142, 2018.

[26] M. Mohammadi, A. Al-Fuqaha, S. Sorour and M. Guizani, "Deep Learning for IoT Big Data and Streaming Analytics: A Survey," in *IEEE Communications Surveys & Tutorials, 2018*.