Application Of Deep Learning using different CNN Structures for Real-time Image Classification in Industries.

Vyom Agarwal[†] 1
Junior Undergraduate, Indian Institute of Technology (BHU) Varanasi

Abstract - Deep Learning has emerged as a new area in machine learning and is applied to a number of image applications. The main purpose of the work presented in this paper, is to apply the concept of a Deep Learning algorithm namely, Convolutional neural networks (CNN) in image classification for application in Industries. This paper presents an approach for real-time training and testing for image classification for the real-time detection and recognition of targets in Industrial Setup to monitor employees. In this paper, we propose to use deep learning algorithms to achieve the expected results in industries for monitoring Attendance. Starting with use the Digit of MNIST data set as a bench mark for classification of grayscale images in general. Moreover, the proposed method can be used to detect and recognise targets in streetscape videos with high frame rates and high definition.

Keywords - **Deep Learning, Deep CNNs, Image Classification, Moving target detection, MNIST Image Dataset and Image Net.**

I. INTRODUCTION

INDUSTRIES should have a proper mechanism for verifying and maintaining or managing the real-time attendance record along with number of hours of working on regular basis. On that note, Deep Learning has been proved as a popular and powerful method to create Image Classification Models. In practice, the manual system also needs more time for recording and calculating the average attendance of each employee.[‡]

In deep learning, networks of artificial neurons analyse large dataset to automatically discover underlying patterns, without human intervention. With the increasing performance of convolutional neural networks (CNN) during the last years, it is more straightforward to classify images directly without extracting handcrafted features from segmented objects.

II. RELATED WORK

Recent progress in this area has been due to two factors: (i) End to End learning for the task using a convolutional neural network (CNN), and (ii) the availability of very large scale training datasets. Deep learning algorithms are a subset of the machine learning algorithms, which aim at discovering multiple levels of distributed representations. Keras has five Convolutional Neural Major networks that have been pre-trained on the ImageNet[‡] dataset are VGG16, VGG19, ResNet50, Inception V3 and Xception.

Various techniques, other than deep learning are available enhancing computer vision. Though, they work well for simpler problems, but as the data become huge and the task becomes complex, they are no substitute for deep CNNs.

TABLE I - CNN models and their configurations[‡]

Method (s)	Year	Configuration
AlexNet	2012	Five convolutional layers+three fully connected layers
Clarifai	2013	Five convolutional layers+three fully connected layers
SPP	2014	Five convolutional layers+three fully connected layers
VGG	2014	Thirteen/fifteen convolutional layers+three fully connected layers

[†] vyom.agarwal.met16@iitbhu.ac.in

*This Paper was drafted as a part of an Assignment in **LLG-309,** Conclusive Research Draft is currently in progress under aegis of **Dr. Sunil Kumar**, Mathematical Sciences, IIT(BHU) Varanasi.

GoogLeNet	2014	Twenty-one convolutional layers+one fully connected layer

Aziza Ahmedi, et al.[‡] proposed an Automatic Attendance System Using Image processing" the system used a camera that captures a video and sends to administrator server using web service. Features of face are extracted using Local Binary Pattern (LBP) and Histogram of Oriented Gradients (HOG), the features are eyes, nose, and mouth, and then it is subjected to the Support Vector Machine (SVM) classifier. Advantage of this system is it marks the attendance of the recognised faces automatically, Disadvantage is the recognition is carried out one by one and not in parallel so it requires a lot of time. [‡]

III. Basic Theory And Methodology

A. Neural Networks and Working Principle

Neural networks [‡] are made up of a number of layers with each layer connected to the other layers forming the network. A feed-forward neural network or FFNN can be thought of in terms of neural activation and the strength of the connections between each pair of neurons In FFNN, the neurons are connected in a directed way having clear start and stop place i.e., the input layer and the output layer. The layer between these two layers, are called as the hidden layers. Learning occurs through adjustment of weights and the aim is to try and minimise error between the output obtained from the output layer and the input that goes into the input layer. The weights are adjusted by process of back propagation (in which the partial derivative of the error with respect to last layer of weights is calculated). The process of weight adjustment is repeated in a recursive manner until weight layer connected to input layer is updated.

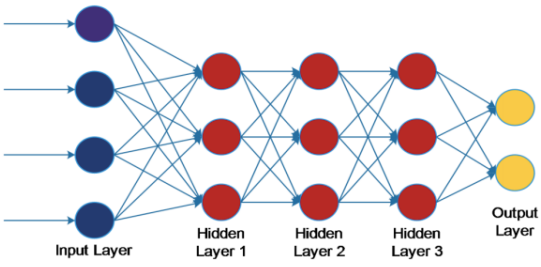

Fig. 1 Architecture of Artificial Neural Networks (ANNs) [‡]

Artificial neural networks (ANNs)[‡] are networks of simple processing elements (called 'neurons') operating on their local data and communicating with other elements. The design of ANNs was motivated by the structure of a real brain, but the processing elements and the architectures used in ANN have gone far from their biological inspiration. There exist many types of neural networks, e.g. see [‡], but the basic principles are very similar. Each neuron in the network is able to receive input signals, to process them and to send an output signal. Each neuron is connected at least with one neuron, and each connection is evaluated by a real number, called the weight coefficient, that reflects the degree of importance of the given connection in the neural network.

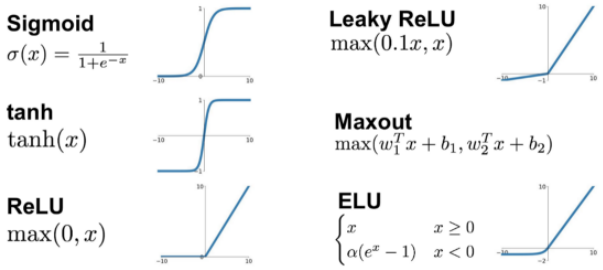

Fig. 2 Activation functions for ANNs [‡]

B. Deep Learning

After obtaining the feature vectors from the image, the image can be described as a vector of fixed length, and then a classifier is needed to classify the feature vectors.

In general, a common convolution neural network consists of input layer, convolution layer, activation layer, pool layer, full connection layer, and final output layer from input to output. The convolutional neural network layer establishes the relationship between different computational neural nodes and transfers input information layer by layer, and the continuous convolution-pool structure decodes, deduces, converges, and maps the feature signals of the original data to the hidden layer feature space [‡]. The next full connection layer classifies and outputs according to the extracted features.

C. Convolutional Neural Network

Our system uses grayscale images as input image having 28x28 sizes. The first layer in CCN applied 32 filters on input images, each image size is 3x3 producing 32 feature maps of size 26x26. The second layer is applying 64 filters, each of size 3x3 producing 64 feature

maps of size 24x24. Max pooling layer is act as third layer which is used to down sampling the images to 12x12 by using subsampling window of size 2x2. The layer 4 is fully connected layer having 128 neurons and uses sigmoid activation function for classification of images and produce the output image.

$$z(x,y) = f(x,y) * g(x,y) = \sum_{t=0}^{t=m} \sum_{h=0}^{h=n} f(t,h)\, g(x-t, y-h)$$

Convolutional Neural Networks[‡] basically consists of three major layers, these are Convolutional Layer, SubSampling or Pooling Layer, and Fully-Connected Layer. After merging all of these layers a CNN architecture has been set up. Let will try to see each layer shortly. Firstly, suppose there is a[32x32x3] inputted image in a raw pixel, this indicates an image with a width 32, height 32, and colour RGB (Red, Green, Blue) format. By deciding the filter size, the Convolutional layer will calculate a dot product between the weights with each and every neuron. For example, the result of this CONV layer might be [32x32x12] if we used 12 filters. RELU layer will use an activation function called Rectified Linear Unit (RELU).[‡]

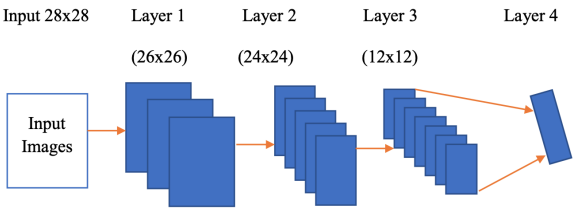

Fig. 3 Architecture of Convolutional Neural Networks[‡]

Each neuron in the Convolutional layer uses these activation function to train the CNN several times faster. The RELU function is defined as a = max (z,o) where, a and z indicates the input, whereas o indicates the output of the activation function. The Sub-Sampling or POOL layer is used to reduce the dimensions (width, height), then the volume is like [16x16x12]. Finally, the Fully-Connected (FC) layer calculate the final values of each neuron and compare them, and resulting in a volume of with the size of [1x1x10] among the 10 categories. In FC layer, each neuron will be connected to all the numbers in the preceding volume [‡].

Different layers of the convolutional neural network used are:

[ß] **Input Layer:** The first layer of each CNN used is 'input layer' which takes images, resize them for passing onto further layers for feature extraction.

[ß] **Convolution Layer:** The next few layers are 'Convolution layers' which act as filters for images, hence finding out features from images and also used for calculating the match feature points during testing.

[ß] **Pooling Layer:** The extracted feature sets are then passed to 'pooling layer'. This layer takes large images and shrink them down while preserving the most important information in them. It keeps the maximum value from each window, it preserves the best fits of each feature within the window.

[ß] **Rectified Linear Unit Layer**: The next 'Rectified Linear Unit' or ReLU layer swaps every negative number of the pooling layer with 0. This helps the CNN stay mathematically stable by keeping learned values from getting stuck near 0 or blowing up toward infinity.

[ß] **Fully Connected Layer:** The final layer is the fully connected layers which takes the high-level filtered images and translate them into categories with labels.

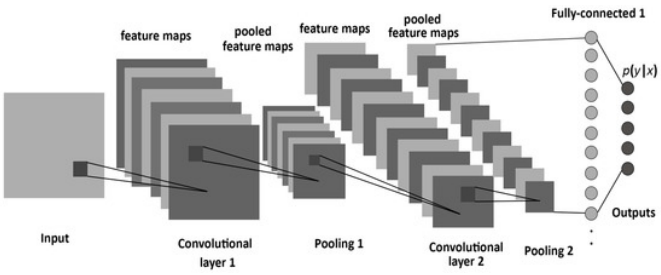

Fig. 4 A Typical Architecture Map with Hierarchical Units [‡]

D. Algorithm

1. Batch size =128 , no of classes 10, number of epochs = 5,
2. Dimension of input image 28 ×28,
3. Loading the input images from MNIST data set
4. Variable exploration: X=test data set (10000,28,28,1), Train data set (60000,28,28,1)
5. Creating and compiling the models
6. Training the network.

In General networks each neuron is connected to all neurons of previous layer. In real time it is impractical for high-dimensional inputs such as images. For instance, the input volume has size 32x32x3 and the receptive field size is 5x5. Then each neuron of conv

layer will have weights to a 5x5x3 region in the input volume for a total of 5*5*3=75 weights (and +1 bias parameter). The extent of connectivity along the depth axis must be 3, since it is the depth of the input volume. In CNN parameter sharing is used to reduces the number parameter in entire process.

Every neuron performs dot product by receiving some input and using bias it follows non-linearity. The whole convent still expresses a distinct score function, from the raw pixels on one end to class scores at the other end.

Convolutional Neural Networks (CNNs) have taken the computer vision community by storm, significantly improving the state of the art in many applications. One of the most important ingredients for the success of such methods is the availability of large quantities of training data. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [‡] was instrumental in providing this data for the general image classification task. More recently, researchers have made datasets available for segmentation, scene classification and image segmentation [‡].

E. Training Details

We train a full AlexNet on a large dataset to provide a useful feature extractor for the ELM and then train the ELM on the target dataset. Specifically, we train AlexNet on a dataset, which contains images from 16 classes. Therefore, the number of neurons in the last fully-connected layer of AlexNet is changed from 1, 000 to 16. All, but the last network layer are initialised with an AlexNet model1 that was pre-trained on ImageNet. The training is performed using stochastic gradient descent with a batch size of 25, an initial learning rate of 0.001, a momentum of 0.9 and a weight decay of 0.0005. To prevent overfitting, the sixth and seventh layers are configured to use a dropout ratio of 0.5. After 40 epochs, the training process is finished. The caffe framework [‡] is used to train this model dataset [‡] which contains images from 10 classes. The images are passed through the CNN stub and the activations of the fifth pooling layer were passed.

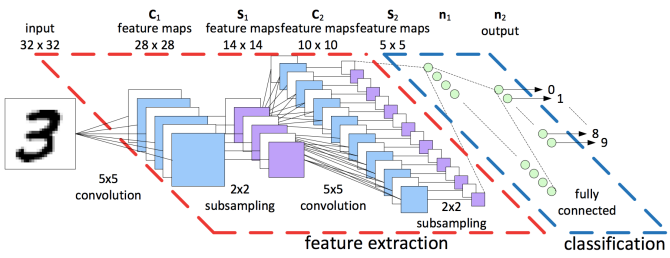

Fig. 5 A Typical Architecture along with the dimensions [‡]

IV. PROPOSED ARCHITECTURE

The proposed system captures images at regular intervals of time during the complete lecture, Faces are detected from the captured images after detection by using Face recognition technique (convolution neural network) the system will recognise the faces and mark the attendance of the recognised students. [‡]

V. CONCLUSIONS

We have addressed the problem of real-time training for document image classification. In particular, we present a classification approach that trains in real-time, i.e. a millisecond per image and them classify images for tracking of an employee in an industry. We used Convolutional Neural Networks (CNN) for image classification using images form hand written MNIST data sets. By training the images using CNN network we obtain the 68.7% accuracy result in the experimental part, These data sets used both and training and testing purpose using CNN. Images used in the training purpose are small and Grayscale images. The computational time for processing these images is very high as compare to other normal JPEG images.

Paper is based on the deep learning approach called Convolutional Neural Network (CNN) by summarising different literature studies. It is also discussed on the handcrafted feature extraction and the deep learning CNN model. It is also used to model and represent the data in a high-level abstraction. These all make the system more accurate, even in a challenging situation, i.e., illumination, occlusion, rotated faces, etc. We compared the deep learning based CNN approach with the previous state-of-the-art face recognition methods. The result specifies that the deep learning based CNN approach is more advanced than the other hand-crafted feature extraction techniques in various circumstances.

The system can be scaled out to be used in big Industries where the attendance and work hours are to be recorded and maintained where employees are awarded salary on overall performance of their work hours.

The proposed system improves the performance of existing **Attendance Management Systems** in the following ways:

i) Automatic tracking of the records of the employees and accurate marking of the attendance.
ii) Calculating the actual number of hours worked by an employee on a certain day.
iii)Increasing Efficiency and Security of the overall system.

*This Paper was drafted as a part of an Assignment in **LLG-309,** Conclusive Research Draft is currently in progress under aegis of **Dr. Sunil Kumar**, Mathematical Sciences, IIT(BHU) Varanasi.

VI. References and Reading List [‡]

- TensorFlow (2019). Image Recognition. [ONLINE] Available at: https://www.tensorflow.org/tutorials/image_recognition.
- J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. FeiFei, "ImageNet: A large-scale hierarchical image database," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255.
- Aziza Ahmedi , 2Dr Suvarna Nandyal, "An Automatic Attendance System Using Image processing", The International Journal Of Engineering And Science (IJES), vol. 4, issue
- W.S. McCulloch, W. Pitts, A logical calculus of ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115-133. [2] S. Haykin, Neural Networks - A Comprehensive Foundation, Macmillan, 1994.
- Krizhevsky, I. Sutskever, and G. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, NIPS 2012
- Rui Wang, Wei Li, Runnan Qin and JinZhong Wu "Blur Image Classification based on Deep Learning", IEEE, ISBN 978-1-5386-1621-5 pp. 1-6, 2017
- Teny Handhayani, Janson Hendryli, Lely Hiryantyo "Comparison of Shallow and Deep Learning Models for Classification of Lasem Batik Patterns", ICICoS, ISBN 978-1-5386-0904-0, pp. 11-16, 2017
- Laila Ma˝rifatul Azizah, Sitti Fadillah Umayah, Slamet Riyadi, Cahya Damarjati, Nafi Ananda Utama "Deep Learning Implementation using Convolutional Neural Network in Mangosteen Surface Defect Detection", ICCSCE, ISBN 978-1-5386-3898-9, pp. 242-246, 2017
- Sebastian Stabinger, Antonio Rodr´ıguez-S´anchez "Evaluation of Deep Learning on an Abstract Image Classification Dataset", IEEE International Conference on Computer Vision Workshops (ICCVW), ISBN 978-1-5386-1035-0, pp. 2767-2772, 2017
- Sachchidanand Singh, Nirmala Singh "Object Classification to Analyze Medical Imaging Data using Deep Learning", International Conference on Innovations in information Embedded and Communication Systems (ICIIECS), ISBN 978-1-5090-3295-2, pp. 1-4, 2017
- K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In Proc. BMVC., 2014.
- I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. 2014.
- Cole Murray, "Building a Facial Recognition Pipeline with Deep Learning in Tensorflow", Accessed on October 5, 2017
- CS231n, "Convolutional Neural Networks for Visual Recognition", Accessed on March 28, 2018 http://cs231n.github.io/convolutional-networks/

VII. Acknowledgement [ß]

The author would like to express his deep gratitude towards Dr. Sunil Kumar and Professor P.K. Panda for their support and encouragement during this work.