Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning

Manjunath Jogin
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
manjunathrjogin@gmail.com

Mohana
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
mohana.rvce@gmail.com

Madhulika M S
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
madhulikashivaprasad08@gmail.com

Divya G D
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
divyagdinesh@gmail.com

Meghana R K
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
meghana.rk.1997@gmail.com

Apoorva S
Department of Telecommunication
R. V. College of Engineering
Bengaluru, India
apoorva.shanky@gmail.com

Abstract— **The Image classification is one of the preliminary processes, which humans learn as infants. The fundamentals of image classification lie in identifying basic shapes and geometry of objects around us. It is a process which involves the following tasks of pre-processing the image (normalization), image segmentation, extraction of key features and identification of the class. The current image classification techniques are much faster in run time and more accurate than ever before, they can be used for extensive applications including, security features, face recognition for authentication and authorization, traffic identification, medical diagnosis and other fields. The idea of image classification can be solved by different approaches. But the machine learning algorithms are the best among them. These algorithms are based on the idea proposed years ago, but couldn't be implemented due to lack of computational power. With the idea of deep learning, the models are trained better and are able to identify different levels of image representation. The convolutional neural networks revolutionized this field by learning the basic shapes in the first layers and evolving to learn features of the image in the deeper layers, resulting in more accurate image classification. The idea of Convolutional neural network was inspired by the hierarchical representation of neurons by Hubel and Wiesel in 1962, their work was based on the study of stimuli of the visual cortex in cat. It was a fundamental breakthrough in the field of computer vision in understanding the working of visual cortex in humans and animals. In this paper feature of an images is extracted using convolution neural network using the concept of deep learning. Further classification algorithms are implemented for various applications.**

Keywords—Convolution Neural Network, kNN Classifier, Linear Classifier, Soft-max Classifier, Fully Connected Layer, Pool Layer, Activation Function Layer.

I. INTRODUCTION

The abstract idea of convolutional neural networks was first proposed by LeCun and was further improved in [1] [2]. LeCun proposed the idea for identifying the numbers on US postal cards using the networks of artificial neurons with local connections. The network couldn't be scaled further due to lack of computational power and small training dataset. In 2012, convolutional neural network took off to a whole new level and really proved to work on the general- purpose domain of image recognition with the invention of AlexNet. The reason for the success of AlexNet is, with deep neural networks we have an architecture that is flexible, with layers of neurons being added, which increases the learning capacity of those networks [3]. This resulted in the increased amount of training data. The next advantage was the use of machinery computational methods on GPUs to fetch the training data. Convolutional neural network (CNN) is a form of artificial neural networks that has specialization for being able to detect patterns and make sense of them; this pattern detection makes CNN useful for image analysis [5]. CNN is a sequence of layers and every layer performs some unique functions on the data fed to it. The five layers that form the CNN are Input layer-to hold the raw data, Convolutional layer-performs dot product between image patch and all the filters and computes the output volume, activation function layer- will apply activation function on every element of the output of the convolution layer, Pool layer- it makes the output of the previous layer memory efficient, so that the computation costs reduces , Fully-Connected layer- takes input from the previous layer and outputs the computed 1-D array class-scores. Object detection and tracking algorithms are described by extracting the features of image and video for security applications [10] [11]. Classifiers are used for image classification and counting [8] [9]. YOLO based algorithm with GMM model by using the concepts of deep learning will give good accuracy for feature extraction and classification [12].

II. RELATED THEORY AND FUNDAMENTALS

This section describes the theory and some of the fundamental concepts of classifiers along with its advantages and disadvantages. Classifiers are used for image classification and counting

A. kNN-Classifier

K-Nearest neighbor algorithm is a type of instance-based algorithm where the function is only approximated locally and all computation is delayed until classification. It is the fundamental and one the simplest classification techniques when there is little or no prior knowledge about the distribution of the data. Advantages of kNN is that, it is robust

978-1-5386-2440-1/18/$31.00 ©2018 IEEE

in noise training data, effective if the training data is large, there is no training phase and learns complex models easily. The disadvantages are that it is hard to determine the nearest number of neighbors and it is hard to apply in high dimensions which lead to low computational efficiency, large data sparsity and large amount of storage requirement. It is also not clear which type of distance metric to use and the computational cost is quite high.

B. Linear Classifier

Linear Classifier is a frontier that best segregates the two classes, hyper-plane or a line. It is best suited for extreme cases. They are one of the most used classification methods for practical applications which are linearly separable. Advantages include their effectiveness in high dimensional spaces. They are effective in spaces where number of dimensions is greater than the number of samples. They have different kernel functions for various decision functions. The kernel functions can be added together to achieve more complex hyperplanes. The disadvantages are, if the number of features is greater than the number of samples, it leads to the poor performance. They do not provide probability estimates.

C. Softmax classifier

Softmax classifier has a different loss function. It is a generalization to multiple classes of the binary Logistic Regression classifier. In SoftMax the outputs are treated like the scores of each class. The results are slightly more intuitive and will have a probabilistic interpretation. This classifier minimizes the cross-entropy between the estimated class probabilities and the "true" distribution. Mapping function and soft-max function is

$$f(Xi, W) = W * X_i \quad (1)$$

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (2)$$

Advantages of Soft-max classifier include efficient classification of basic problems that are linearly separable using soft-max algorithm and its simple model structure makes it easy to train and predict. Disadvantage of this classifier is that it only works on linearly separable data and it does not support null rejection.

D. Convolution Neural networks

In machine learning a feature map for data is created and the classifier is applied on this to solve the problem. And every problem has unique set of data and strategies that are applied are different to different problems. Hence, to overcome this CNN is used to automatically generate features and combine it with the classifier. Advantages of CNN classifier include, out of all the classifiers the list of layers that transform input volume to output volume are the simplest in this method. There are very few distinct layers and each layer transforms the input to output through a differentiable function. Disadvantage is that they do not encode the position and orientation of the object into their predictions. A convolution is a significantly slower operation than, say max pool, both forward and backward. If the network is deep, each training step is going to take much longer [3].

III. DESIGN, IMPLEMENTAION AND RESULTS

The convolutional neural network architecture is implemented with many layers such as convolutional, ReLu, max pooling. The architecture is inspired by AlexNet as shown in fig. 1. It consists of six layers of Conv2D, ReLu, Max-pooling and fully connected layer. Further layers like dropout added to the network to enhance the performance during training. The dropout layer is activated only during training. The dropout layer randomly drops certain number of neurons during forward pass (input to the function) and remembers the neurons that are left during the forward pass. And only updates the non-dropped during backward pass. The dropout is a feature that brings the regularization. The dropout layer makes the model to learn robust features that are independent to the neurons which in turn avoids the overfitting during the training phase.

Fig. 1. The layers of the convolutional neural network.

Fig. 1 shows the section of the architecture for convolutional neural network implemented. It involves the repetition of (Conv2D – ReLu – Maxpooling) layers six times with different filter size, gradually reducing the number of neurons. The number of neurons before the output layer is reduced to around 1000 from the input neurons of 3072 (32X32X3). It helps in reducing the number of weights at the fully connected layer at the end.

In this paper CIFAR-10 dataset is used for training the algorithm. The CIFAR-10 dataset consists of around 50000 images with each image having 32X32X3 dimensions. The dataset has 10 object classes of car, bird, ship, deer, truck etc. These images are labelled with numbers 0 to 9 for training the algorithm. These set further divided into test and training set for better performance. The accuracy is calculated over the test set, while the algorithm is trained over the training set.

A. kNN classifier

K-Nearest Neighbor is the non-parametric algorithm. During training, the classifier takes the training data and simply remembers it. During testing, kNN classifies every test image by comparing to all training images and transferring the labels of the k most similar training examples. Although kNN are not used in image classification anymore, but it is better (around 35% accuracy) than random guessing (around 10%). Since kNN is a non-parametric algorithm, it doesn't iterate for tuning parameters. The nearest neighbor each test image is calculated with each test image, a matrix of these distance values is obtained. For the value of k, the k number of nearest neighbors is selected and the least distance rule is selected to judge the nearest neighbor, in case of tie any rule can be applied to select the nearest neighbor. The dimensions of the input dataset are

Training data shape of the input image is: 50000X32X32X3
Training labels shape of the input image is: 50000
Test data shape of the input image is: 10000 X 32 X 32 X 3
Test labels shape of the input image is: 10000

```
In [5]: runfile('C:/Users/jogin/Downloads/winter1516_assignment1/assignment1/knnclasifier.py',
wdir='C:/Users/jogin/Downloads/winter1516_assignment1/assignment1')
Reloaded modules: cs231n, cs231n.classifiers.linear_classifier, cs231n.data_utils,
cs231n.gradient_check, cs231n.classifiers.k_nearest_neighbor, cs231n.classifiers,
cs231n.classifiers.softmax, cs231n.classifiers.linear_svm
Training data shape:  (50000L, 32L, 32L, 3L)
Training labels shape:  (50000L,)
Test data shape:  (10000L, 32L, 32L, 3L)
Test labels shape:  (10000L,)
(5000L, 3072L) (500L, 3072L)
10
Accuracy is  0.282
```

Fig. 2. Results of kNN algorithm on CIFAR10 dataset with k=10

Although kNN is not a good image classifier, it's better than a random guess. An accuracy of 28.2 percent is obtained by using kNN on CIFAR10 dataset. The fig. 2 shows the result of the kNN algorithm executed using Spyder IDE.

B. SVM classifier

The Multiclass Support Vector Machine (SVM) loss is different from the linear classifier loss. The support vector machine has a different loss function than the L2 loss. The score in SVM makes sure that the correct class gets a more score than the incorrect classes by a margin of delta. This way the SVM makes sure that the model learns to give high score to the correct class than the incorrect class, making sure that the model makes a better accuracy. For a given input image Xi and the correct label y_i, the model predicts scores for all the classes. The loss functions than calculates the loss by taking the correct class score S_j and all other incorrect class scores S_{yi} over each iteration. The Multiclass SVM loss for the i^{th} example is:

$$\boldsymbol{L_i = \sum_{j \neq y_i} max(0,\ s_{j-}\ s_{y_i} + \Delta)} \quad (3)$$

Where, L_i = Loss for the i^{th} iteration.
S_j = score of the j^{th} neuron
S_{yi} = score of the correct class
Δ (delta) = small value by which the correct class score should be greater than incorrect class. The input layer is the values of pixels of the image with 3072 neurons (32X32X3). And the output layer consists of 10 neurons (10 classes).

```
linear SVM on raw pixels final test set accuracy: 0.374000
```

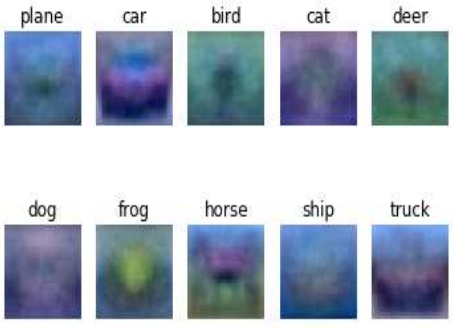

Fig. 3. Visualization of weights of the SVM algorithm.

The SVM performs better than kNN classifier, an accuracy of 37.8 percent is obtained and the visualization of the weights is shown in fig. 3 with the test set accuracy.

C. Softmax classifier

The Softmax classifier has a different loss function. It is generalization to multiple classes of the binary Logistic Regression classifier. The major difference between SVM and SoftMax classifier is the loss function. While the scores in the SVM are numerical values of any order, the Softmax scores are limited between (0, 1). Although there is no significant improvement in the accuracy for Softmax, it is preferred in the output layers of many models; it gives an institution of probability for a correct class. The loss is cross entropy as shown

$$\boldsymbol{L_i = -log\left(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}}\right)} \quad (4)$$

Where the function notation fj means the j-th element of all the class scores f. The total loss for the dataset is the average of Li over all training examples together with a regularization loss R (W).

```
Softmax on raw pixels final test set accuracy: 0.341000
```

Fig. 4. Visualization of weights for SoftMax classifier

The Softmax performance doesn't vary much with that of SVM classifier, an accuracy of 34.10 percent is obtained, and the fig. 4 shows weight visualization.

D. Fully-connected neural network

The neural networks in the SoftMax and SVM consist of only input and output layer. The idea of deep neural layer came into picture after the inclusion of hidden layers into the architecture. The multiple hidden layers in the network made the model learn better. Here a single hidden layer net is implemented. The number of neurons in the hidden layer is also a hyperparameter. It was found that the around 50 to 60 neurons in the hidden layer, the model performs better. The output layer in the fully connected layer consists of SoftMax loss. A two-layered neural network is implemented with ReLu as activation function in the hidden layers and Softmax as activation function in the output layer. The backpropagation is implemented to update the weights in the training period.

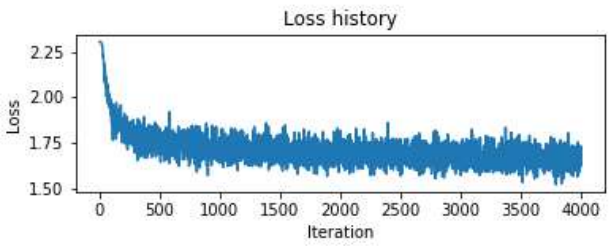

Fig.5. The loss history curve of fully connected neural network.

The fig.5 depicts the loss for the fully connected layer over 4000 iterations. The loss curve decreases rapidly at first and saturates gradually.

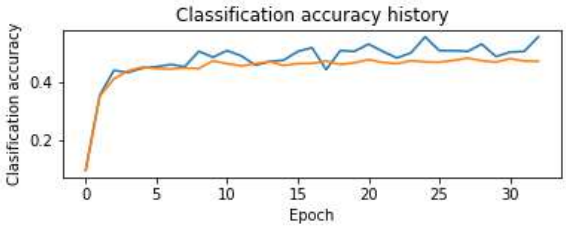

Fig.6.Classification accuracy curve of fully connected neural network.

The fig 6 shows the training and validation set accuracy for the fully connected model. Training set accuracy hasn't overshot the validation set accuracy; it means the model hasn't been over fitted.

```
test accuracy: 0.464
(3072L, 57L)
(57L, 32L, 32L, 3L)
```

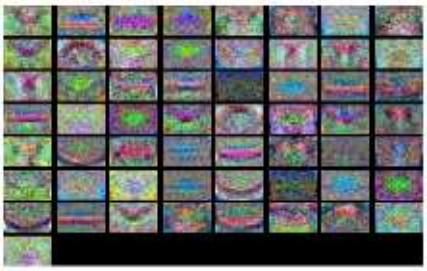

Fig. 7. The visualization of weights of the first layer.

The fully connected network tended to give maximum accuracy over the number of neurons in the hidden layer being around 60. The visualization of the weights pf these 60 neurons are shown in the fig 7. For instance, the model has learnt multiple view point of the car images, thus able to produce better test accuracy than SVM and SoftMax. The accuracy increases to over 50 percent with two-layered fully connected neural network as shown in fig 5. Classification accuracy history of training and validation as shown in fig 6. And the visualization of weights as shown in fig 7.

E. Convolutional Neural Network.

The architecture overview of CNN is discussed here; the architecture is inspired by AlexNet. The network for classification of images on CIFAR-10 dataset has the following stack. (INPUT-IMAGE - CONVOLUTIONAL - RELU – MAX-POOL – FULLY CONNECETD)The images of CIFAR-10 dataset contain 10 classes, namely cat, ship, horse and others. Each image is of the dimension 32X32X3 that holds the raw pixel values of the image, with three color channels R, G, B. Convolutional layer computes the local dot product with input and the weights. The resulting dimensions of the subsequent layer depend upon the number of filters used and the filter size. ReLu is the activation layer which doesn't change the dimensions of the layers. Max-pooling layer down samples values and reduces the dimension [4]. Fully connected layer will compute the class scores, hence the output layer consists of 10 neurons, the class scores are further mapped to the log probability using SoftMax activation layer [6] [7].

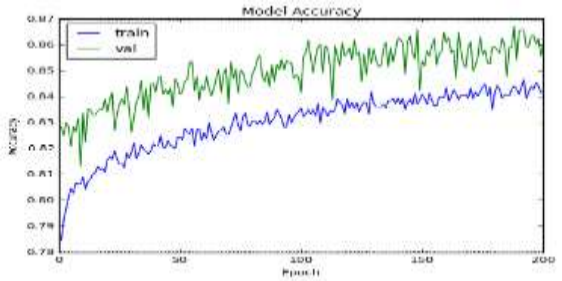

Fig.8. The accuracy result for convnet in training and validation phase.

Fig. 8 shows the plot of training and validation accuracy over the iterations for the convnet model. The accuracies tend to saturate after certain iterations, thus the increase in iterations doesn't increase the accuracy anymore.

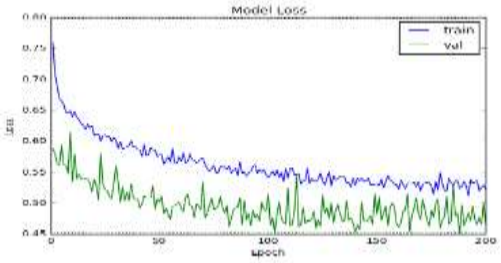

Fig. 9. The loss for convnet in validation and training phase.

Fig. 9 shows the loss curve in training and validation phase for the convnet model. A ConvNet is implemented and the accuracy obtained is 85 percent, the fig. 8 and 9 shows the accuracy and loss for the convnet trained on CIFAR-10 image dataset, further by changing the size of the filters and the number of filters the accuracy can be increased.

F. Result Analysis

Various algorithms are implemented for the image classification using CIFAR-10 dataset.

TABLE 1: CLASSIFICATION ALGORITHM AND OBTAINED ACCURACIES

Classification Algorithm	*Accuracy in %*
KNN Classifier	28.2
SVM Classifier	37.4
SoftMax Classifier	34.1
Fully connected Neural Network	46.4
Convolutional Neural Network	*85.97*

Table 1 depicts various algorithm and their respective test accuracies. Various algorithms are implemented to solve the problem of image classification and the results are compared. The kNN classifier is a rudimentary solution for image classification. The kNN classifier is non- parametric, hence it can't be trained, but still it performs better than random guessing. Next algorithms implemented were SVM and SoftMax, both these algorithms are similar except the scoring in SVM is categorical cross entropy and that in SoftMax is probabilistic. The performance of both these classifiers has no

much difference. The shortcoming in the SVM and SoftMax classifiers is there inability to identify the different angle of views of images of the object, animal of same class. This can be overcome by implementing fully connected layer with multiple hidden layers according to the necessity. The test accuracy improves drastically in FNN to around 50 percentage. But still this doesn't match the human accuracy of classification greater than 90 percentage. The convolutional neural networks outperform humans in image classification due to the addition of convolutional layer to the deep network. The convolution layer can alter the shape of the output (Unlike other models, where it is fixed). Thus, enabling to learn the basic object shapes in the primary layers, and learn to build on to identify much complex objects in the deeper layer. The convolutional neural network is adoption of the human visual cortex, hence able to perform in par with human. The convolutional neural network implementation drastically reduces the error rate. Hence the convolutional neural networks are chosen for feature extraction of images.

IV. CONCLUSION

Deep Learning has a wide range of applications in various domains of Home Automation, Industrial Automation, Facial Recognition, Surveillance and Crowd Management, Autonomous Navigation of Unmanned Aerial Vehicles (UAVs) in both indoor and outdoor environments, Medical Diagnostics, Self-organized Network Management in 5G technology. CNN is also used in Object Classification and Detection in Photographs. It is used in Recommender Systems, Natural Language Processing, Video Analysis, Game of Checkers and Drug discovery Deep learning has exhibited nice performance in many applications. The CNNs is an often-used architecture for deep learning and has been widely used in computer vision and audio recognition. The Convolutional neural networks from AlexNet to GoogLeNet have tremendously developed and have accuracy greater than human. With faster run-time classification finds many applications in autonomous navigation and many such applications. In this paper various classification algorithms are implemented using the concept of deep learning. Obtained results show that convolutional Neural Network will give 85.97 % accuracy for image classification. Further it can be used for various video survillance and security related applications for feature extraction and classification.

References

[1] Y Lecun, L Bottou, Y Bengio et al., "Gradient-based learning applied to document recognition[J]", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.

[2] A Krizhevsky, I Sutskever, G E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems* 2012, vol. 25, no. 2, 2012.

[3] M. D. Zeiler, R. Fergus, "Visualizing and understanding convolutional networks", ECCV, 2014.

[4] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition", ICLR, 2015.

[5] Guo lili, Ding shifei, "Deep learning research progress [J]", 2015.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with Convolutionals", Co RR, vol. abs/1409, pp. 4842, 2014.

[7] "Deeply supervised nets", *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics* AISTATS 2015, pp. 9-12, 2015.

[8] Mohana and H. V. R. Aradhya, "Elegant and efficient algorithms for real time object detection, counting and classification for video surveillance applications from single fixed camera," *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, Bangalore, 2016, pp. 1-7.

[9] H. V. Ravish Aradhya, Mohana and Kiran Anil Chikodi, "Real time objects detection and positioning in multiple regions using single fixed camera view for video surveillance applications," *2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO)*, Visakhapatnam, 2015, pp. 1-6.

[10] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, " Object Tracking Algorithms for video survilllance applications"*International conference on communication and signal processing (ICCSP),* India, 2018, pp. 0676-0680.

[11] Apoorva Raghunandan, Mohana, Pakala Raghav and H. V. Ravish Aradhya,"Object Detection Algorithms for video survilllance applications"*International conference on communication and signal processing (ICCSP),* India, 2018, pp. 0570-0575.

[12] Arka Prava Jana, Abhiraj Biswas, Mohana, "YOLO based Detection and Classification of Objects in video records" *2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT)* 2018, India.