Human Head Counting and Detection using Convnets

Savita S M
Department of Electronics and Communication
Dayananda sagar college of engineering
Banglore, India
Savitasm993@gmail.com

Shahla Sohail
Department of Electronics and Communication
Dayananda sagar college of engineering
Banglore, India
Shahla400@gmail.com

***Abstract*—** Now days, Detection of human head play an very important role in pedestrian counting. Machine learning is one platform, where human being can train a machine to act without being explicitly programmed and gives more accurate result, even when there is no enough data. Convolution neural network is one which works well for multimedia communication such as Text, Audio and Video. In this paper convnets play an important role in human head detection. In this paper it's going to explain the less number of layers with more accuracy in the results with less time consuming.

Keywords- *Hyperparameter,Feature maps,Convnets(CNN) and SGD.*

*****

I. INTRODUCTION

Convolutional neural networks are invented by D H Hubel and T N Wiesel. These networks are similar to all other neural network but it differs in convolutional layer only. In this network neurons are the preprocessing element, these play very important role in CNN. Convnets are feedforword multiperceptron layer where weights are going to share in each layers so weights play very important role in CNN. Where bias the term its act like threshold to avoid over fitting of images. In each layer hyper parameters are introduced. In Machine learning we have two types of learning processes one is supervised learning [2] and another one is unsupervised learning[1]. In supervised learning we have to train our network in the area of interest only, in this learning stage you are the teacher to the network you are forcing you to the network to learn only those features. Unsupervised nothing but randomly we are extracting features.

II. CONVNET ARCHITECTURE

Convnets are very specialized network where it consists of convolutional layer, pooling layer and fully connected layers at the end. There are many layers in the network and each of them repeatedly using in network to reduce the size of the image (In this application). Convents consist of input layers and output layers in between them we find five layers and those layers are called 'hidden layers' .where in this paper we are using number of layers will be seven but most important layers will be only five layers. Fig 1 is example for convents how it looks likes. Selection of input image should be multiples of two's and filter size should be odd numbers so filter should convolve around the whole image.

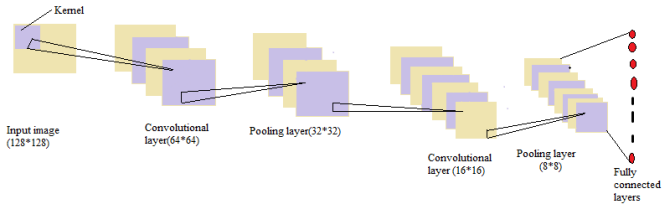

Fig 1.Convnets architecture

As shown in the above figure we are using 128*128 image sizes as our input to the network after normalization only. Then we are converted into Gray scale image so it's easy for calculations and our understanding also. Here in each layer the size of image will reduced to half of that previous image size. Hyper parameters are nothing but filter size and strides what we used in each layers .Filters are designed in such a way that each feature in the input image it going to learn each and every feature of the of that image .In convolutional layer a filter is introduced so that filter will convolves around the our input image so it will learn may be image edges or image corners or may be curves we can't predict the feature map, but we can predict the size of feature maps using formula as given below

$$F.M_{(i)} = (I - K + 2P) / S + 1 \text{ ----- } (1)$$

Above equation represents expecting output image size for next layer here layer nothing but a set of particular operations that take place, such as consider in convolutional layer the input image size is 128*128 and the output image size is 64*64, this output image is the input to next layer such as pooling layer these alternative operations going on in the overall network. Where i represents the no of layer ,I represents the input image size for network ,K represents the kernel size in convolutionals layers ,P is padding and S is stride in each layers. But in fully connected layer we are not using any hyper parameters they become one dimensional vectors. After fully connected we have classification in softmax layer.

III. FEATURE EXTRACTION

A. *Definition*

Extracting remarkable properties from the input which will easily understood for networks.

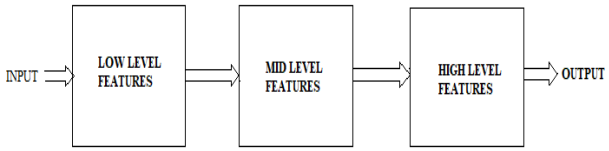

Fig.2. Feature extraction procedure.

Feature extractions take place in three types those are shown in following figure 2. when we given input as image to network it extracts the first low level features such as edges and curves.mid level feature[3] nothing but object parts(may be combinations of edges) and high level features are object models. We get the features in each layer of convents.

B. *Features maps* .

Feature map are the number of each kernel what we obtained in layer by layer output. These feature maps will increase in convolutional layer but these are remains same in pooling layer. These maps also called as activation maps.

IV. PROPOSED MODULE

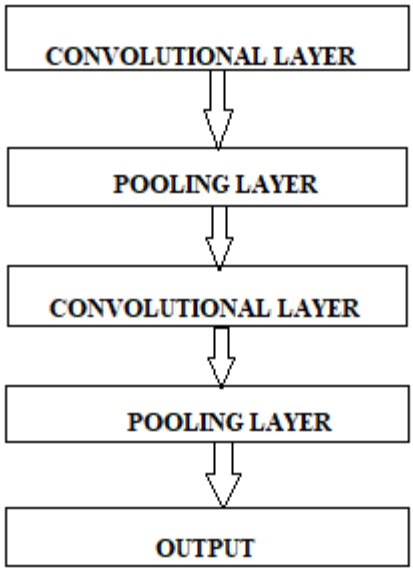

Fig.3. Framework for proposed methodology.

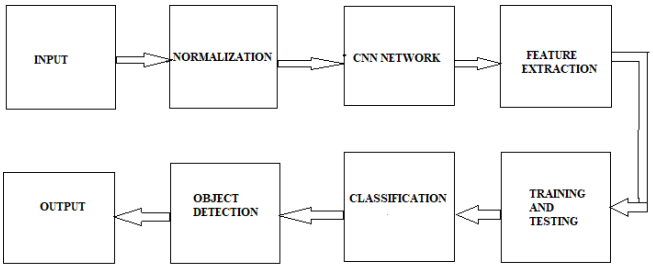

Fig .4.Typical block diagram for head detection.

Fig 3 its explain about the network is divided into five layers.CNN consist of input layer, hidden layers and output layer. Hidden layers are nothing but other than input layer and output layers. So the proposed method consists of four hidden layers. Here each layer consist of API'S .Those API'S help to operate each layer properly. Layer means set of operations.

Each convolutional layer consists of filter or kernel it will learn the each and every feature of an image. One important feature of CNN is to instead of detecting feature of given image it will learn the feature of that image it makes this network will fully different from conventional networks[4] .By learning those features of given image it will helpful get more accuracy.

A. *Abbreviations and Acronyms*

CNN- convolutional neural network (convnets), API-application programming language, NN –neural network.

ML - machine learning.

B. *Selection of hyaperparameters*

- Hyper parameter means selection of kernels and strides in each layer. In this module filter size is 3*3 and number of layers are five.

C. *Equations*

Cross correlations: let assume that where I is the input image and K is the kernel what we used in convolutions layers and there dimensions are K1*K2 it's given by the equation

$$(I * K)_{ij} = \sum_{m=0}^{k1-1} \sum_{n=0}^{k2-1} I(i+m, j+n) K(m,n)$$

-- (1)

Convolution operation given by below expression

$$(I * K)_{ij} = \sum_{m=0}^{k1-1} \sum_{n=0}^{k2-1} I(i+m, j+n) K(m,n)$$

-- (2)

$$(I * K)_{ij} = \sum_{m=0}^{k1-1} \sum_{n=0}^{k2-1} I(i+m, j+n) K(-m, -n)$$

-- (3)

$$E = \frac{1}{2} \sum_p (t_p - y_p)^2$$

-- (4)

From equation (3) we come to know that cross correlation is same as convolution. Where – means flipped kernel, i and j are the iteration of the input image and m and n rows and column.

There are two propagations take place in this whole process one is forward propagation and another one is backward propagation. If the any error in the forward pass and it's represented by the equation (4). 'P 'represents predicted output in forward propagation, t is target and y forward propagation output. This error is also known as Cost function.

In back propagation is works on SGD algorithm. In this algorithm each and every feature learns by the weights present in kernel based on the target given to the network. Where the each and every weight adjusted in such way that to minimized the errors and tries to get more accuracy is same as the target. When we achieved target or nearby accuracy of target the classification take place based on labeling what we did initially while initiating CNN network. For more information about the back propagation go through the references. As earlier only it mentioned that there are four stages in initialization of CNN, training, testing and finally classification take place.

As shown in Fig 4.when input is before giving to the network it normalized into some particular size i.e. multifications two's (128*128) and it converted into gray color image for understanding purpose. After normalization it given to CNN network and it extracts the features those features are shown in below Fig.5.

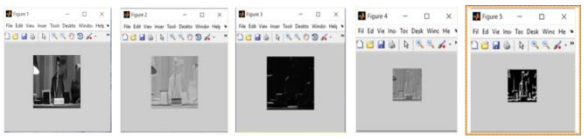

Fig.5.Obtained features from CNN

After feature extraction next step is training.

A. Training

The back propagation is called as training. There are totally four stages take place. Forward passes, loss function, backward pass and finally weight updating. During forward pass the whole image of their dimension m * n pass through the network. It consist of number of layers and it predicts the output image .whatever the prediction take place in at the end of the network it compare between the actual output image if both are same then no need of back propagation.

If both images are different those outputs consider as error function and difference of those functions calculations take place. I.e. called as cost function. Then backward pass take place and it adjust the weights according to the actual (target) output.

B. Testing

After training testing take place. We get accuracy by calculating MSE vs. NO of epoch. In this testing we come to known that for our requirement of any other business logic our network is working properly or not. Some experiments results as shown in below graph. Testing result is is giving more accuracy is about nearly 99.8%. So this network works awesomely while compare to the other network.

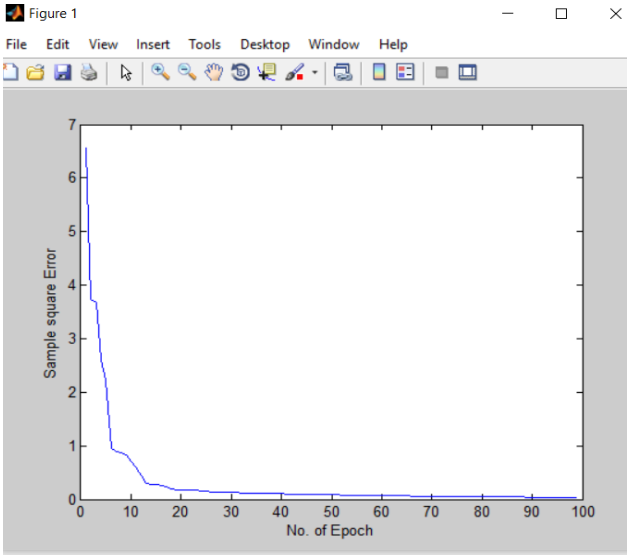

Fig.6.Testing results

C. Classification

In classification we have lot of algorithms[5] are there so we take help from those algorithms. Where finally we get the how any labeling we done at the initially .at the output of this softmax layer[6] we come to know that the results. This is final most steps of convents. And we have lot of API's based on that whole network will work more expectedly.

V.RESULTS

Both figures are represents the detection of human head and the counting of those heads. The results will help in get exact accuracy of people travelling in metro stations, railway stations, and pedestrian's detection and in fairs. Overall we can predicts the people accuracy and arrange things according to that number.

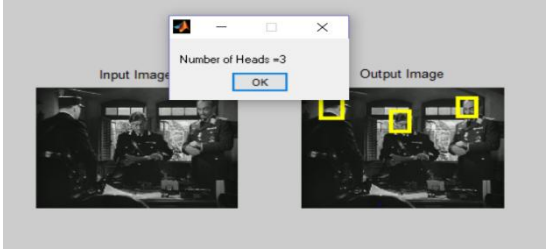

Fig.7.Obtaining output from gray color imag

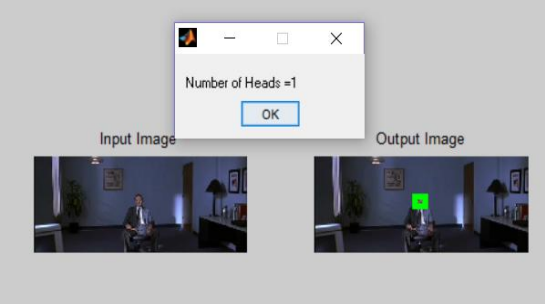

Fig.8. Obtaining output from color image.

VI.CONCLUSION

CNN play very important role in multimedia communication. Now a day's CNN is hot-topic because if we trained the network by some dataset and its has ability to detect those objects which is never we trained the things in the network. Nothing but it detects new objects such as our brain does. If we assume that our eyes are the filters it convolving around the whole image and try to extract the many important features and send into the number of nodes in the present in our brain, and finally store such things in our brain (hard disk).

VII. FUTURE WORK

Computer vision play very important role in digital life. As we know that demand increases for person centric information so try to put more efforts on this person centric information so labeling of each every dataset from some particular group we applied and out of that group entrance to the some restricted area it will going identified that person so it will helpful avoid the unauthorized persons entry.

REFERENCES

[1]. Pierre $ermanet Koray Kavukcuoglu $oumith Chintala Yann LeCun" Pedestrian Detection with Un$upervised Multi-Stage Feature Learning"2013 IEEE conference pp.3626-3633,2013.
[2].Iqbaal muhamamda and Zhu yan "supervised machine learning approaches "ijsc 2015.0133.
[3].Quin Tin,Bo Zhou, wen hua zhao, Yun Wei and Wei-wei Fei"Human head detection using HOG features of head and shoulder based on depth map.2013-academy publisher.
[4].Hui xu, 0Peng Cui,Lifeng sun, Shigiang Yang, Xifeng Ding "a cascade SVM approaches for head and shoulder detection using HOG"IEEE 2009.
[5]Andreas Kolsch, Muhammad Zehan Afzal, Markus Ebbeck and Marcus Liwicki "Real time image classification using extreme CNN".arxiv- 1711.2017.
[6].Deepika jaswal, soumya v, kp Soman"Image classification using deep CNN"IJRTC-2014.