OBSERVATION OF STANFORD DOG DATASET USING

CONVOLUTIONAL NEURAL NETWORK

Thota Naveen Babu

1904347

*University of East London*

*u1904347@uel.ac.uk*

**Abstract:** The Coursework is mainly focusing on the Classification of images and to find the accuracy of Stanford Dog dataset which contains 120 breeds of dogs from the world .The Dataset is categorized by the images of dogs, annotations from the image Net of the fine gained images. From those 120 breeds of dataset we chose 15 breeds which are having 1900 images. The main aim is to find the classification of images by using image processing and to find the accuracy of Dog dataset we used Convolution Neural Network (CNN) algorithm.

1. **Introduction:**

The Stanford dog dataset has been downloaded from the Kaggle.com, dataset contains 795MB. Dataset contains 120 different breeds of dogs having over 20000 images, from that we have chosen 15 breeds of dogs having over 1900 images and loaded the dataset into MATLAB. All images are having different pixels sizes so, we introduced image processing technique and converted all images into 120x120 pixels. All the images are in three dimensional arrays and converted them into two dimensional arrays by using RGB2Gray function. Later on, all the Grey scale images are stored in the image data store and we created the labels for all the classes. Each class having 149 images and out of them 115 images are used for training and remaining 34 images to testing. For more image classification and feature extracting of an object we used Convolutional Neural Network algorithm to train the images. In CNN we used convolution and relu layer , maxpool layer as a input layers for all the images to extract the features, and batch normalization layer, classification layers act as output layers of CNN network are used for the classification of all images. In convolution2dlayer filter size is given as 3 and number of filters is given as 24,48,64,128 respectively and in max pooling layer step size is given as 2 stride. After training all the images in CNN network, validation accuracy has been shown by using the training network graph and we got the final accuracy of 80% after 30 iterations.

1. **Methodology:**

Convolutional neural network is deep learning neural network mainly used for classification of images. CNN is also very useful for finding the patterns in all the images and recognize the objects accordingly. From the image data they learn directly and by using the image patterns they do classification and eliminate the need of manual feature extraction.

Convolutional Neural networks are built using different layers and each layer is made up of different no of neurons depending upon its weight. All neurons are inter connected to its previous layer and post layer neurons. The first one is input layer. The last one is output layer, this will be shown as predicted result. The remaining layers will be in between input and output layer. These will do the process and pass data to one after other. Every neuron will send data to its next layer neuron this will be the dot product. Neural networks are defined as its no of layers and no of neurons involved it.

**2.1. Architecture of Convolutional Neural Network (CNN):**

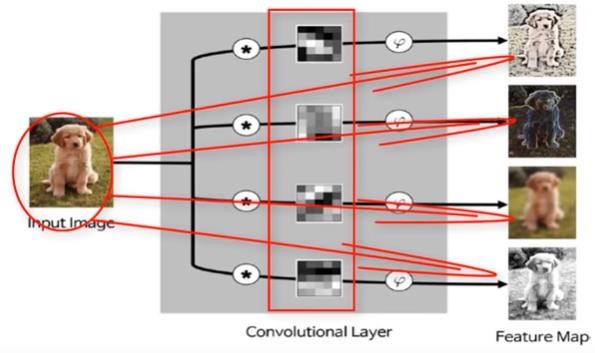
CNN has different layers, convolution and relu layer , max pool layer as a input layers for all the images to extract the features, and batch normalisation layer, classification layers act as output layers of CNN network are used for the classification of all images.

1. **Feature Learning**: The input images enter into the feature extraction Network the

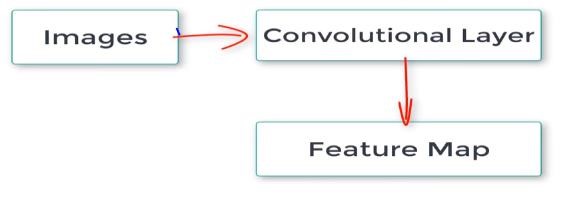
extracts feature signals enter the classification Neural Network

2 **Classification Section**: The Classification Neural Network operates based on the feature of the images and generates the output. The feature extraction Neural Network consists of the pile of the convolution layer and pooling layer pairs. The convolution layer, as its name implies converts the image using the convolution of operation. it can be thought of as collection of as a collection of digital filters. The pooling layer combines the neighboring pixels into single pixels therefore the pooling layer reduces the dimension of the image. As the primary concern of CONVNET is the image the operation of the convolution and pooling layer s are conceptually in 2 dimensional planes. This is one of the diff b/w ConvNet and another neural network. Soft max layer is also used for image classification.

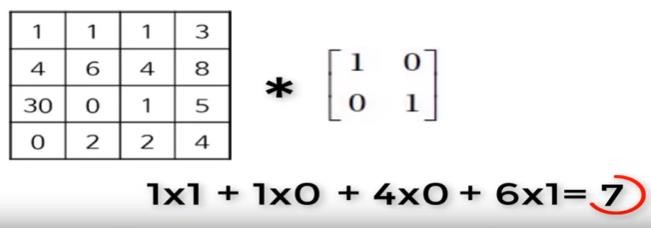
**Convolution layer:** Convolution is one of the main building blocks of a CNN. The term convolution refers to the mathematical combination of two functions to produce a third function. It merges two sets of information. CNNs is powerful, and these are not perfect for visualizations.

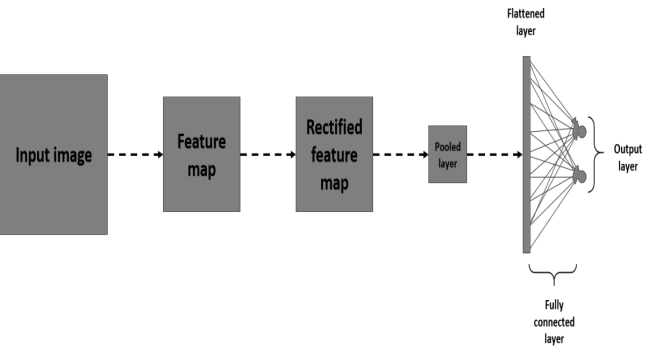


Feature map is to be produced with an filter by using input data and it is done by convolution neural network.We execute a convolution by using filter to the input. In every point, a matrix multiplication is performed and sums the result onto the feature map. The area of our filter is also called the receptive field, named after the neuron cellsThe size of this filter is 3x3. It generates the feature maps from the images. The working principle of this layer is different than other Neural Network layers.

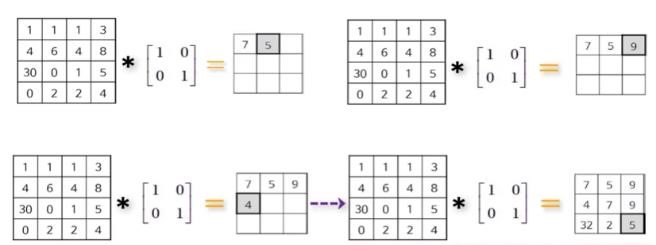


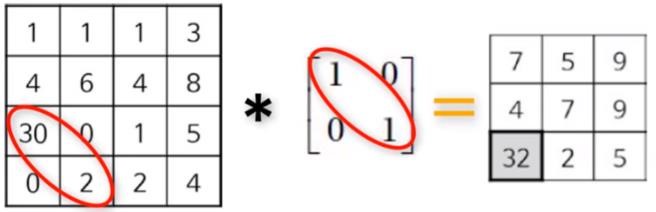
it does not employ connection weights and a weighted sum instead, it contains filters that converts image. We call this filters, convolution filters the number of feature map and the no of convolution filter is same. That means if there is 4 conventional layers then it generates 4 feature maps the filter of convolution layer are 2 dimensional matrices. To understand how convolution works let us show some example of the mathematical matrix multiplication of convolution layer by taking 4x4 pixel image with 2 convolution filters.



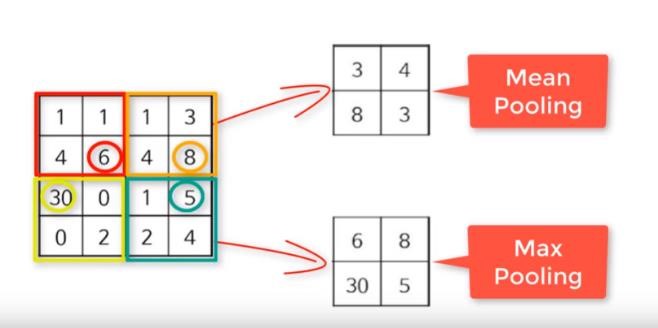


I am taking these filters arbitrarily, but in actual ConvNet , these filters are determined through training process. Let’s start with the 1st filter the convolution operation begins at the upper left corner of the sub matrix that is the same size as the convolution filter. Let’s apply convolution this 1 will be multiplied with this 1. This 0 will be multiplied with this and this 1 will multiplied by this 6 the result of the multiplication will be added together and the value of the addition will replace this entire block . And by doing this matrix multiplication for all blocks we will get a result that 4x4 pixel image will be converted into 3x3 pixel.





**Pooling layer:**The task of the pooling layers the reduce the size of the image the operation of pooling layer is very simple and straightforward to understand how it is works let’s take same the image to perform convolution operation. There are 2 types of max pooling and mean pooling. From mathematical perspective calculating the mean pooling and max pooling is convolution operation. The difference between the convolutional layer and this convolution operation is stationary convolution filter and the convolution area do not overlap. therefore, if we take 2\*2 matrix as the convolution area, there are 4 convolution areas for the mean pooling, the calculation are done by taking the means of the convolution areas. For example, the mean of 1st area is (1+1+4+6)/4=3the mean value of next convolution area is (1+3+4+8)/4 which means 4.



In this same way the mean value of next area is 8 and 3 respectively. This is how the mean pooling values are generated. Calculating the value of max pooling matrix is very easy simply take the largest value of each convolution area and form the matrix. The highest value of the first convolution area is 6 ,2nd convolution area is 30, 3rd is 30 and 4th is 5. The max pooling layer is generated using the values.

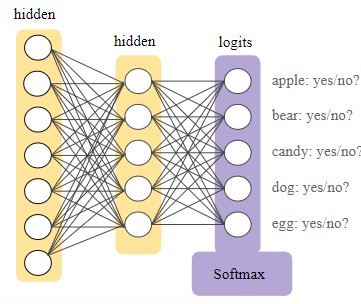
**ReLU** is act as a activation function and to contrast the difference by using sigmoid and rectified function, we use sigmoid and rectified function respectively to see very good result.

**Stride** is the size of the step the convolution filter moves each time. A stride size is usually 1, meaning the filter slides pixel by pixel. By increasing the stride size, your filter is sliding over the input with a larger interval and thus has less overlap between the cells.

**Classification layer:**

After learning features in many layers, the architecture of a CNN shifts to classification. The next-to-last layer is a fully connected layer that outputs a vector of K dimensions where K is the number of classes that the network will be able to predict. This vector contains the probabilities for each class of any image being classified. The final layer of the CNN architecture uses a classification layer such as soft max to provide the classification output.

**Soft max layer** act as a logistic regression which helps us to provide categorical output of a image, and this function is better than RELU to give classified output. Soft max function will calculate the probability of each class.



The soft max layer has the same nodes as a output node. By using logistic regression, it will find all the probabilities of each class. This soft max function is mainly used to reduce the losses of the training data.

Fully connected layer: Each neuron receives input from every element of the previous layer. The number of inputs of all input image classes are equal to the number of outputs of a fully connected layer.

1. **Simulations:**

* 1. **Introduction of dataset:** The stand ford dog dataset contains 120 different breeds of dogs having over 20000 images, from that we have chosen 15 breeds of dogs having over 1900 images. Each class having 149 images and out of them 115 images are used for training and remaining 34 images to testing and all images are in different folders which have different images count and having different pixels, so it is difficult to train the dataset in the Neural network. By using image processing we made the dataset flexible and used it to solve the neural network problem by using convolution neural network.

* 1. **Input encoding:**

Step 1: A new path is created and stored our dataset and loaded that file in Matlab. images=('C:\Users\naveen\Documents\MATL

AB\latest'); filenames=dir(fullfile(images,'\*.jpg')); count=numel(filenames); **Step 2**: By using image processing technique we resized all images into 120x120 pixels and converted into Grey scale images, because the Convolutional neural network cannot work if all images are in different pixel sizes.

for k=1:count thisimage= filenames(k).name; filename=[trained\_images '/' filenames(k).name]; Img1=imread(filename);

I1=imresize(Img1,[120 120]); Gray = rgb2gray(I1); figure; imshow(Gray);

title('Gray image’);

end

**Step 3:** All images are stored in image Datastore, all image data is converted into computer vision format like as numbers. And are been created for all 15 classes which is having a count as 148 and total of all images in 15 classes are about 1900 images. In each class we have used 115 images for training and remaining 34 for testing by using split function.



imds = imageDatastore(images,'IncludeSubfolders',tr ue,'LabelSource','foldernames'); label = countEachLabel(imds); img = readimage(imds,1); size(img) mincount = min(labelCount{:,2}); numimages = 148; minSetCount = min(numimages,mincount); imds1 = splitEachLabel(imds, minSetCount,

'randomize'); countEachLabel(imds1) numTrainFiles = 115; [imdsTrain,imdsValidation] =

splitEachLabel(imds1,numTrainFiles,'random

ize');

Step 4: Convolution neural network is created and we have used 3 layers as a input feature extraction and 2 output layers for image classification. Image input layer is nothing but pixel size of all images which is 120x120 pixels and in convolution layer we used 3 filters and we doubled the size of filter in other layers as well. And we used one activation function that is rectified linear unit also a pooling layer is used with an step size

of 2 layers = [ imageInputLayer([120 120 1]) convolution2dLayer(3,24,'Padding','same')

batchNormalizationLayer reluLayer maxPooling2dLayer(2,'Stride',2)

convolution2dLayer(3,48,'Padding','same')

batchNormalizationLayer reluLayer maxPooling2dLayer(2,'Stride',2) convolution2dLayer(3,64,'Padding','same')

batchNormalizationLayer reluLayer maxPooling2dLayer(2,'Stride',2) convolution2dLayer(3,128,'Padding','sa

me') batchNormalizationLayer reluLayer fullyConnectedLayer(15) softmaxLayer

classificationLayer

];

**Step 5:** Validation accuracy graph is shown using training Options graph by giving the epoch 30 and we got the final accuracy of 80%.options = trainingOptions('sgdm','InitialLearnRate',0.01 ,'MaxEpochs',30,'Shuffle','everyepoch','ValidationData',imdsValidation,'Valid ationFrequency',6,'Verbose',false,'Plots','traini ng-progress'); net =

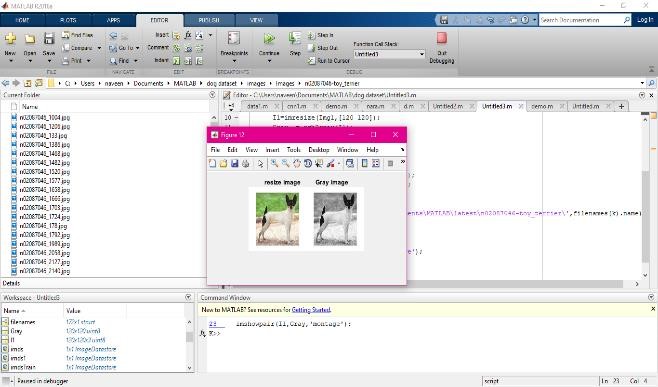
trainNetwork(imdsTrain,layers,options); YPred = classify(net,imdsValidation); YValidation = imdsValidation.Labels; accuracy = sum(YPred == YValidation)/numel(YValidation)

**3.3 Procedure:**

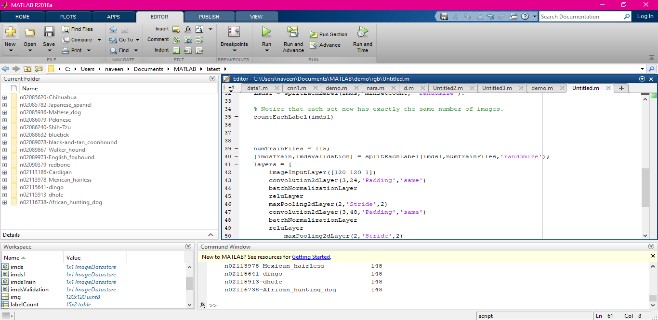
The dog dataset has 120 different classes having of 20000 images, out of that we have chosen 15 breeds having 1900 images and loaded all the dataset in the matlab by creating a new path. Labels are created for each class by giving count as 149 each, we stored the data in image data store and then all images are converted into number pixels. Here, Image processing methods are used to resize all the images and all the images are converted into color to Gray scale images. By using split function from 149 images we used 115 images for train and 34 images for test. Moreover, here matrix multiplication will take place among pixels of images and convolution filters. The input image pixel is dot product with filters and to get the result where the input image is matched to the output image of the network, the position of input and output images should be same and by using dot product it should be highest value. Convolution Neural Network are used to train the dataset. In CNN, we used three layers for feature extracting of all the images and two layers for classification. In convolutional layer 3 filters are used and increased the filter size as 24,48,64,128 and for preventing the feature map we used ‘padding’. RELU is act as a activation function and to get very good result we used sigmoid method. Max pooling layer is used to reduce the image size and here there are two types of pooling one is max pooling and second one is mean pooling and we give the step size as 2 stride. Fully connected layer is a output layer and classification as well. Softmax function is a output layer where it will give the better result than Rectified linear unit function and give the stable output. By using training network graph we have shown the final validation accuracy of 80% and 30 epochs with an 30 iterations.

1. **Results Obtained:**

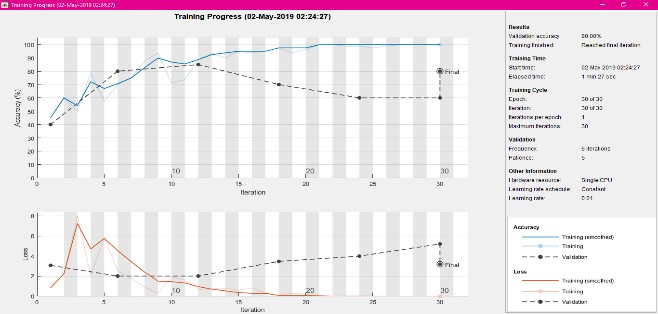
Using Image processing all images converted into 120x120 pixel and Grey scale images.



Labels are created for each class having count of 148 and 115 images used for training and 33 images for testing.



Final validation accuracy is 80% with an epoch of 30 and iteration 30.



**4.1 Critical analysis of results:**  Initially we faced an issue to create a labels, because in every class of our dataset we have different count in each folder, we sort this problem by creating label for each class and given the minimum count of 148 of each class and we used 115 images to train the network and 33 are used to test our network.

In convolution layer at first, we have chosen

2 filters and filter size as 12 but later we used 3 filters and increased filter size as 24 and doubled in the next layers respectively, and also, we used pooling layer as step size 2. For colored images we got very less accuracy because all the images are in threedimensional plane so it was difficult to train the network in CNN, so we have converted all the images into Grey scale and we got a validation accuracy of 80.00% with less iteration of 30 and epochs used are 30.

1. **Conclusions:**

Image classification is done by using the Convolution Neural Network and it is a

deep learning methodology having more hidden layer where image classification can be very easily by using the Neurons of the input image and also feature extraction can be done by using the input weights of all class. Mainly to get the better result we converted all the images into 120x120 pixels, and by using image processing method all images converted into three dimensional to two dimensional images where as training process can be done easily for all the images. If all images are in colored then the training network cannot detect the feature of images so the information might be loss. Grey images are very useful to get better output. Basically, in our dog data set contains different categories of dogs with less similarities, so it is difficult to train our network then we increased the number of neurons by adding more layers and more filters are used in convolution layer and filter sizes also increased accordingly. By using the option training network we got the final validation accuracy of

80%.Moreover we used 30 epochs and 30 iterations to get stable accuracy of our output. And the loss gets decreased at the end of the network. Image processing technique is most useful us to get better accuracy because, here it was unbale to detect if all imaged are colored and we converted them into Grey scale images and finally we got 80% accuracy.

1. **Reference:**

International Journal of Engineering & Technology Website:

www.sciencepubco.com/index.php/IJET Research Paper.

Convolutional Neural Network for Image

Classification Chen Wang Johns Hopkins University Baltimore, MD 21218, USA cwang107@jhu.edu Yang Xi Johns Hopkins University Baltimore, MD 21218, USA yxi5@jhu.edu.

Simard, P.Y., Steinkraus, D. and Platt, J.C. 2003 Best practices for convolutional neural networks applied to visual document analysis, volume 1. IEEE.

Thomas M. Mitchell. . 1997. . Machine Learning. , McGraw-Hill, Inc., New York,

NY, USA, 1 edition. Paul J. Werbos. . 1974. .

Beyond Regression: New Tools for

Prediction and Analysis in the Behavioral Sciences. , Harvard University Press.

Stanford dog dataset from [www.kaggle.com](http://www.kaggle.com/)

[https://www.kaggle.com/jessicali9530/stanford-dogsdataset/kernels](https://www.kaggle.com/jessicali9530/stanford-dogs-dataset/kernels)

Classifying the images having more classes from free code camp

[https://medium.freecodecamp.org/how-to-classifyphotos-in-600-classes-using-nine-million-openimages-65847da1a319](https://medium.freecodecamp.org/how-to-classify-photos-in-600-classes-using-nine-million-open-images-65847da1a319)

Deep Learning with MATLAB – matlab expo2018

Introducing Deep Learning with the MATLAB – Deep Learning E -Book provided by the mathworks.

CNN concepts from free code camp. [https://medium.freecodecamp.org/anintuitive-guide-to-convolutional-neuralnetworks-260c2de0a050](https://medium.freecodecamp.org/an-intuitive-guide-to-convolutional-neural-networks-260c2de0a050)

[https://uk.mathworks.com/solutions/deeplearning/convolutional-neural-network.html](https://uk.mathworks.com/solutions/deep-learning/convolutional-neural-network.html) [https://uk.mathworks.com/help/deeplearning /ref/nnet.cnn.layer.convolution2dlayer.html](https://uk.mathworks.com/help/deeplearning/ref/nnet.cnn.layer.convolution2dlayer.html) [https://www.quora.com/Why-is-aconvolution-neural-network-harder-to-trainon-color-images-compared-to-grayscaleimages](https://www.quora.com/Why-is-a-convolution-neural-network-harder-to-train-on-color-images-compared-to-grayscale-images)

Deep learning concepts for Neural network in Math work. [https://uk.mathworks.com/discovery/neuralnetwork.html](https://uk.mathworks.com/discovery/neural-network.html)

Object recognition in Math work. [https://uk.mathworks.com/solutions/deeplearning/object-recognition.html](https://uk.mathworks.com/solutions/deep-learning/object-recognition.html)

[https://hackernoon.com/visualizing-parts-ofconvolutional-neural-networks-using-kerasand-cats-5cc01b214e59](https://hackernoon.com/visualizing-parts-of-convolutional-neural-networks-using-keras-and-cats-5cc01b214e59)

[https://www.youtube.com/watch?v=ZOXOw](https://www.youtube.com/watch?v=ZOXOwYUVCqw)

[YUVCqw](https://www.youtube.com/watch?v=ZOXOwYUVCqw)