Shri Vile Parle Kelvani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

Approved by AICTE and Affiliated to the University of Mumbai

Department of Electronics & Telecommunication Engineering

Mini Project On

Title: **Cat and Dog Recognition**

SUBMITED BY:

Name of Student	SAP ID
Sanjeet Krishna	60002160050
Viren Baria	60002160005

Teacher's Name: Prof. Vishakha Kelkar

Shri Vile Parle Kelvani Mandal's

DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING

Approved by AICTE and Affiliated to the University of Mumbai

Department of Electronics & Telecommunication Engineering

CERTIFICATE

This is to certify that Mr. Viren Baria,

SAP ID 60002160005 of TE EXTC 1: has submitted their

Mini Project for Subject Name for the Academic Year 2018-2019.

Guide

Examiner

Head of Department

EXTC Department

Index

Introduction:

Image classification is a fundamental problem in computer vision. Deep learning provides successful results for machine learning problems. Many algorithms like minimum distance algorithm, K-Nearest neighbour algorithm, Nearest Clustering algorithm, Fuzzy C - Means algorithm, Maximum likelihood algorithm are used for the purpose of image classification. Here, image classification is performed using convolutional neural network which is became standard after since AlexNet (Alex Krizhevsky, Geoff Hinton and Ilya Sutskevar) won Image Net in 2012. Generally convolutional neural network uses GPU technology because of huge number of computations but, in proposed method we are building a very small network which can work on CPU as well. The network is trained using a subset of Kaggle Dog-Cat dataset. This trained classifier can classify the given image into either cat or dog. The same network can trained with any other dataset and classify the images into one of the two predefined class.

Software Used: Python

Flowchart:

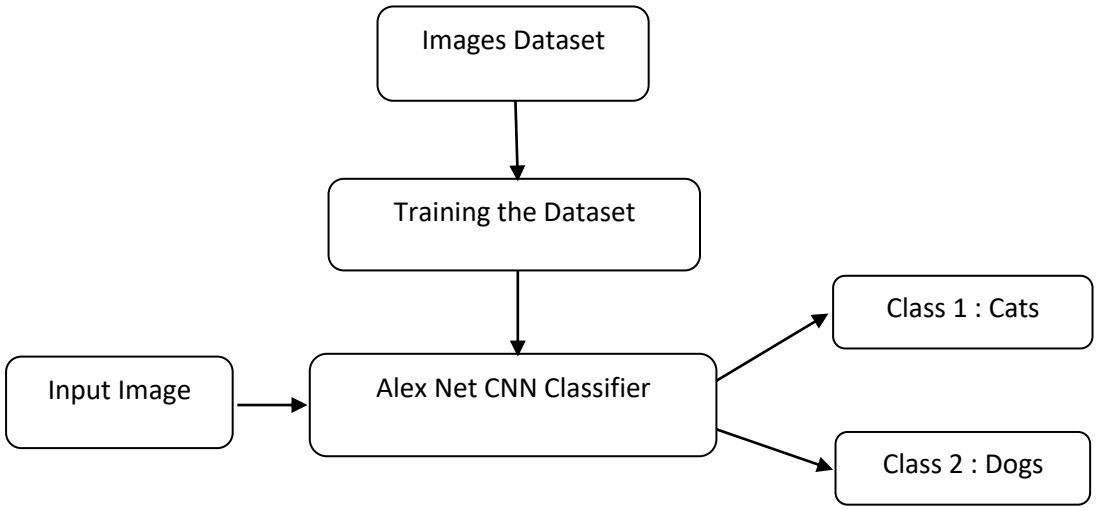


Fig.1. AlexNet CNN Model

Theory:

AlexNet solves the problem of image classification where the input is an image of one of 1000 different classes (e.g. cats, dogs etc.) and the output is a vector of 1000 numbers. The i^{th} element of the output vector is interpreted as the probability that the input image belongs to the i^{th} class. Therefore, the sum of all elements of the output vector is 1.
The input to AlexNet is an RGB image of size 256×256. This means all images in the training set and all test images need to be of size 256×256.
If the input image is not 256×256, it needs to be converted to 256×256 before using it for training the network. To achieve this, the smaller dimension is resized to 256 and then the resulting image is cropped to obtain a 256×256 image.

If the input image is grayscale, it is converted to an RGB image by replicating the single channel to obtain a 3-channel RGB image. Random crops of size 227×227 were generated from inside the 256×256 images to feed the first layer of AlexNet. Note that the paper mentions the network inputs to be 224×224, but that is a mistake and the numbers make sense with 227×227 instead.

AlexNet Architecture

AlexNet was much larger than previous CNNs used for computer vision tasks (e.g. Yann LeCun's LeNet paper in 1998). It has 60 million parameters and 650,000 neurons and took five to six days to train on two GTX 580 3GB GPUs. Today there are much more complex CNNs that can run on faster GPUs very efficiently even on very large datasets.

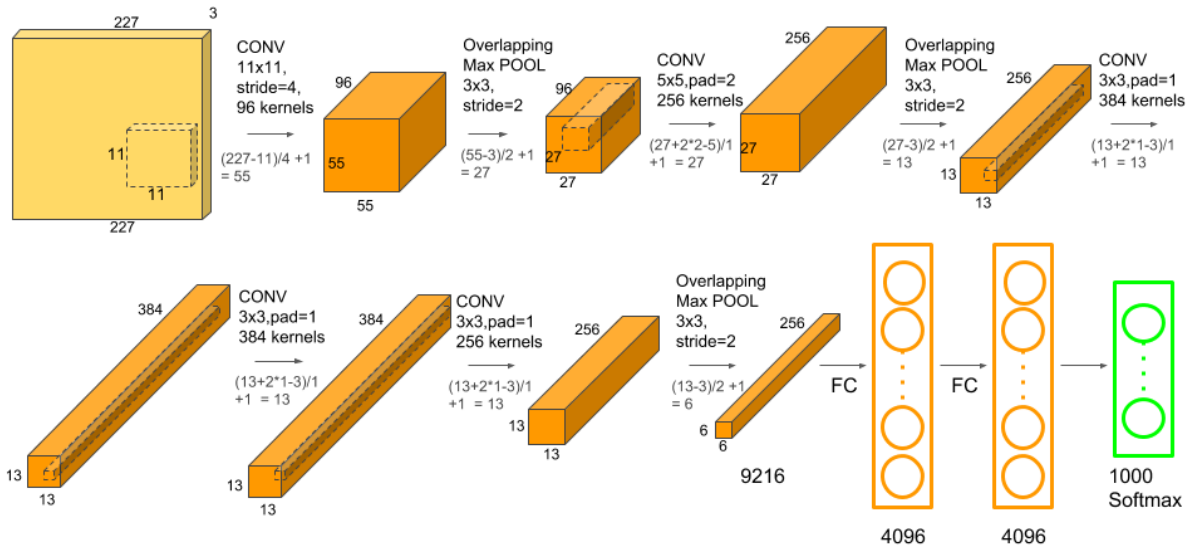


Fig.2.AlexNet CNN Architecture

AlexNet consists of **5 Convolutional Layers** and **3 Fully Connected Layers**.

Multiple Convolutional Kernels (a.k.a filters) extract interesting features in an image. In a single convolutional layer, there are usually many kernels of the same size. For example, the first Convolution Layer of AlexNet contains 96 kernels of size 11x11x3. Note the width and height of the kernel are usually the same and the depth is the same as the number of channels.
The first two Convolutional layers are followed by the Overlapping Max Pooling layers that we describe next. The third, fourth and fifth convolutional layers are connected directly. The fifth convolutional layer is followed by an Overlapping Max Pooling layer, the output of which goes into a series of two fully connected layers. The second fully connected layer feeds into a softmax classifier with 1000 class labels.

ReLU (Rectified Linear Unit) nonlinearity is applied after all the convolution and fully connected layers. The ReLU nonlinearity of the first and second convolution layers are followed by a local normalization step before doing pooling.

Results:

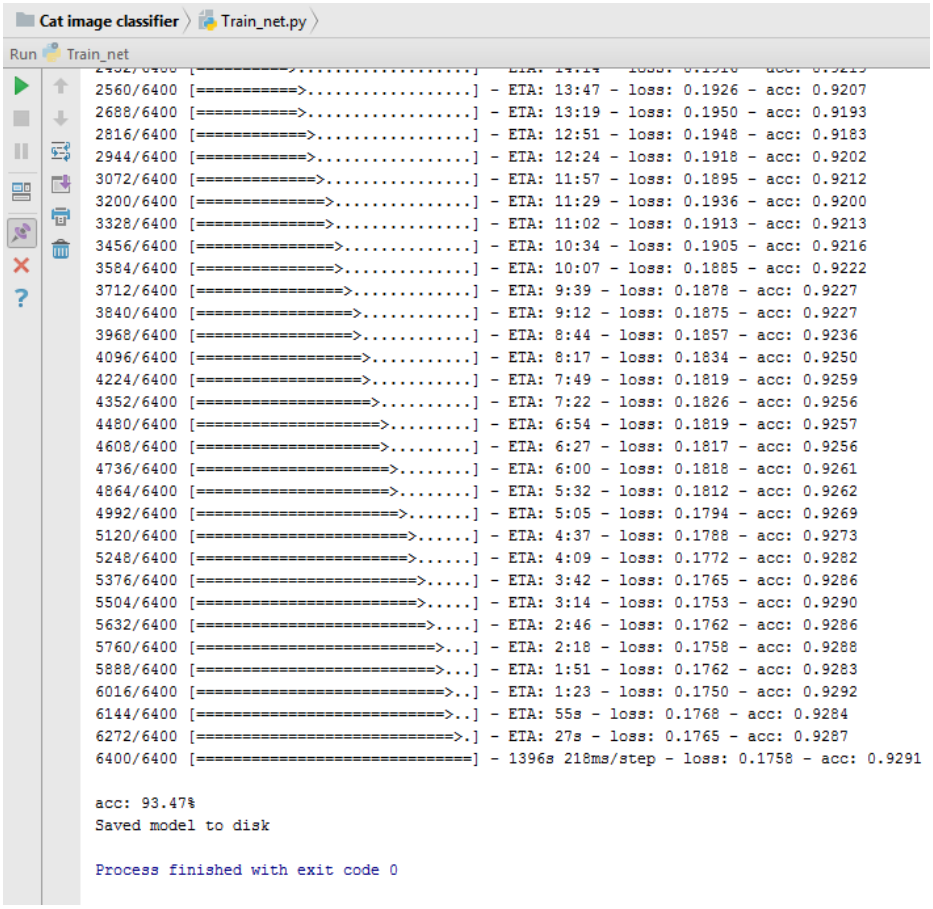
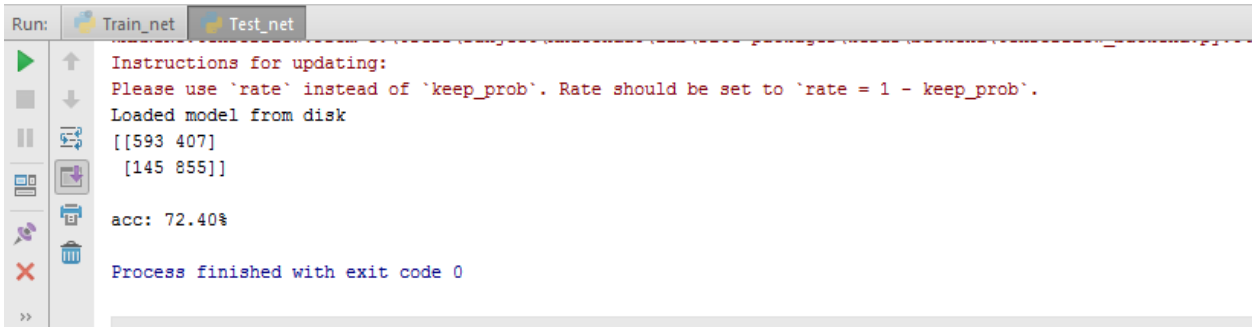
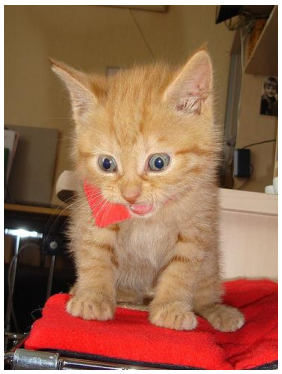


Fig.3. Training the Dataset





```
Loaded model from disk
dog.jpg
[[0.869275]]
dog

Process finished with exit code 0
```

```
Loaded model from disk
cat1.jpg
[[0.21600792]]
cat

Process finished with exit code 0
```

Fig.4. Testing the Dataset

Conclusion:

We have successfully performed a classification process using Alex Net convolutional neural networks. Although there are many algorithms that perform image classification, convolutional neural network is considered to be a standard image classification technique. Convolutional neural network uses GPU technology because of large number of layers which increases the number of computers. Therefore, we presented a very small convolutional neural network which can work on CPU as well. This network classifies the images into one the two predefined classes say Cat and Dog.

References:

1. Shanmukhi, M., et al. "CONVOLUTIONAL NEURAL NETWORK FOR SUPERVISED IMAGE CLASSIFICATION." *International Journal of Pure and Applied Mathematics* 119.14 (2018): 77-83.
2. You, Yang, et al. "100-epoch imagenet training with alexnet in 24 minutes." *ArXiv e-prints* (2017).
3. Gonzales and Woods, "Digital Image Processing", Pearson Education, India, Third Edition.