Classification of Dogs and Cats

Rajeev Kanth Reddy
rmeda@umail.iu.edu

Karthik Sreenivas
kartsree@umail.iu.edu

Introduction

We take object classification problem of determining whether the animal is a cat or a dog from an image. We applied and compared the following techniques for image classification.

1. ***Support Vector Machines***
2. ***Bag of Words***
3. ***Convolutional Neural Networks***
4. ***Classification based on region growing***

For our problem took data from a Kaggle competition named Dogs Vs Cats. It consists of 12500 images of each cats and dogs for training and 12500 test images.

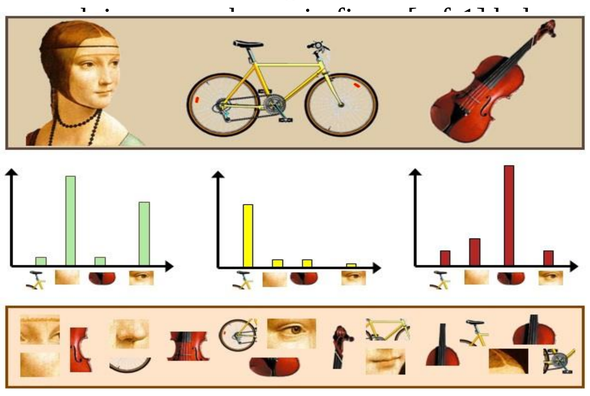
Bag of Words

Build a visual vocabulary

- First we resize all the images into either 32*32 and get SIFT descriptors for all each images in the train folder.
- Then cluster all the 128D SIFT descriptors into 100 clusters using K-Means algorithm.
- We save the centers of these 100 clusters. These represent the visual vocabulary for our model.

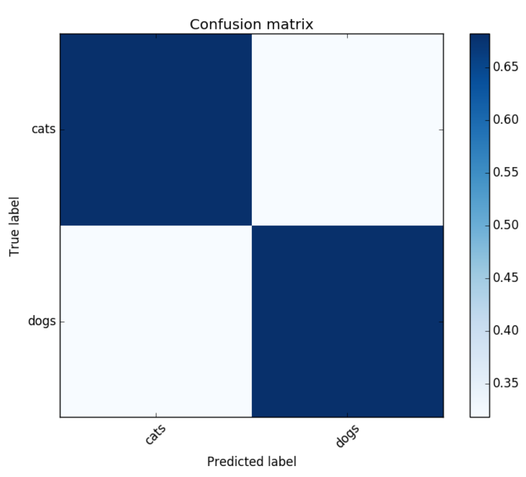
Represent each image as a histogram of the visual vocabulary.

- First we calculate the SIFT descriptors for the image.
- Then look for the nearest visual word for each descriptor and generate a histogram for



Finally use SVM to train the vectors containing histograms of each image and use the trained model to classify any new image given.

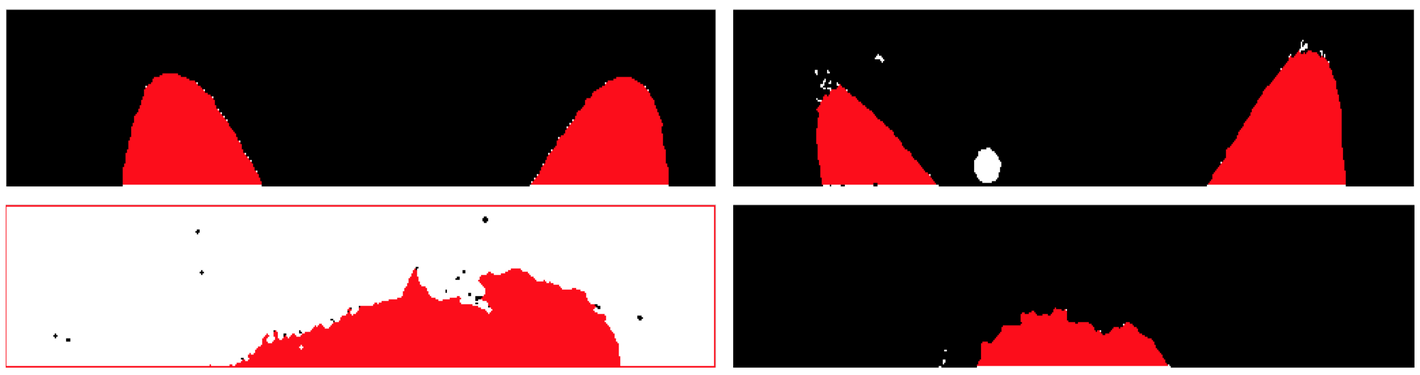
Confusion Matrix



Classification based on Region Growing

- ***Feature identification***: We try to first find a unique identifier which can classify cats from dogs. Here we have selected the "ears" as that feature which can differentiate cats from dogs. The criteria behind this is that the ears of cats are pointed upwards whereas most dogs have their ears inclined towards their face.
- ***Pre-processing***: We convert the image to black and white and also applied the median filter to eliminate noise. Since our region of interest is the portion of the ear we extracted the ear portion of the image. The next step was to take count of the number of black and white pixels in the extracted portion of the image. We identified that the count of pixels for whichever color was greater formed the background or the undesired part of the image the pixel count of the color in smaller number contained our region of interest.
- ***Region growing***: Suppose the number of black pixels are greater than the number of white pixels, we scan the extracted portion of the image from the very first pixel until we encounter a white pixel. When a white pixel is encountered we check its neighbors to see if there are any other white pixels in its vicinity. If present we move towards that particular direction and successively check its neighbors until we reach a point where a white pixel has no more white pixels in its vicinity. This forms one region. Then from the point where we first encountered the white pixel we continue in search of other white pixels which have not been visited. This requires us to keep track of the co-ordinates of the white pixels visited. When we encounter a white pixel which has not been previously visited we again apply the same region growing principle. We continue this till all pixels in the entire image has been scanned. Then we take count of the number of regions. Since cats have the presence of ears in the extracted portion of the image the number of regions will be more for cat images when compared to dog images.

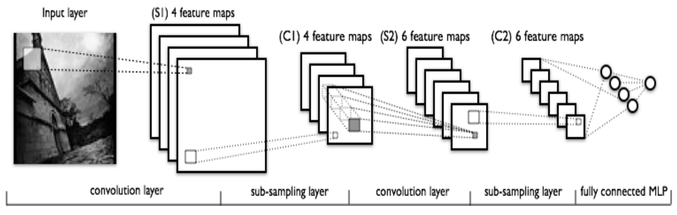
Here are some example cases.
These are the successful cases. The two images in upper portion are classified as cats and the bottom two as dogs.



Here is an unsuccessful case.

Convolutional Neural Networks

We implemented LeNet 5 architecture using caffe deep learning library. This 5 layer model has layers as following [CONV-POOL-CONV-POOL-CONV-FC]. Figure[ref-2] shows the model in detail.



Support Vector Machines

We resize each image to 40x40 and take it as 1600 D vector and apply SVM to train images. Using this trained SVM we predict classes.

We got an accuracy of 58% with this model.

Acknowledgments

- Poster template taken from http://colinpurrington.com/
- Packages used – opencv, SVM Multiclass, Caffe
- Recognizing and Learning Object Categories Li Fei-Fei, Rob Fergus, Antonio Torralba ICCV 2005 – [ref-1]
- LeNet 5 model diagram taken from deeplearning.com [ref-2]