**Abstract:**

In this assignment, our task was to classify images of cats and dogs using Bag of words model. We have trained 500 images of each class, and achieved average accuracy of 96%. We have used K-fold cross validation. We used both local and global features to classify the images and get highest accuracy when we select combination of both. For global feature, we select histogram of each image with 64 bins and used normalize results to make a visual vocabulary. In addition, we also used Haar-like features as our local set of feature. The images were pre-process using Gaussian interest point detectors and all the local and global features have been computed at every interest point.

**Our Approach:**

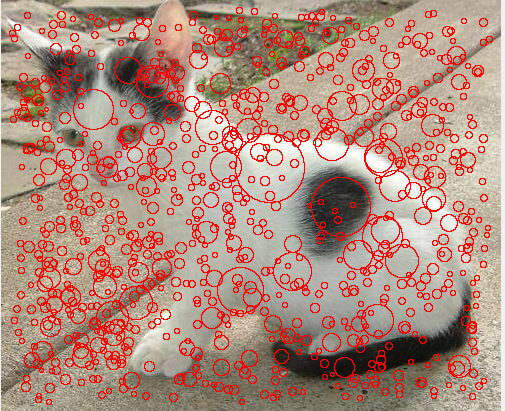
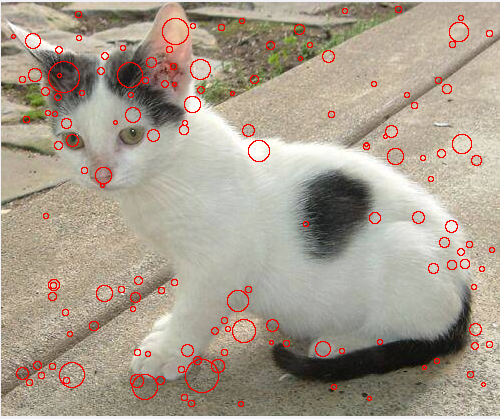
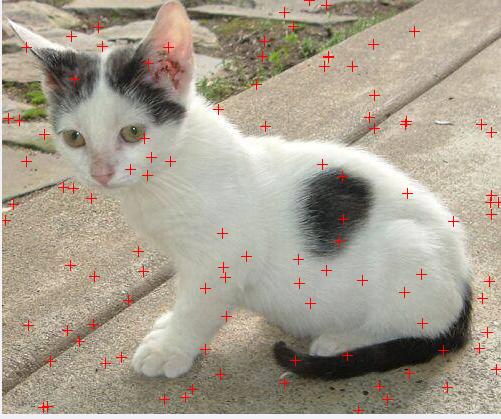
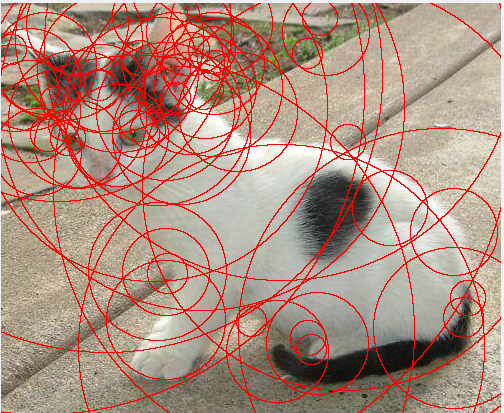
We have used rapid miner as a tool to classify images. We used a plug-in of Image processing and mining which provides around 170 processes related to images. Our approach could be concretely divided into the following steps:

1. **Selection of Pictures:**

We select a sample of images from training set and apply the “Set contrast” operation in order to avoid the noise in images. In addition, we also label the images as Cat and Dog for training.

1. **Interest Point Detection:**

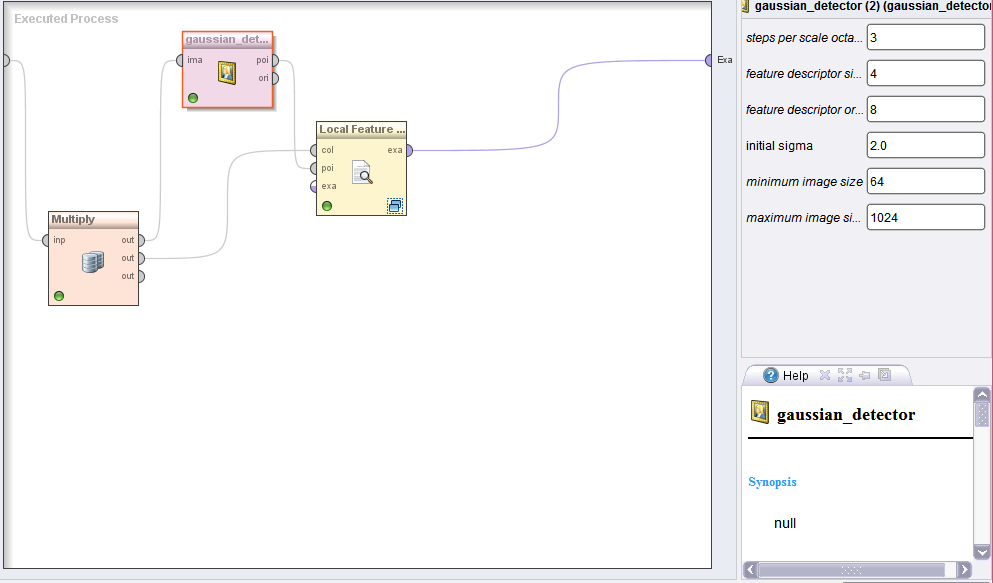
We use Gaussian interest point detection method in order to find interest point in the images. Previously we applied the following interest point detectors and found the Gaussian to be the best among them. Visualization of each interest point on a single image described below:



*Upper Left:* Harris Interest Point, *Upper Right:* Random Interest Point, *Lower Right:* Hessian

*Lower Left:* Gaussian Interest Point detector

Since the Gaussian interest point detector has most appropriate points among four, we choose that.

****

1. **Local Feature Extraction in Points:**

After interest point detection, we extract three global and one local feature from each interest point. The features and their parameters:

1. Histogram:-

We use histogram global feature. We set the bin size of 64 and normalize the output values

1. Edge Haar-Like feature:-

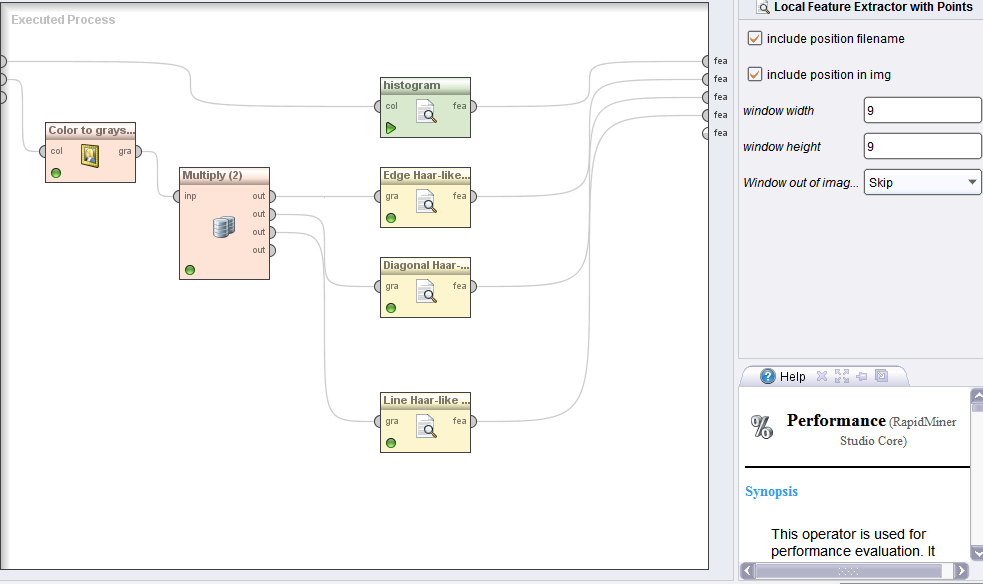
We use edge Haar-like feature and select Horizontal type of feature.

1. Diagonal Haar-like feature:

We use Diagonal Haar-Like feature with window size of 9x9.

1. Line Haar-like feature:

We use Line Haar-like feature with horizontal orientation and size of X-axis as 3 and y-axis as 1.

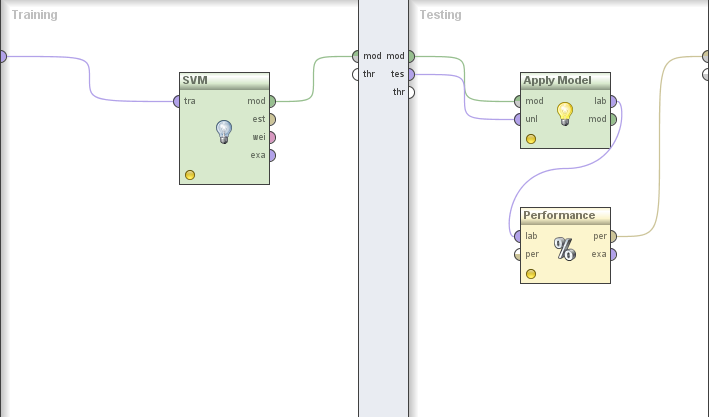


1. **Clustering:**

The above four-process compute the respected features on every interest point of an image and return an example set consisting of patch with 67 features each. We then cluster them in to K-mean and set K equals to 100. Therefore, our visual words would have 100 clusters with each cluster representing a code word and together making a dictionary.

1. **Classification:**

After representing each image as histogram through Cluster Centroids, we did K fold cross validation using SVM (Linear) as classification model.



1. **Results:**

