Restricted Boltzmann Machine

Its Application to Image Classification

Siddhartha Chandra
CVIT, IIIT Hyderabad

Getting the Features

Level 1

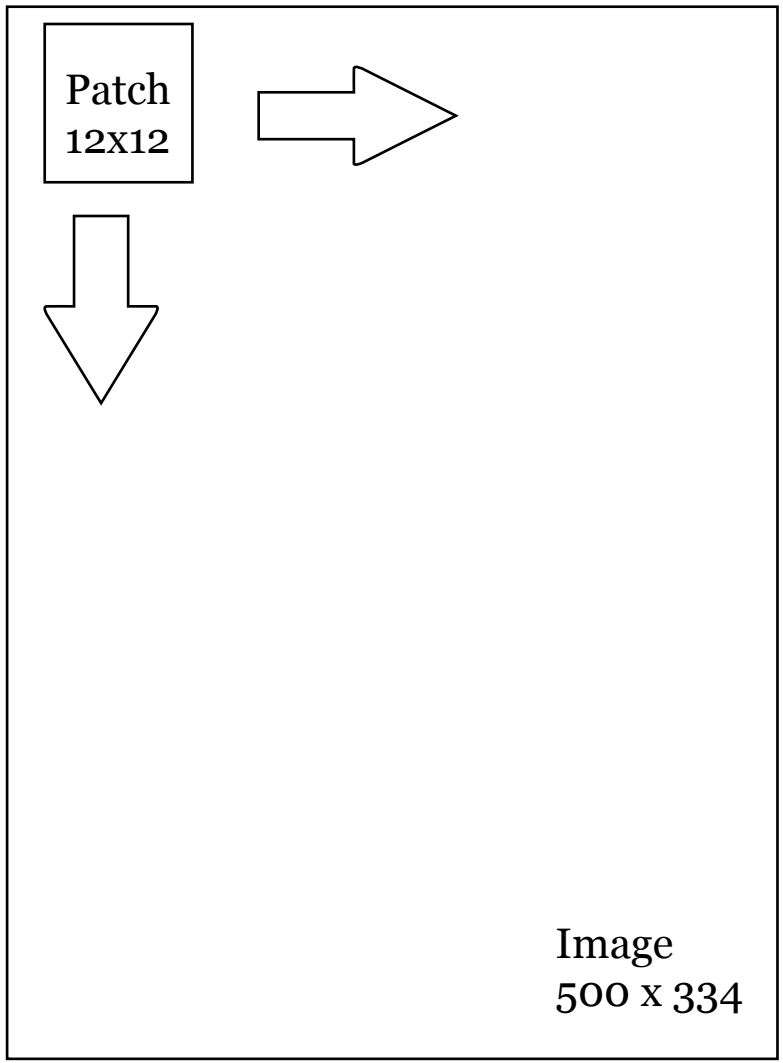


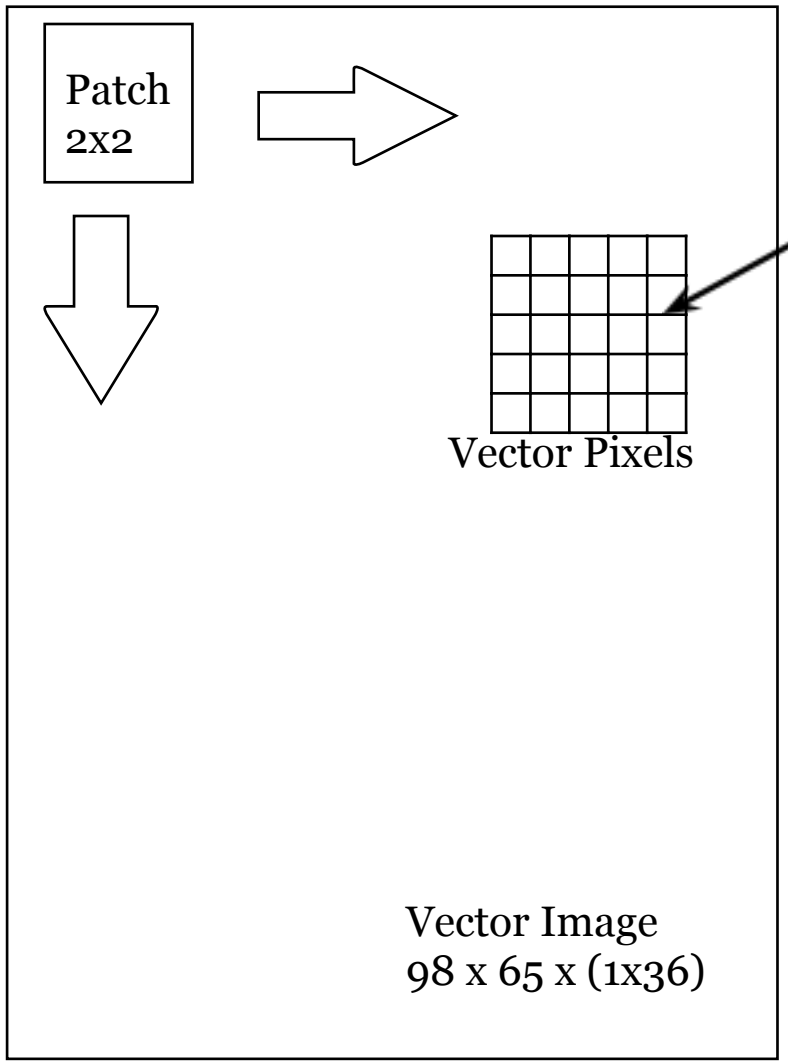
Image Size : 500x334
Level 1.
Patch Size: 12x12 = 1x144
Shift=5.
#Patches per Row=98 length(1:5:500-11)
#Patches per Col=65 length(1:5:334-11)

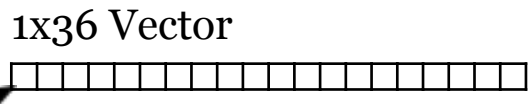
Patches: 98*65 = 6370
Size of Patch Representation of Image = 6370 x 144

After RBM Dimensionality Reduction (1x144 -> 1x36):
Size of Feature Representation of Image = 6370 x 36

Getting the Features

Level 2





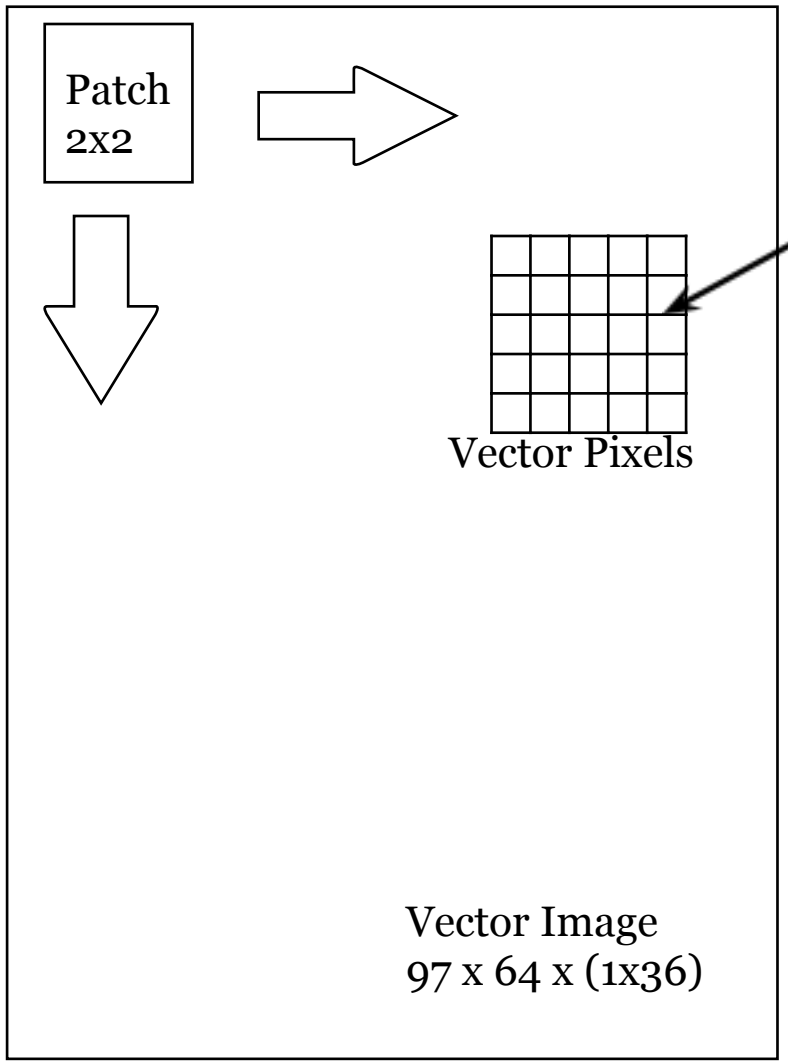
Vector Image, each Pixel is a 1x36 vector
Image Size : 98x65
Level 2.
Patch Size: 2x2 = 4
Shift=1.
Vector Size = 4*(1x36) = 144
#Patches per Row=97 length(1:1:98-1)
#Patches per Col=64 length(1:1:65-1)

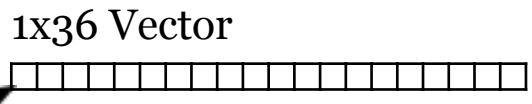
Patches: 97*64 = 6208
Size of Patch Representation of Image = 6208 x 144

After RBM Dimensionality Reduction (1x144 -> 1x36):
Size of Feature Representation of Image = 6208 x 36

Getting the Features

Level 3





Vector Image, each Pixel is a 1x36 vector
Image Size : 97x64
Level 3.
Patch Size: 2x2 = 4
Shift=1.
Vector Size = 4*(1x36) = 144
#Patches per Row=96 length(1:1:97-1)
#Patches per Col=63 length(1:1:64-1)

Patches: 96*63 = 6048
Size of Patch Representation of Image = 6048 x 144

After RBM Dimensionality Reduction (1x144 -> 1x36):
Size of Feature Representation of Image = 6048 x 36

Summary

Image Size: 500x334
Patch Size: 12x12 = 1x144
Patch Representation Size: 6370x144
Level 1 Feature Representation Size: 6370x36
Level 1 Feature Representation Size: 6208x36
Level 1 Feature Representation Size: 6048x36

Bag Of Words Model

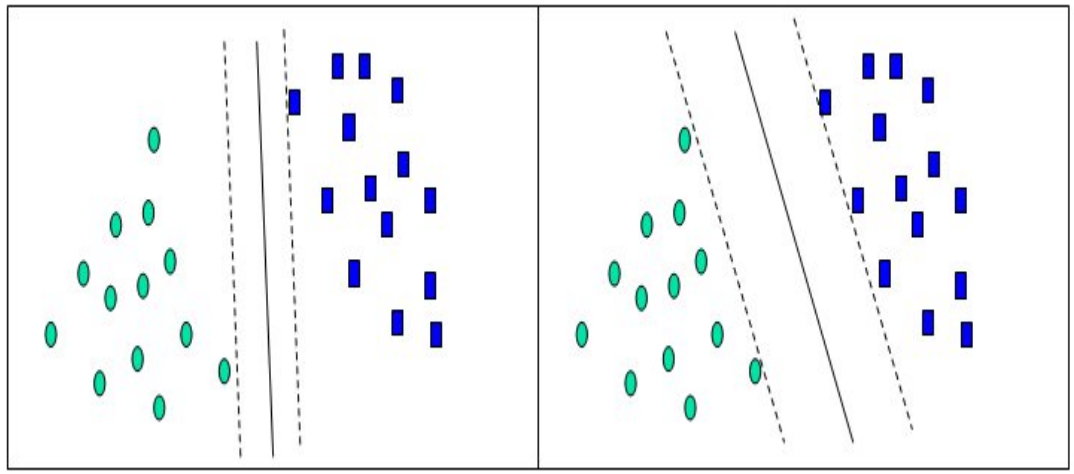
Vocabulary Computation

- Take 30 Random Training Images
- Each Image has a Feature Representation of size nx30
- (n ~ number of 1x36 Features, depends on Initial Size of the Image.)
- Run K-Means on all the 1x36 features. K = Vocabulary Size or number of Visual Words. Say K=1000.
- We get 1000 Cluster Centers / Words. Each Word is a 1x36 Vector.

Histogram Binning

- We use these 1000 words as Bins to bin the features in each Image.
- Binning is basically assigning a feature in an image to its nearest word/cluster center.
- Binning gives us a 1000 vector or a Histogram for each Image. The value of each cell in the vector is the number of features that were assigned to the word corresponding to the cell index.
- We normalize this Histogram so that the sum of all values in the vector is 1.
- **We have the 1x1000 Histogram Representation of each Image.**

Using an SVM as a Classifier



Pegasos SVM

- Each Image has a 1x1000 Histogram Representation.
- We train the SVM using the Histograms of the Training Images.
- We classify the Validation Set Images using the trained Model.
- We get the classifier scores for each of the Validation Set Images.
- We plot the PR Curve using the scores and the Validation Set Images' Class Labels.
- We compute AP as the performance measure of the Classification task.

Experiments

1. Observing Classification Performance using Features at Different Levels
2. Observing the Variation of Classification Performance with size of Reduced Features (number of Hidden Units in the RBM)
3. Comparing the Classification Performance with Bag of Words model trained on SIFT Features.

Status Page

http://researchweb.iiit.ac.in/~siddhartha.chandra/GooglePages/