Scene Recognition

By
Athithyaa and Sunil

Problem Definition

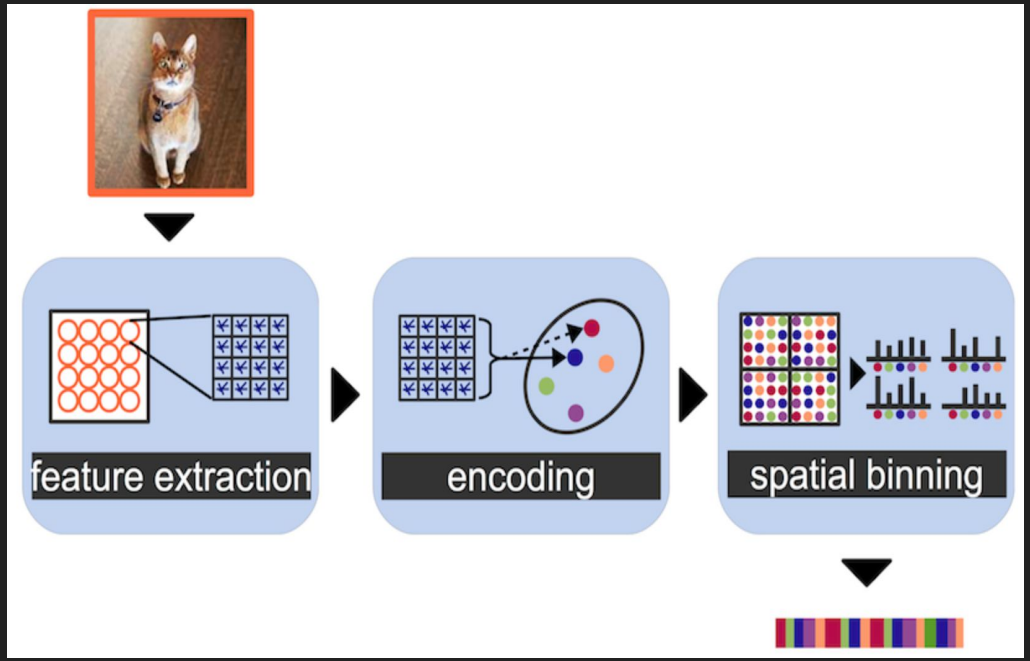
1. Special case of image classification

2. 15 Scene categories

3. Multiple approaches
    a. Bag of Visual Words
    b. Convolutional Neural Networks

Bag of Visual Words

- Feature extraction
- Learning visual words / codebook representation
- Quantization of features
- Visual word frequency - Histogram
- Learning the model - SVM



Bag of Visual Words using GMM

1. Use the same pipeline
2. Employ SIFT/SURF feature extraction
3. Apply k-means to get visual vocabulary cluster
4. For each cluster apply a m component GMM
5. Calculate histogram based on GMM score for each training image

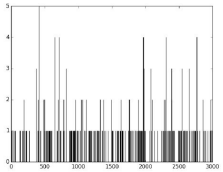
Convolutional Neural Networks

1. Quite different from traditional neural networks
2. Employ convolutional neural network architectures using Theano, Lasagne and Tensor Flow
3. Challenges: Computationally intensive
4. Alternative solution: Use pre trained model like AlexNet, GoogLeNet Inception model etc.

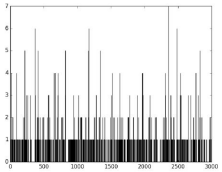
Results

Method	Accuracy
F_0:Bag of Visual words using SIFT	~39%
F_0 + Normalization of histogram	~41.2%
Bag of visual words using SURF (descriptors of 64 dimension with orientation)	~44%
Bag of Visual words using Gaussian Mixture model with a custom scoring scheme	~53.00%
F_1: Bag of visual words using SURF (descriptors of 128 dim without orientation) + Normalization	~56.3%
Bag of Visual words using Gaussian Mixture model	~56.48%
F_1 + 1000 clusters	~58%
F_1 + 2000 clusters	~62.8%
F_1 + 3000 clusters	~63.08%
ConvNets with inception tensorflow	~89.4%

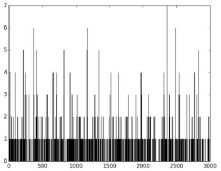
Coast



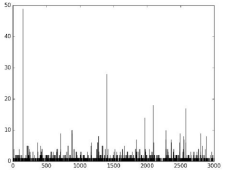
Forest



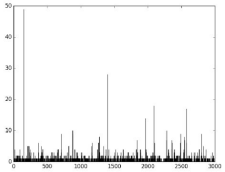
Highway



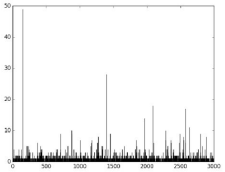
Kitchen



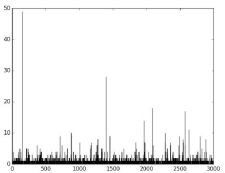
Inside city



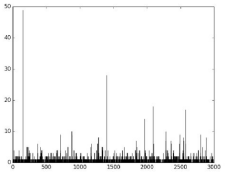
Mountain



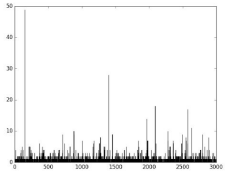
Office



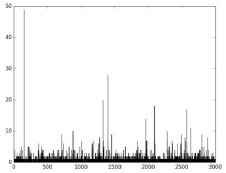
Livingroom



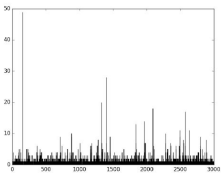
Open Country



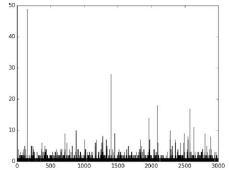
Street



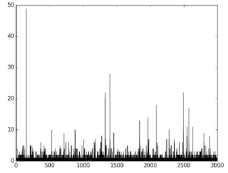
Suburb



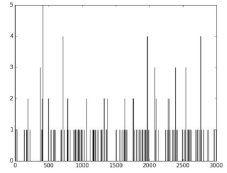
Store



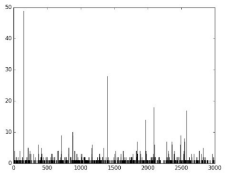
Tall building



Bedroom



Industrial



Demo

Thank You

Questions ?