Team 1: Place Recognition

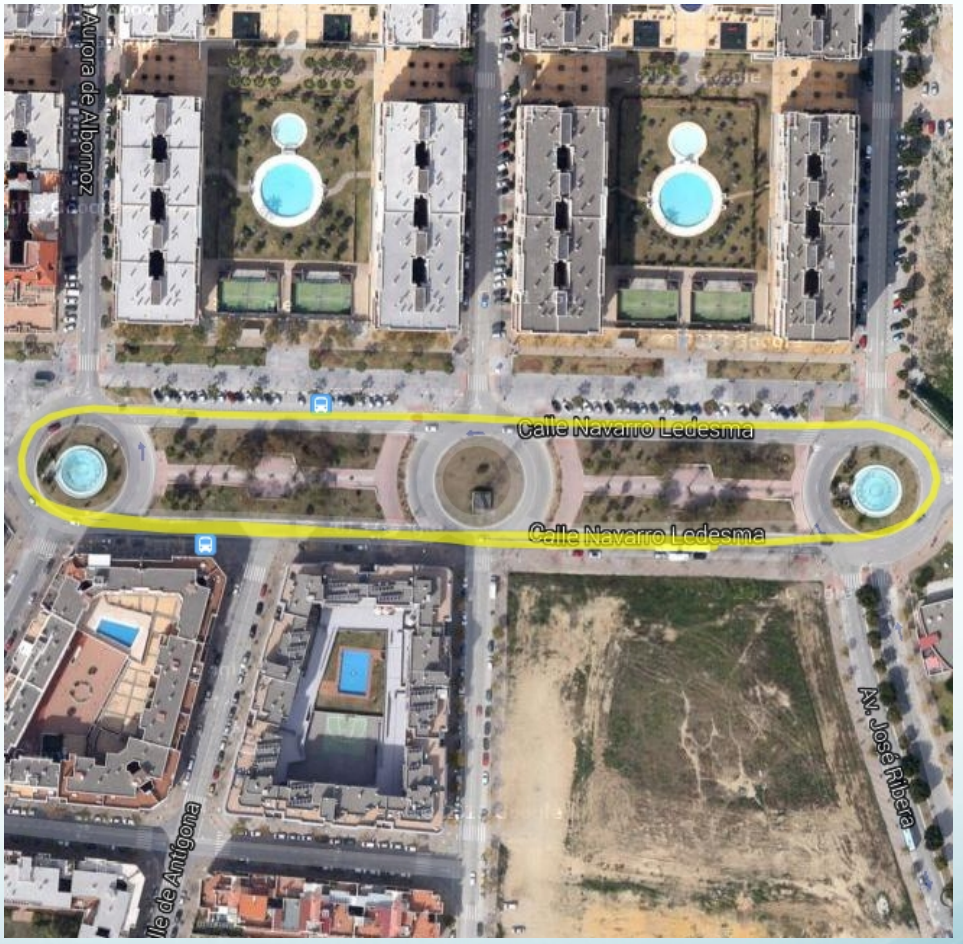
Derek Reiersen, Esther Vasiete, Shuzhe Zhang,
Kelly Kaoudis, Margaret Wheeler, Jeeeun Kim

Project Overview

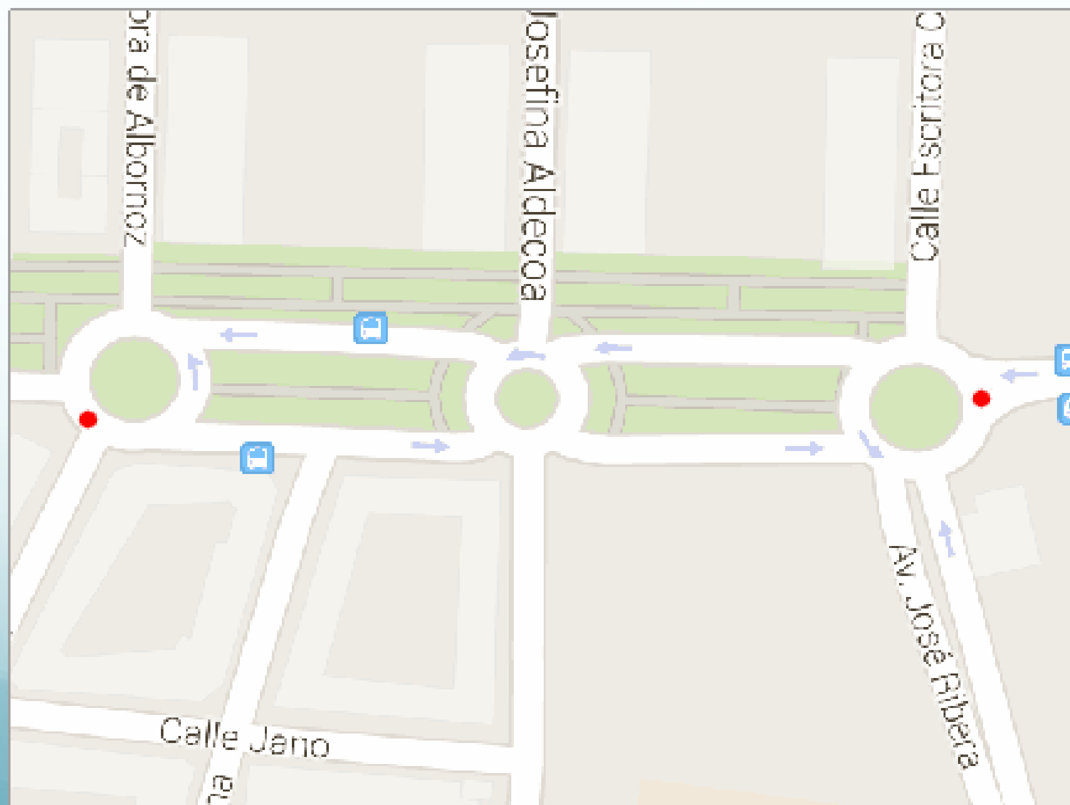
- Place Recognition using two different methods
  - Bag of Words with SIFT descriptors
  - Features extracted from a Convolution Neural Network
- Classify with One-Against-All SVM
- Test for loop closure on Malaga Urban Dataset
- Test for loop closure on custom local Boulder university urban campus dataset (**clbuucd**)

Malaga Urban Dataset Extract #7

- “Short avenue loop closure”
- ~1700 raw stereo images
- Dataset also includes camera data, GPS data, LIDAR data
- Use left camera and GPS data
  - camera for place recognition
  - GPS data to visualize path and loop closure



http://www.mrpt.org/MalagaUrbanDataset



ora de Albornoz
Josefina Aldecoa
Calle Escritora C
Av. José Ribera
Calle Jano

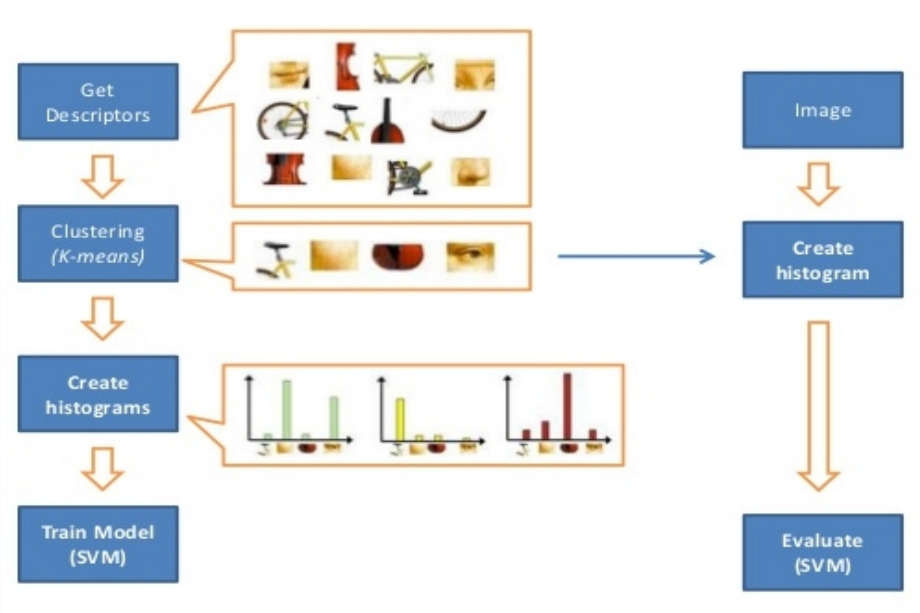
Campus Dataset, **clbuucd**

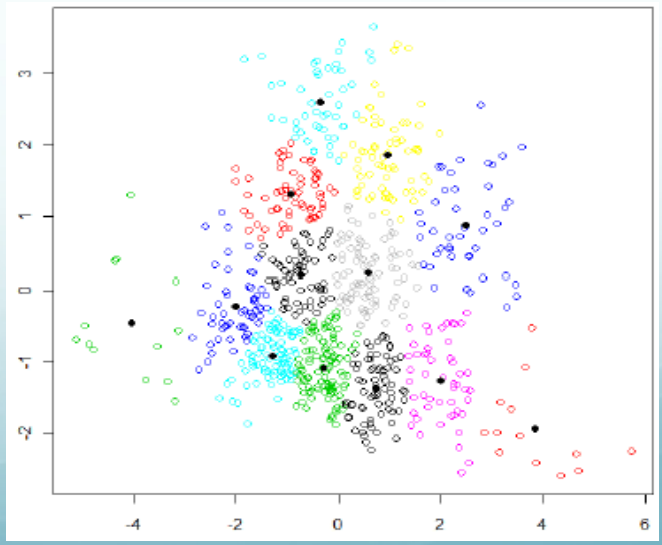
- Short *walking* loop closure
- 248 images
- Images distance ~22 ft
- two complete loops
- GPS data, mono images
- Greater challenge than Malaga Urban Dataset
  - People in images
  - Walking, not driving
  - Fewer images overall


Colorado Ave
Colorado Ave
University of Colorado Boulder
Baker Dr
Cockerell Dr
Regent Dr
Center for Community
Kittredge Loop

Method 1: Bag of Words

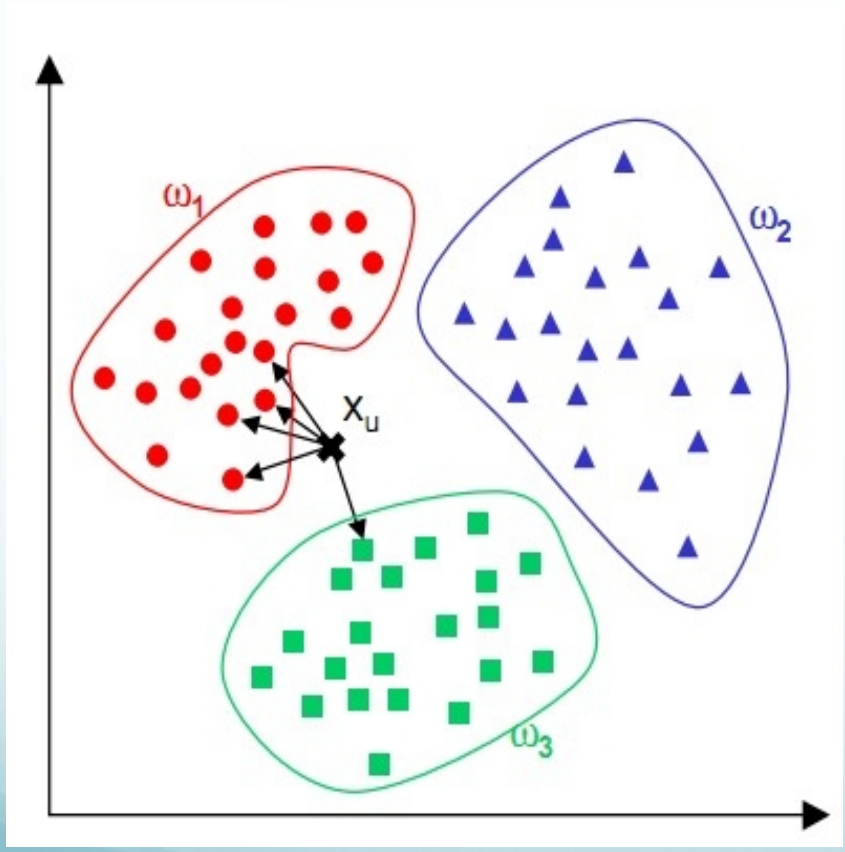
- Def: treat an image like a text document, break up into visual “words”
- Detect and extract *dense* SIFT feature descriptors
- Normalize feature descriptors
- K-Means clustering to create pictorial codebook
- Create histogram(s) of visual words



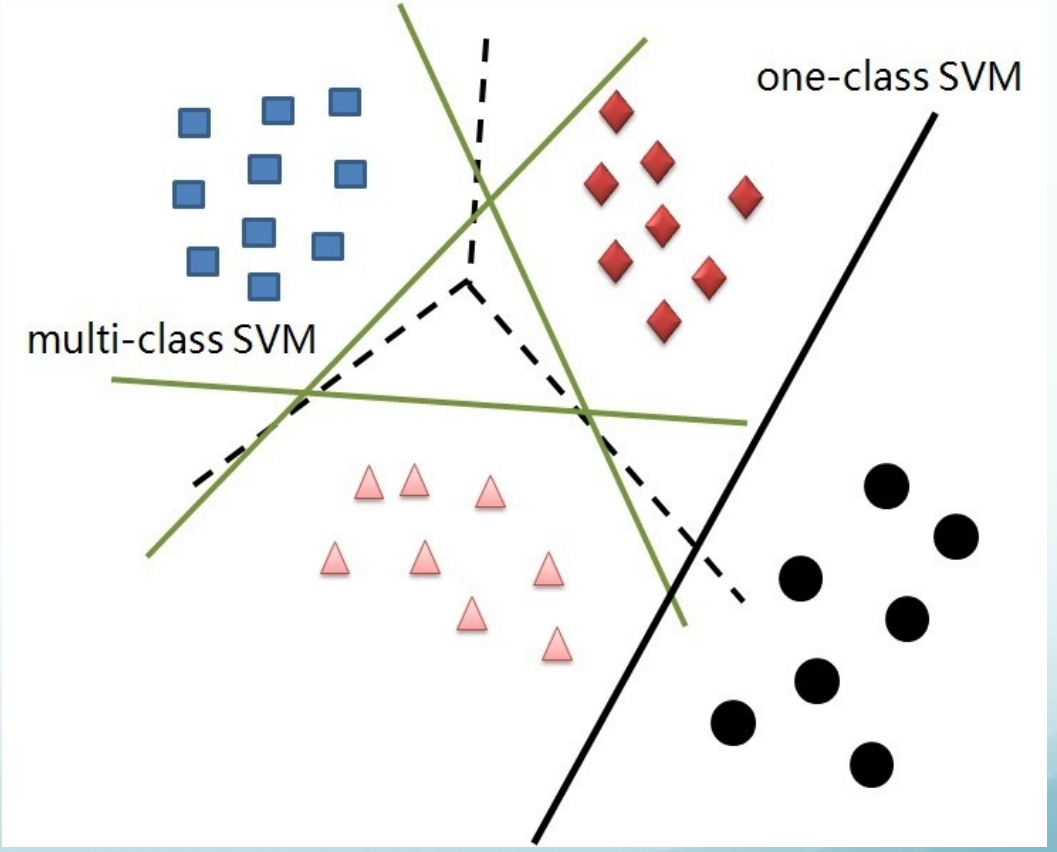


Bag of Words, continued

- Use One-Against-All SVM to classify



K-Nearest Neighbours
(Video Google)



SVM
(Team 1)

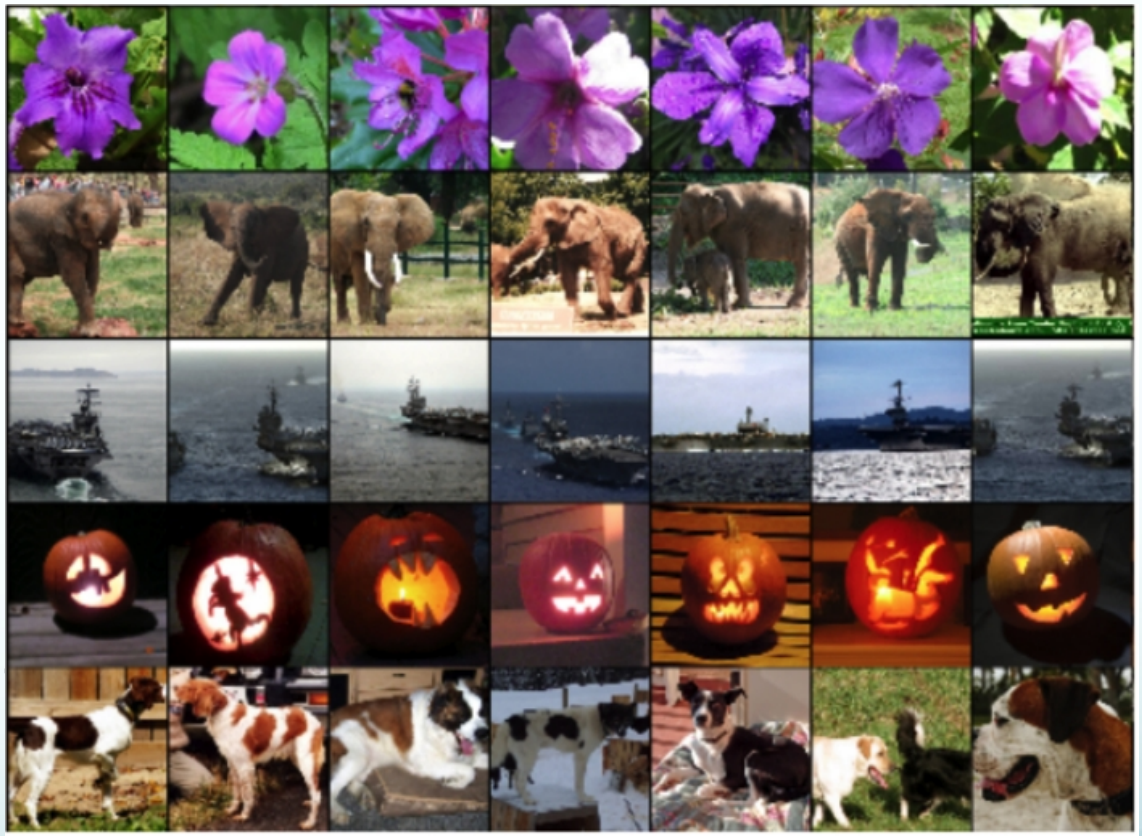
Method 2: CNN feature extraction

- OverFeat

  CNN =
- image classifier, feature extractor

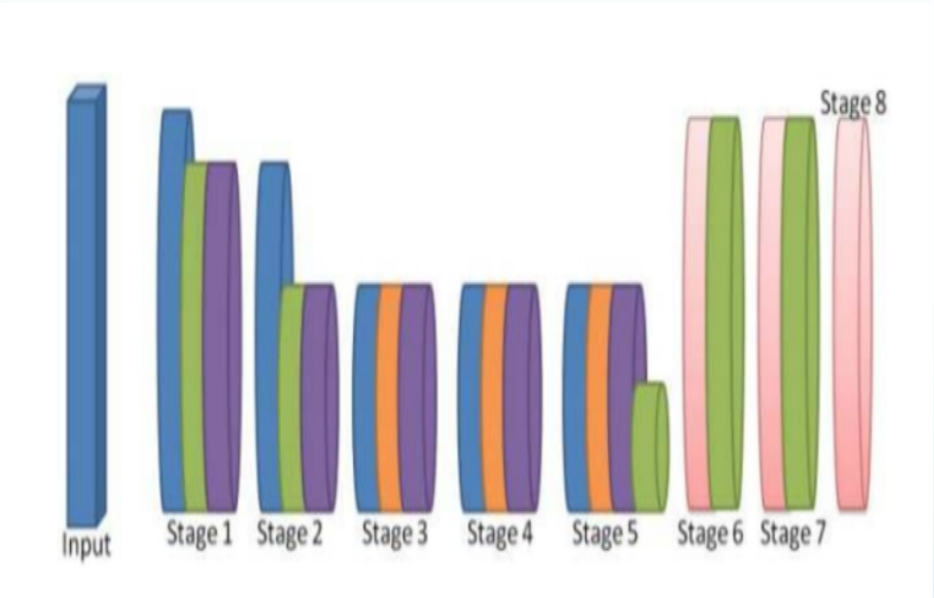
- Trained with the Torch7 package on
- the Imagenet datase

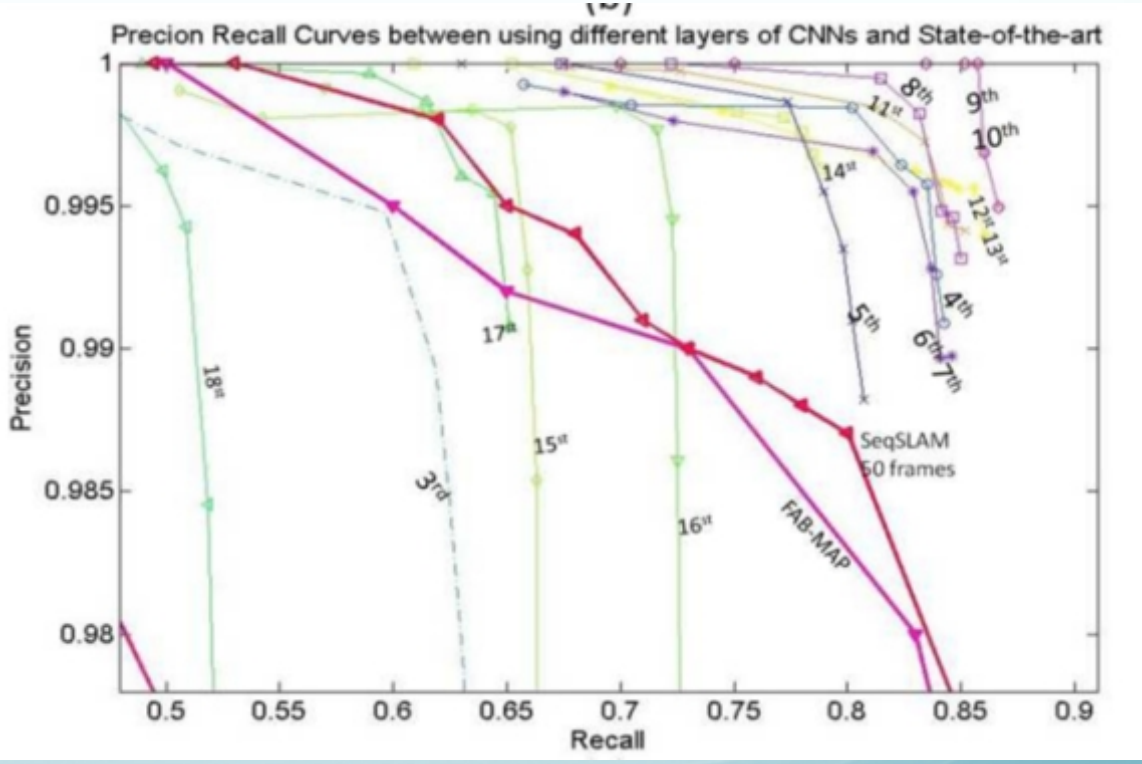
- Linux only...



CNN features, continued

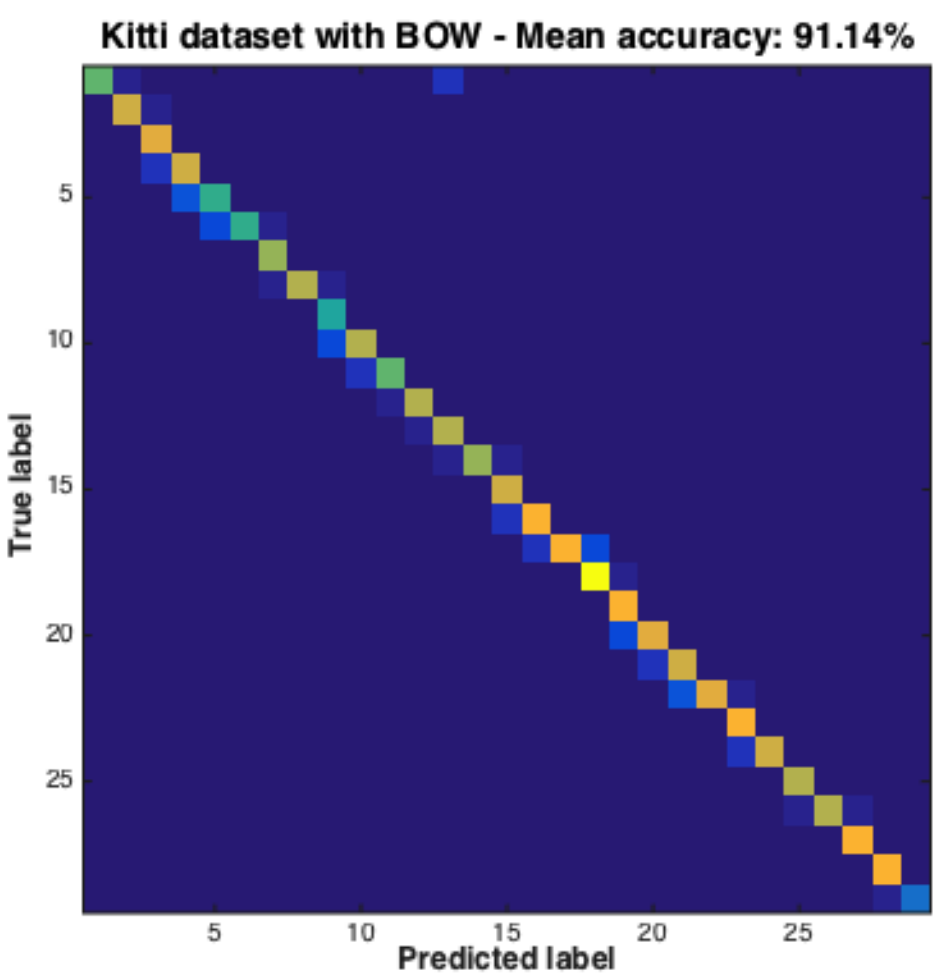
- Feature: vector describing “pictorial word”
- Extract features from layers 10 and 21

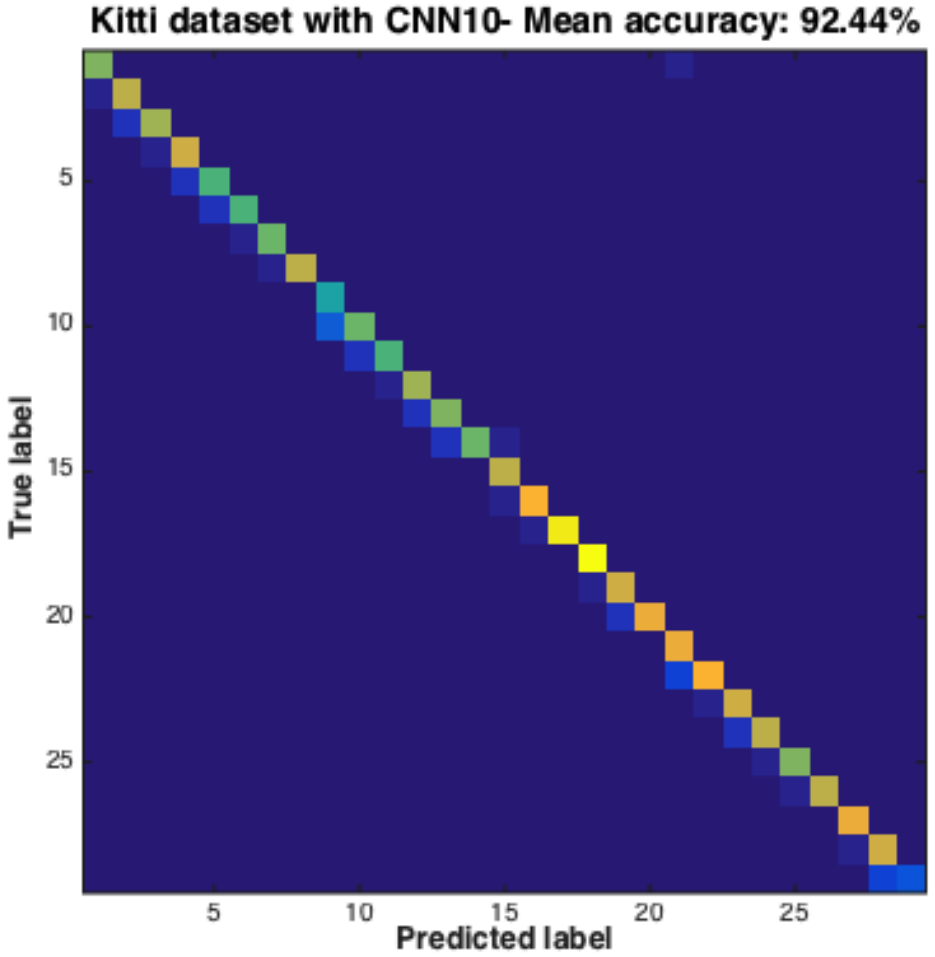




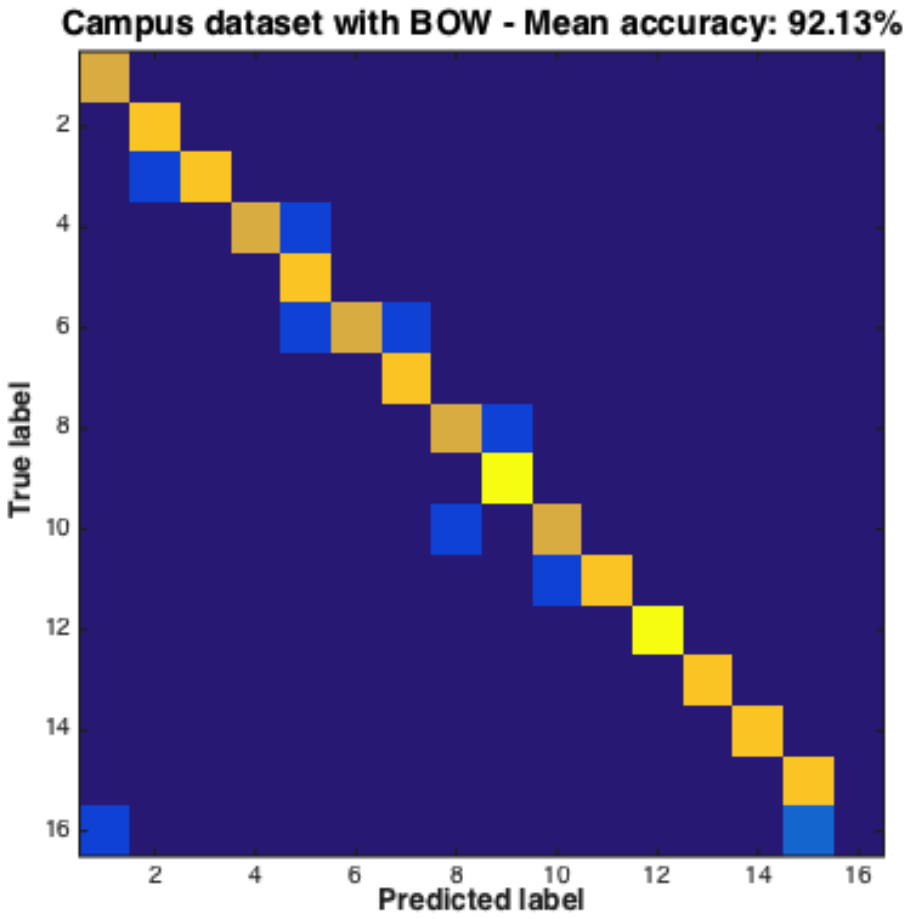
Convolutional Neural Network Based Place Recognition; Chen et al.

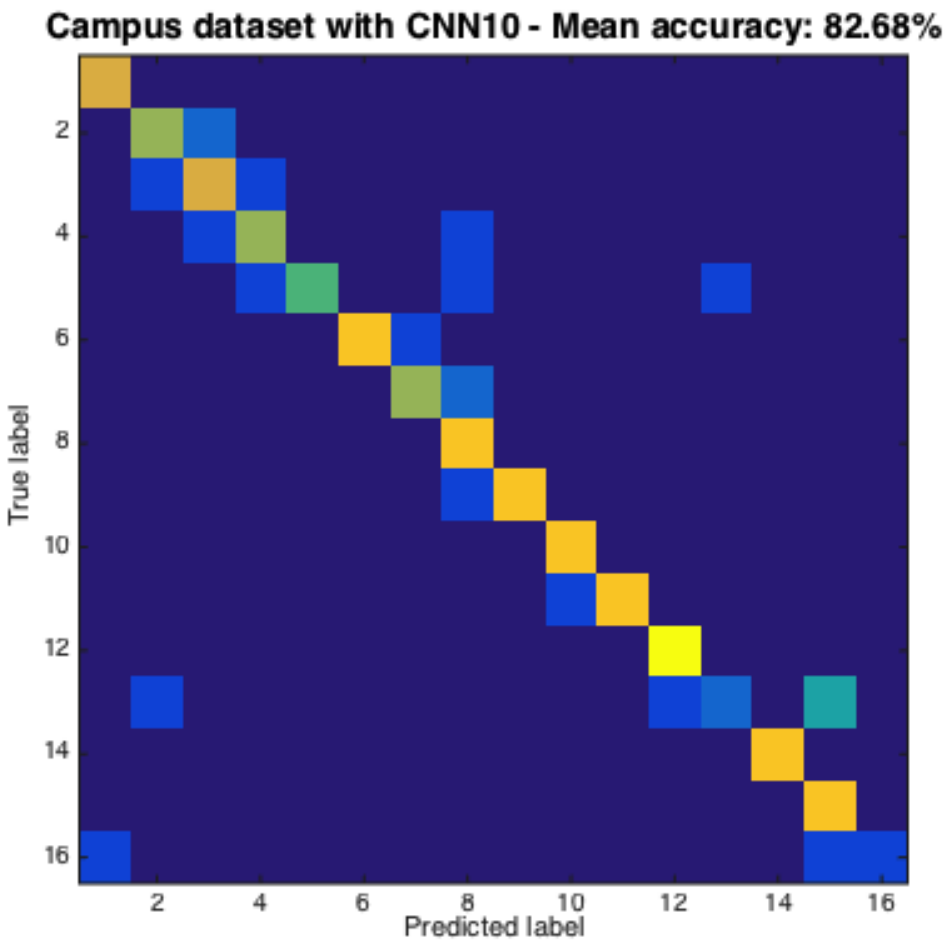
Confusion Matrices: Malaga Dataset





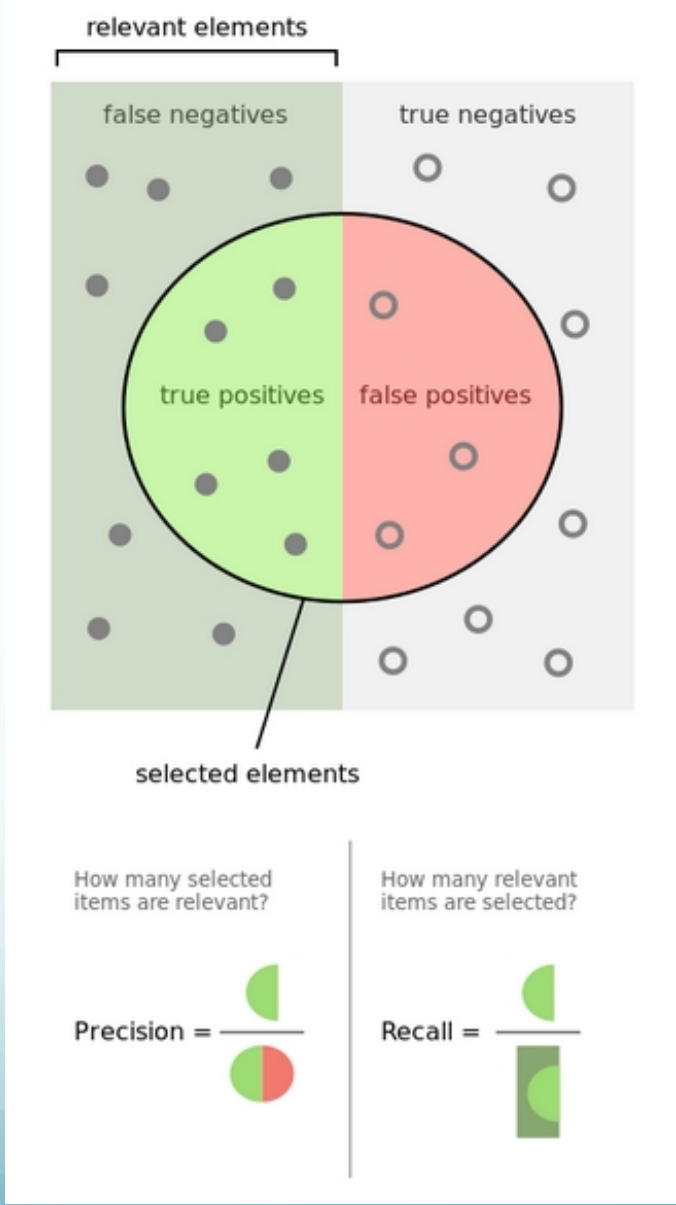
Confusion Matrices: **clbuucd** Dataset



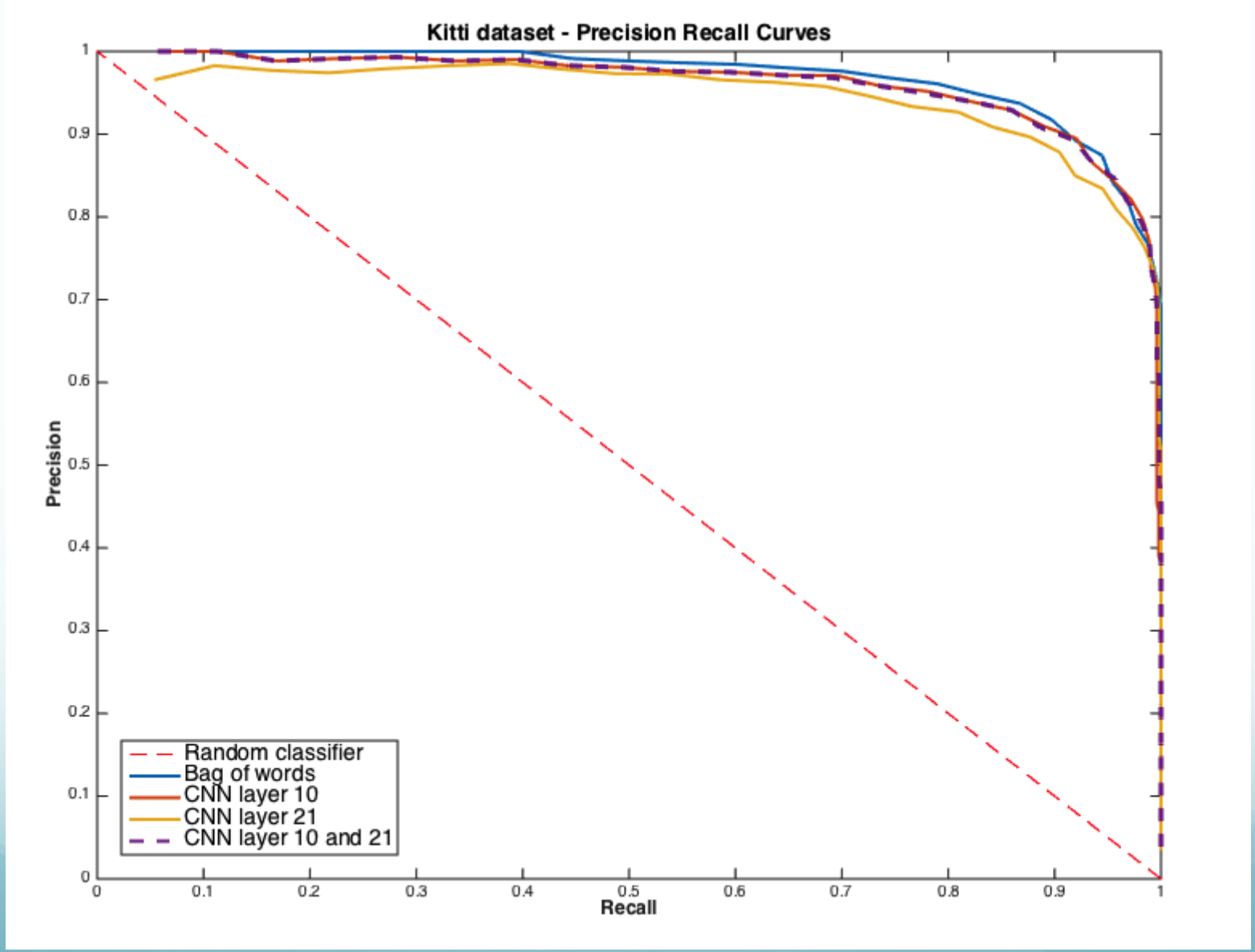


Precision-Recall Curve

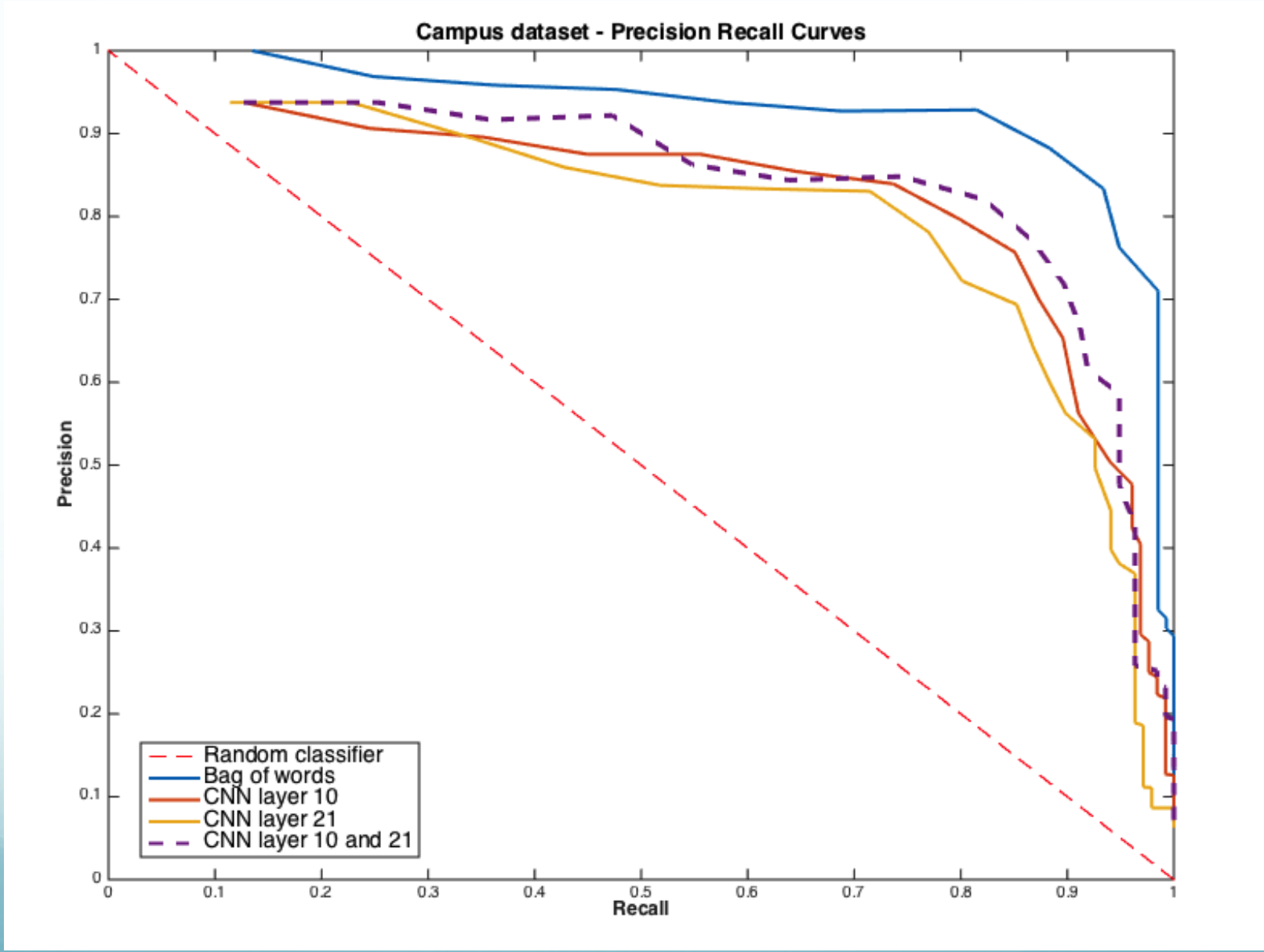
- How well does the model separate the classes?
- Recall =
  num relevant elements retrieved / num relevant elements overall
- Precision =
  num relevant elements retrieved / total # elements retrieved



Malaga Precision-Recall



cluubcd Precision-Recall



Neural network training experiment

- We made a pre-trained neural network work, how about trying to train a CNN on our own data?
- Caffe: deep learning framework out of UC Berkeley
- Imagenet framework on Malaga Dataset Extract #7
  - train on images from first loop
  - classify on images from second loop
- Results may likely improve with a larger dataset

http://caffe.berkeleyvision.org/

Results and Future Work

- Video Google-style BoW worked well on our small datasets
- BoW with OverFeat also worked well
- Caffe needs a little work
- Trying larger datasets?
- Different feature normalization methods?
- HoG and other kinds of descriptors?

Sources

Video Google:
robots.ox.ac.uk/~vgg/publications/papers/sivic03.pdf

Overfeat: github.com/sermanet/OverFeat

CAFFE: caffe.berkeleyvision.org/