Environment recognition from images
using more "traditional" ML techniques (updated)

Fundamentals of Data Science - Winter 2021/2022
Final Project

Robin Münk Marco Muscas Aizirek Doronbekova Sowmya Kommineni Lokanathbabu

The starting point | Dataset & Split

The dataset is from a research paper from Oliva & Torralba (2001)[1]. It contains photos of different environments.

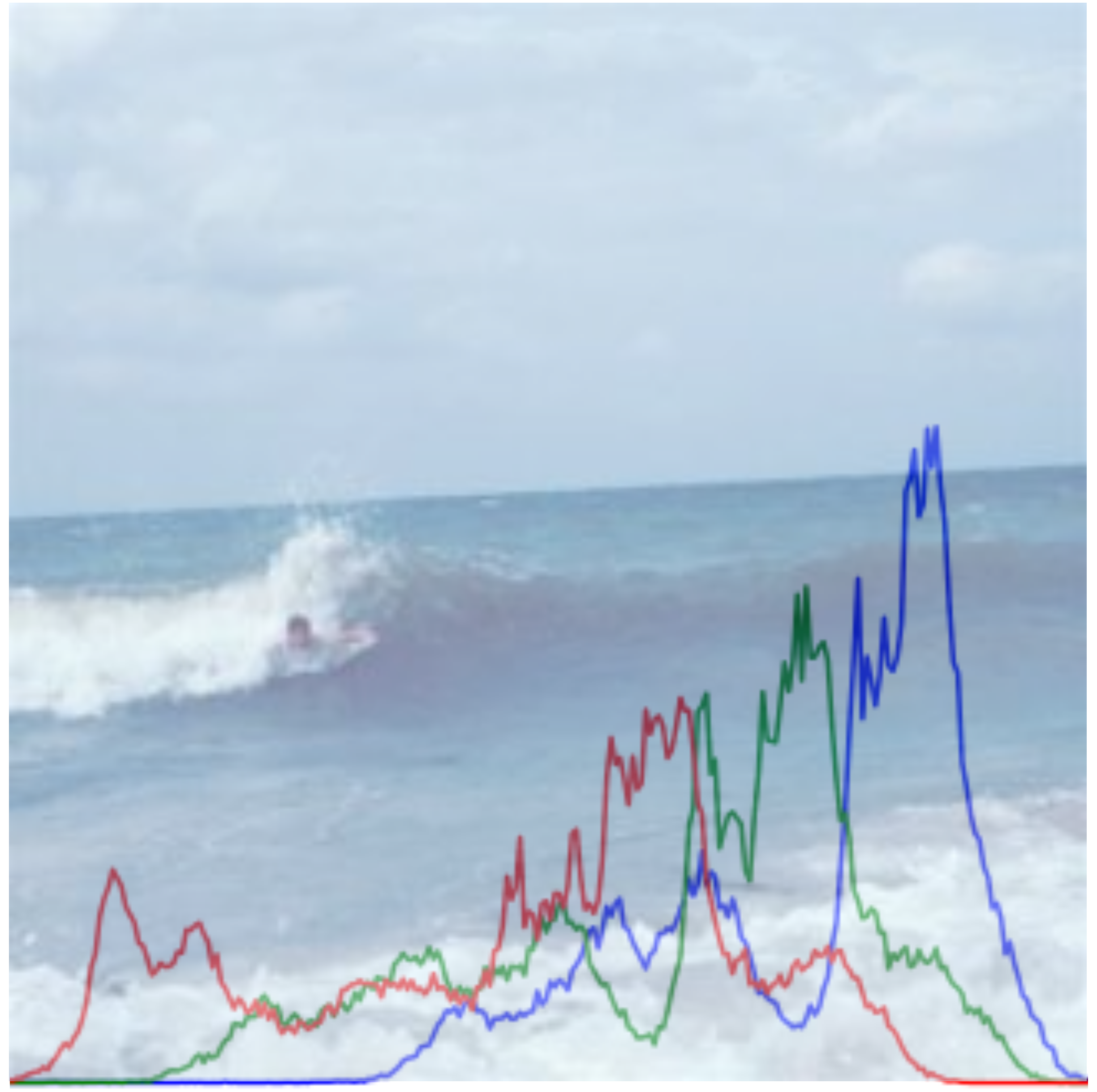
- **FORMAT**: Image path - Label pairs
- **SPLIT**: 64% Training / 16% Validation / 20% Testing
- It required some preprocessing to get the image-label pairs.
- Much of the information in the original dataset is discarded.
- **UNBALANCED:** we have different number of samples for each class

Feature Extraction | RGB & HSV Histograms

We should already know how this ends...

The idea here: each environment has its distinctive colors, right?

- Variable number of bins.
- As tested by a previous assignment, features are not in the colors usually.
- Quickly discarded for image classification.



RGB Histogram

Feature Extraction | ORB

Get the descriptors!

Again, the number of descriptors is variable. But ORB[3] is faster than SIFT!

- Included freely in the OpenCV build installed from pip.
- 32-dims instead of 128.
- (**spoiler alert**) - not as useful as SIFT

ORB Features

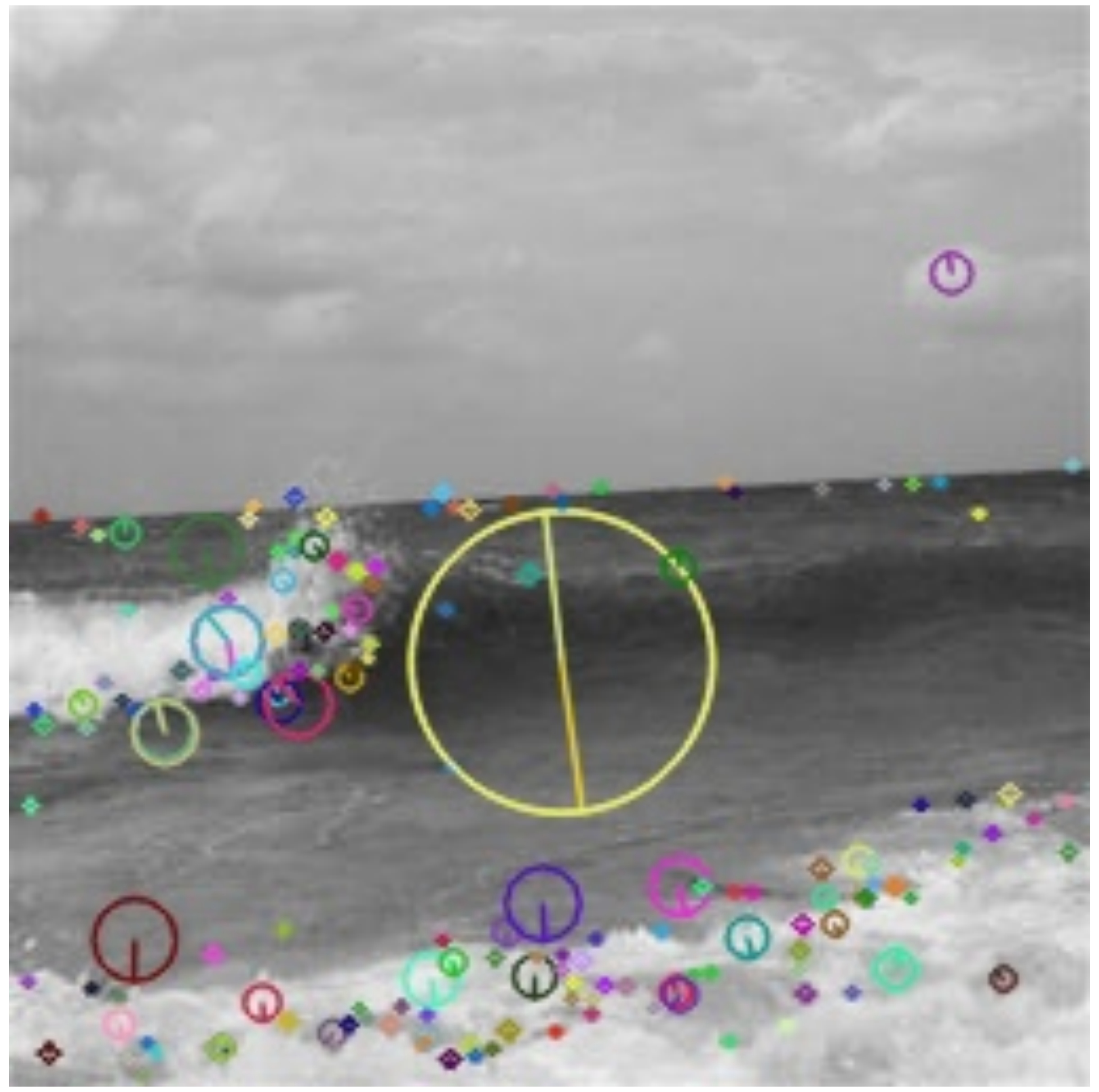
Feature Extraction | SIFT

Get more descriptors!

An image contains several SIFT[2] descriptors, we can imagine them as points of interest.

- The number of descriptors is always variable.
- Each descriptor is a 128-dim vector.
- Slower than the improved counterpart: **SURF**.

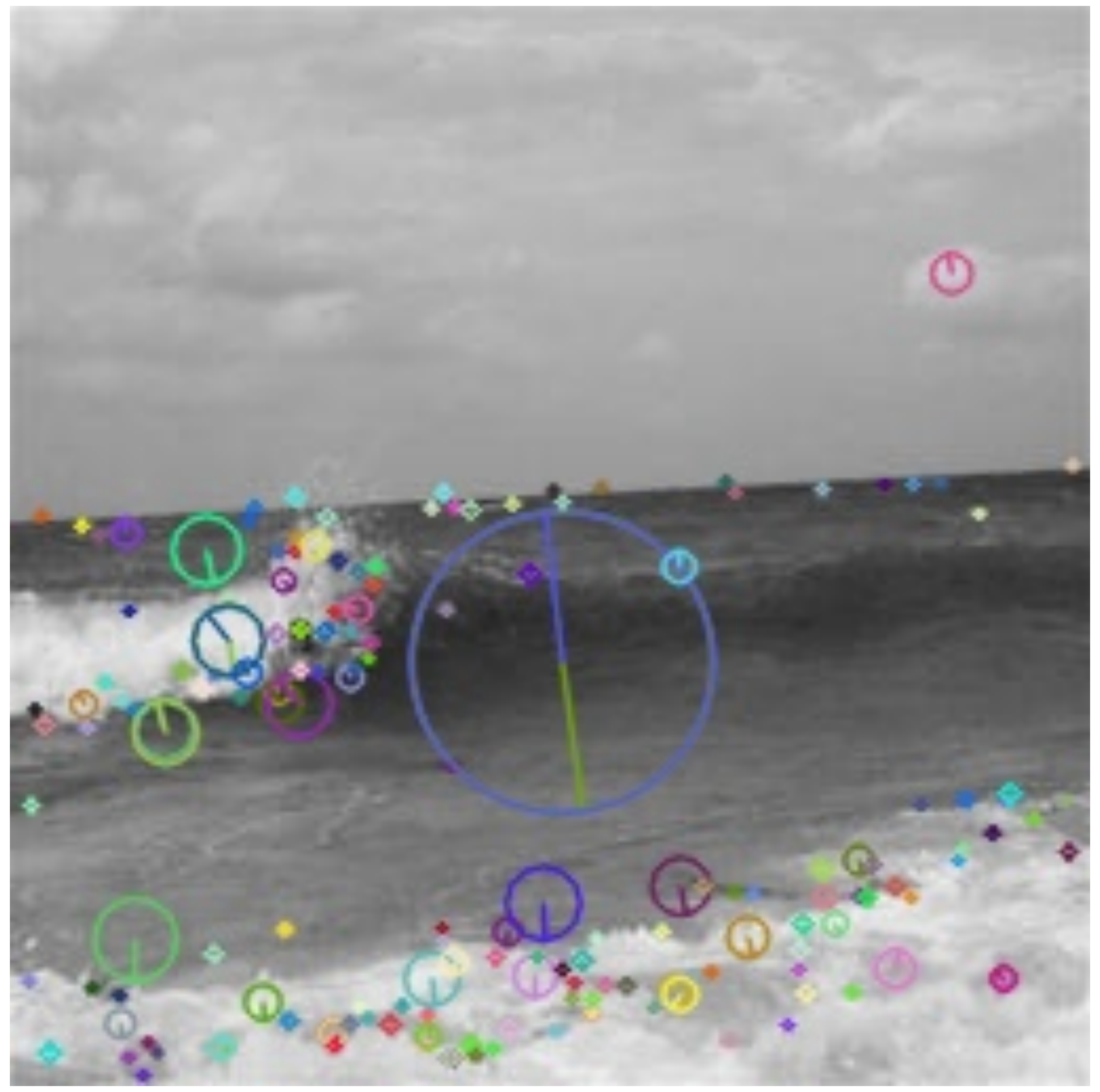
(SURF is patented and not available in the free build of opencv)



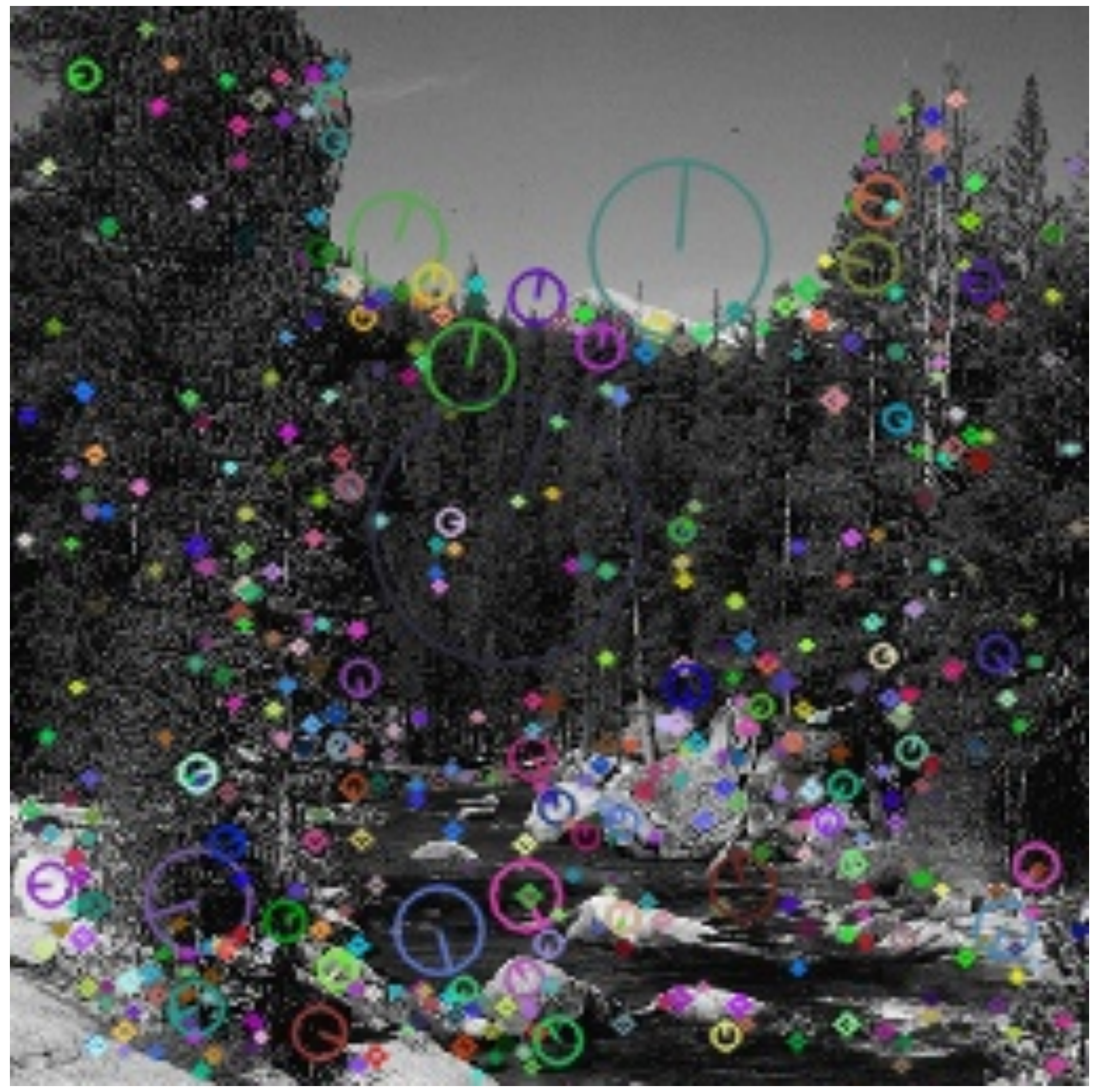
SIFT Features

There's a problem!

- The number of descriptors for each image is different from the rest!



181 descriptors



543 descriptors

997 descriptors

Bag of Visual Words!

Feature mapping | KMeans

How to get the bag of visual words.

A bag of visual words is a one-hot vector indicating the presence of visual "**terms**" (features).

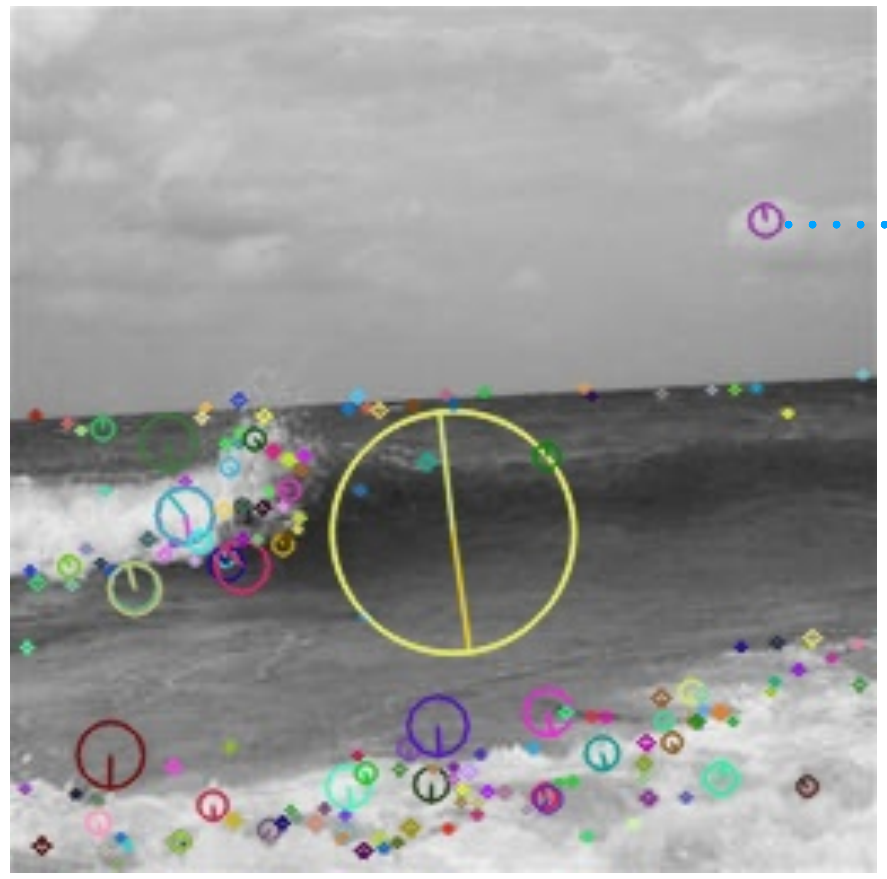
1. Decide how many visual words to report (i.e. number of clusters).
2. Train a KMeans classifier on those visual terms - the feature descriptors.
3. Now you are able to associate a visual feature to a cluster number.

The visual descriptors are nothing but vectors after all

Feature mapping | KMeans

How to get the bag of visual words.

We will see if an image has a certain visual feature, and "tick" its box in a one-hot vector!



181 features

kmeans.predict(descriptor)

0	0	1	0	1	...
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	

Bag of Visual Words

Side note: we actually used MiniBatch-KMeans due to memory constraints.

Prediction | Logistic Reg., SVM and Ridge

- **SVM**

  Find an n-dimensional hyperplane to separate the points into k classes.

  Validation: 70% | Test: 70%

- **RIDGE REGRESSION**

  A linear regression using regularization.

  Validation: 39% | Test: 40%

- **LOGISTIC REGRESSION**

  Logistic model we started with.

  Validation: 69% | Test: 63%

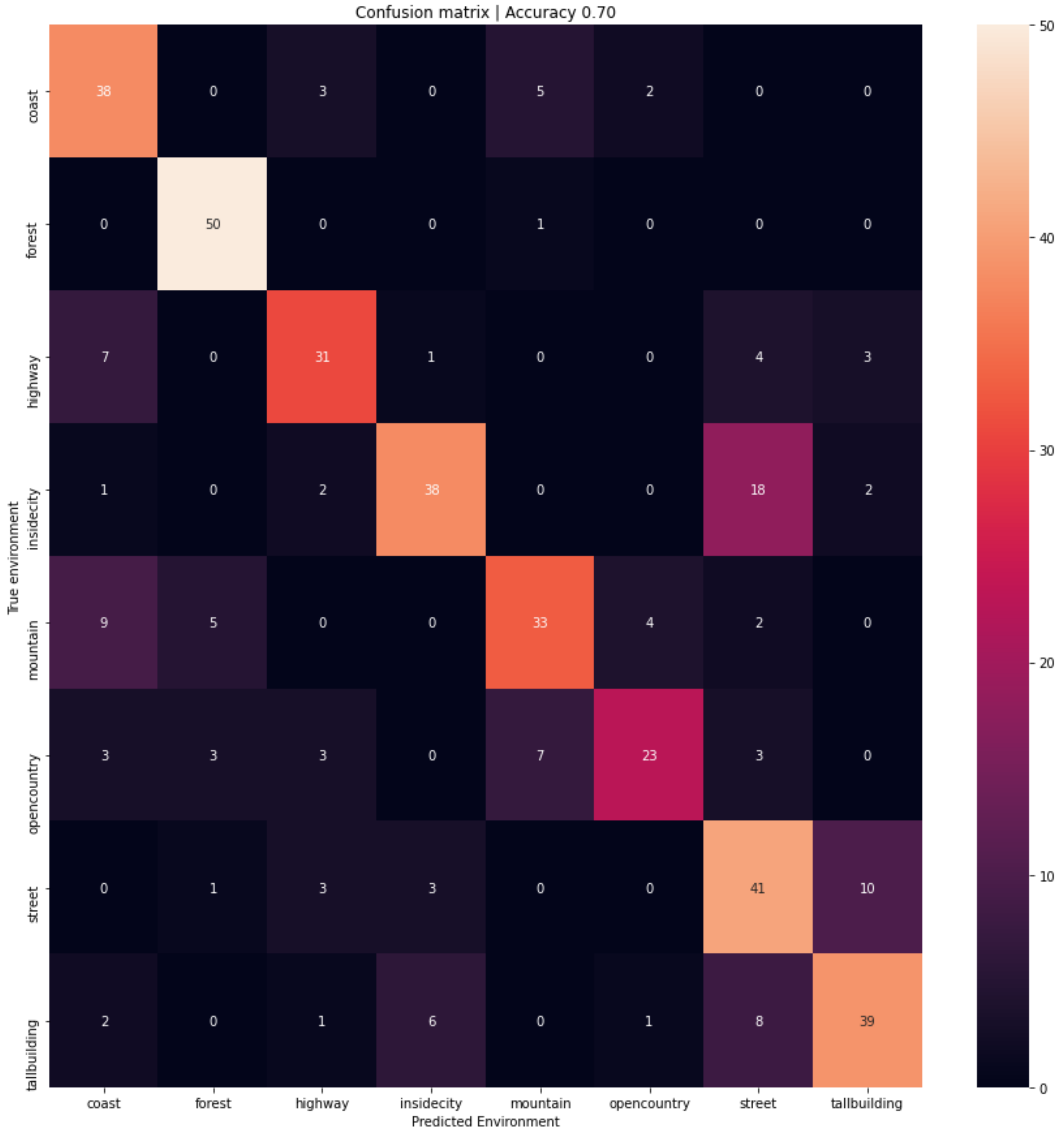
Evaluation | What do we have?

Standard metrics and comparison

- Best test accuracy: 70% using SVM
- Suggested number of features: *n = 2000*
- Batch size for Minibatch-KMeans: *c* > 512
- 260 samples for each class in training

Observations:

- **Some areas are easily confused: highway and street**
- **Execution time is approximately the same (a few minutes)**
- **The number of features influences heavily the accuracy**
- **Training split could go as low as 30% with comparable results to the rest.**



What we could try

A few options...

- Thresholding by number of visual features in the BoVW?
  - Not all features could be relevant, therefore "confusing" the cluster making.

- Data augmentation?
  - We need to be careful, SIFT is scale and rotation invariant...

- More hyperparameter tuning!
  - We can't go wrong with this.

Thank you for your attention!

Bibliography

1. Aude Oliva, Antonio Torralba | Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int. J. Comput. Vis. 42(3): 145-175 (2001)
2. David G. Lowe: Distinctive Image Features from Scale-Invariant Keypoints. Int. J. Comput. Vis. 60(2): 91-110 (2004)
3. Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary R. Bradski | ORB: An efficient alternative to SIFT or SURF. ICCV 2011: 2564-2571