Master in Computer Vision Barcelona

Project: A Hands-On Experience on Visual Object Recognition

Module 5: Visual Recognition

Coordination: Ramon Baldrich

ramon.baldrich@uab.cat

Master in Computer Vision Barcelona

UAB UOC UPC upf.

Project Goal

Goals:

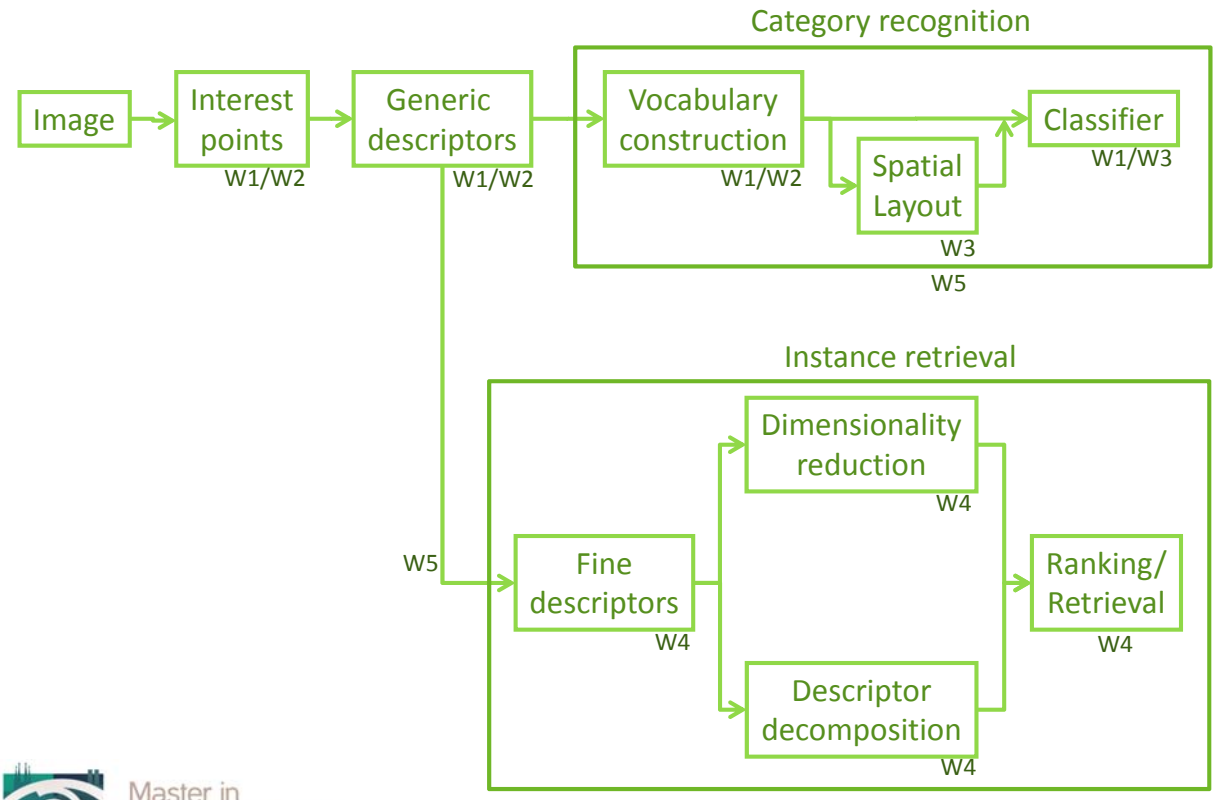
- Estimate the content of images:
  - Category classification
  - Instance Retrieval
- The final project will have to produce labels for images, specifying which is the object in the image and which is the most similar images to a given query.

Data:

- subset of the large hand-labeled Urban and Natural Scene Categories
- subset of the large hand-labeled ImageNet

Master in Computer Vision Barcelona

Project Flowchart



Master in Computer Vision Barcelona

Project Schedule

Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
•BOW •K-Means, •SVM •SVM kernels •LBP	•Keypoints •Early fusión •PCA •Color •SURF/SC •Late Fusion	•Spatial Pyr. •Intermediate Fusion •Multiple SP •Intermediate Fusion	•Fisher descriptor • Inverted file • VLAD • Fast NN • Product Quantization	• Features from CNN • CNN fine tunning	**•Presentation**

Master in Computer Vision Barcelona

Lectures

ASSIGNMENT	LECTURES
Basic pipeline for BOW	The Bag of Words framework
	Experimental setup
Fusing and improving descriptors	Image feature / Color Image
	Augmenting the classification accuracy
Including spatial information	Augmenting the classification accuracy
From recognition to retrieval	Large scale image retrieval
Features from ConvNet (category clasification)	ConvNet for global recognition
ConvNet fine tunning (Image Retrieval)	ConvNet for local recognition

Master in
Computer Vision
Barcelona

Assignments

- Week 1

  Basic pipeline for bag of words

  The goal of this week is (i) to understand the basic principles underlying the pipeline of the Bag of Words approach and (ii) tune the parameters of the vocabulary construction when applied in a subset of a MIT scene database. Subsequently, a more complex classifier based on Support Vector Machine will be trained to learn to discriminate between different object categories..

  - Mandatory tasks: Consider different vocabulary sizes; Substitute the nearest neighbor by a linear SVM.
  - Optional tasks: Experiment with different features like HOG, LBP; Extend the linear SVM to consider kernels, like Chi square or RBF
  - Performance evaluation: Acc., confusion matrix, ROC curves
  - Deliverable: code, 3 slides presentation

Master in
Computer Vision
Barcelona

Assignments

- Week 2

  Fusing and improving descriptors

  The goal of this week is to explore the pros/cons of using key-points selection strategies, to reduce the amount of information in the construction of the code book and to test how to fuse the information obtained by the descriptors tested in Week 1.

  - Mandatory tasks: extend the number of different detectors/descriptors to 'explain' images, reduce the amount of information with PCA and apply a fusion methodology to combine descriptors.
  - Optional tasks: is color information enough to distinguish scenes? consider Opponent SIFT and Color Naming techniques; Late fusion methodology
  - Performance evaluation: Acc., confusion matrix, ROC curves
  - Deliverable: code, 3 slides presentation

Master in Computer Vision Barcelona

Assignments

- Week 3

  Including spatial information

  The goal of this week is to understand the concept of Spatial Pyramids by applying image descriptors at different location and resolution scales. Also, another aim is to implement the late fusion strategy (concatenation of classifiers) for combining different image descriptors

  - Mandatory tasks: Evaluate the performance of the Spatial Pyramid schema. Compare Early Fusion vs. Late Fusion.
  - Optional tasks: explore different SP setups; apply a soft enconding (GMM) architecture.
  - Performance evaluation: Acc., confusion matrix, ROC curves
  - Deliverable: code, 3 slides presentation

Master in Computer Vision Barcelona

Assignments

- Week 4

  From recognition to retrieval

  The goal of this week is to begin working with instance recognition instead of category classification, for the ImageNet dataset. The first part will be devoted to use the Fisher vectors. The second part will be focused on implementing the inverted file technique for performing efficient retrieval.

  - Mandatory tasks: Test the Fisher vector descriptor in a soft-assignment ranking approach; include the inverted file technique.
  - Optional tasks: use the VLAD descriptor; implement the fast/approximate NN.
  - Performance evaluation: Accuracy Top 1, Top 5, Top 10
  - Deliverable: code, 3 slides presentation

Master in
Computer Vision
Barcelona

Assignments

- Week 5

  Recognition and retrieval with convolutional networs

  The goal of this week is two fold: on one hand, to see how features can be extracted from a CNN architecture and applied to Image retrieval. On the other the goal is to explore the usage of existing learned CNN and adapt them to different task to the ones they were thought for (fine tunning).

  - Mandatory tasks: Use an exsiting architecture, analyze which layer is the best for Image Retrieval. Using AlexNet, adapt it to the problems is weeks 1,2 and 3..
  - Optional tasks: TBA
  - Performance evaluation: confusion matrix. Accuracy Top 1, Top 5, Top 10
  - Deliverable: code, 3 slides presentation

Master in
Computer Vision
Barcelona

Assignments

- Week 6
  FULL PROJECT PRESENTATION

Master in
Computer Vision
Barcelona

Programming Language

- Python python.org
- Required libraries (from linux repository)
  - python-opencv
  - Python-sklearn
  - python-imaging
  - python-matplotlib
  - python-numpy
  - python-scipy
  - python-skimage

Master in
Computer Vision
Barcelona

Datasets

Scene dataset (MIT http://cvcl.mit.edu/database.htm)

- 8 classes selected: coast, forest, highway, inside-city, mountain, open-country, street, tall-building

TRAINING/VALIDATION DATASET

- ~200 images x class

TEST DATASET

- ~100 images x class

Master in Computer Vision Barcelona

Datasets

- ImageNet (large scale)
  - 2 classes: cars, dog

  - TRAINING / VALIDATION DATASET (subset)
    - 3500 images per class
  - TEST DATASET (subset)
    - 1500 images per class

Master in Computer Vision Barcelona

System Performance Evaluation

- Image classification (Scene dataset)
  - Accuracy
  - Confusion Matrix
  - ROC curve
- Image retrieval (ImageNet)
  - Top 1, Top 5, Top 10

Master in
Computer Vision
Barcelona

Project Evaluation

- The Project Development is evaluated: PD
  - Each week the work is evaluated: V_{CD}
  - The work of the final system is evaluated: V_{CDfull}

$$PD = \frac{0.8}{5}\sum_{i=1}^{5} V_{Cd_i} + 0.2 V_{Cd_{full}}$$

- The final project presentation is evaluated: PP
- Intra-Group Evaluation: IGE
- The final mark is $V = 0.7 \cdot PD + 0.1 \cdot PP + 0.2 \cdot IGE$

Master in
Computer Vision
Barcelona

Master in Computer Vision Barcelona

Project — A Hands-On Experience on Visual Object Recognition

Module 5 — Visual Recognition

Coordination — Ramon Baldrich

Master in Computer Vision Barcelona

UAB UOC UPC upf.