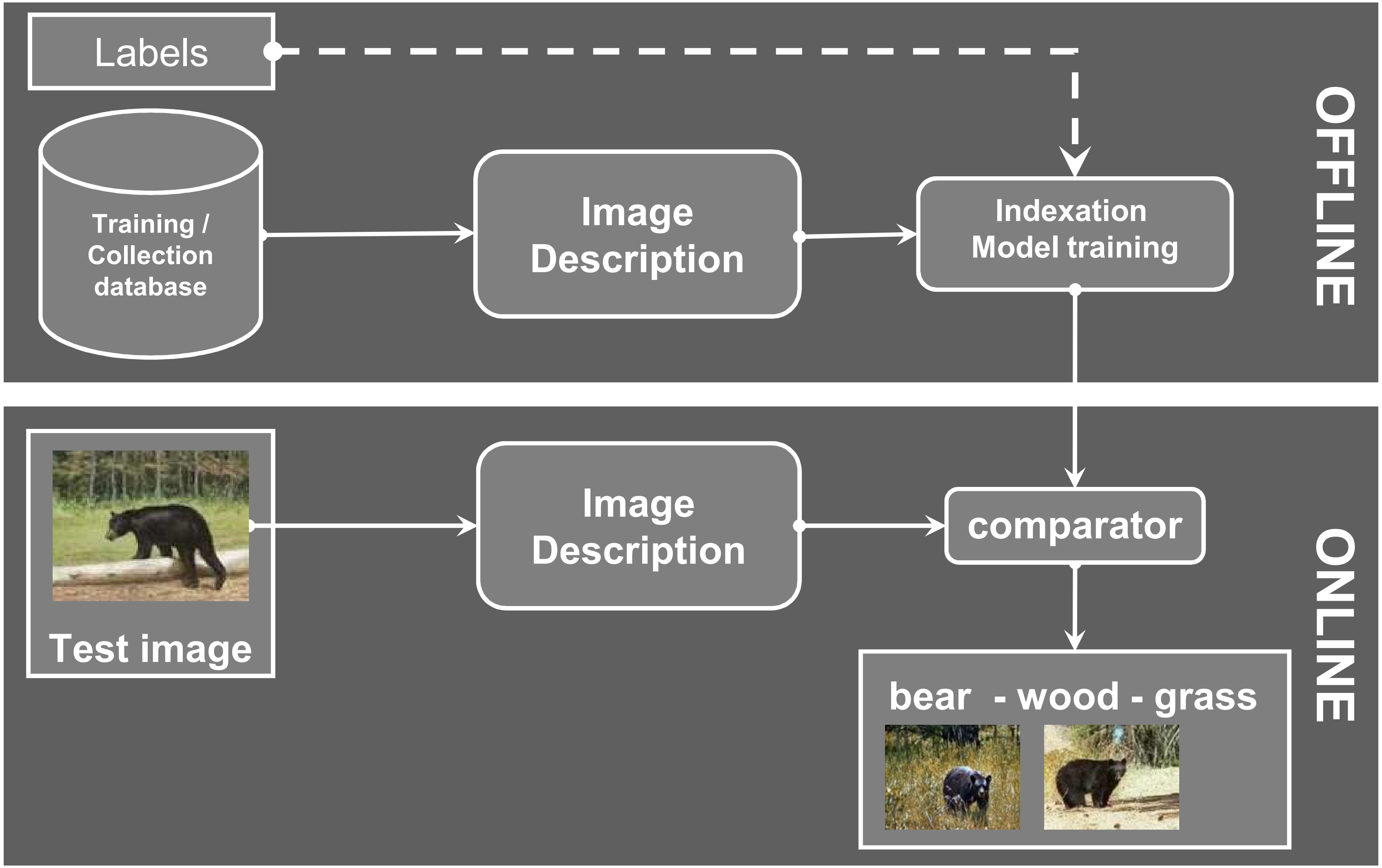
Constrained Local Enhancement of Semantic Features by Content-Based Sparsity

Youssef Tamaazousti, Hervé Le Borgne and Adrian Popescu

youssef.tamaazousti@cea.fr

Classification & Retrieval are linked



Classification & Retrieval are linked

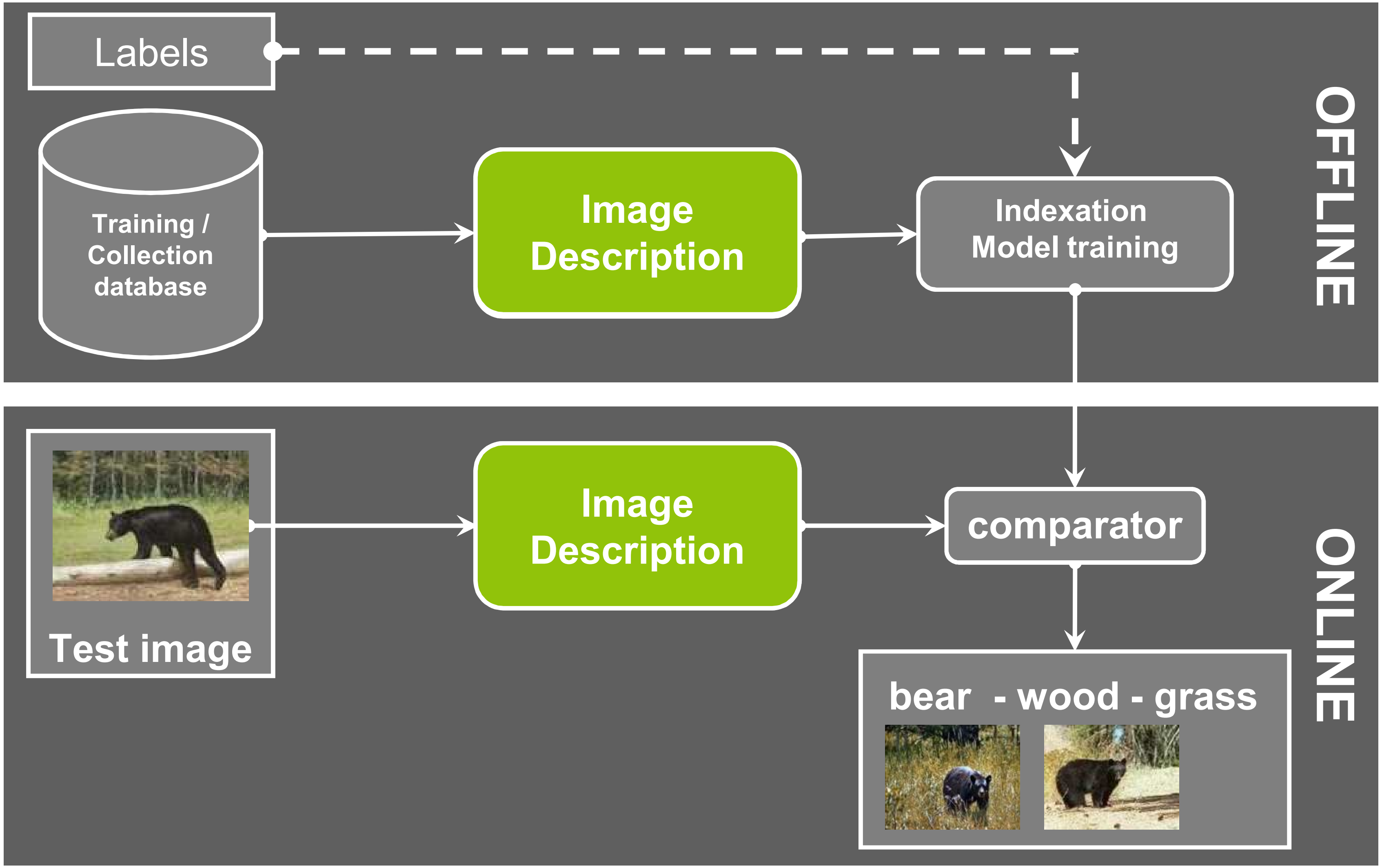
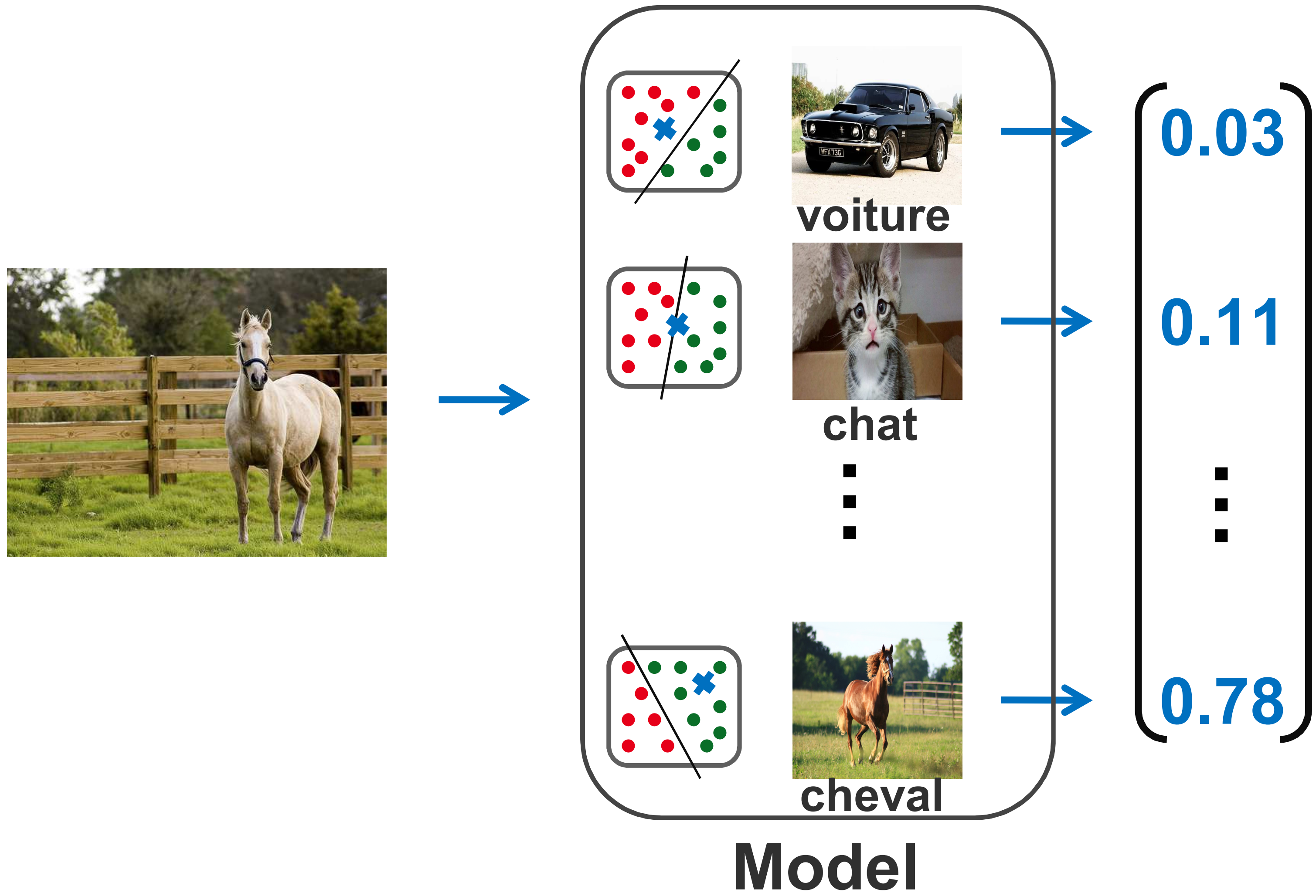


Image Description

- **Low/Mid-Level Features**
  - Bag of Visual Words [Csurska *et al.* 2004]
  - Fischer Kernels [Perronnin & Dance 2007]
  - Fully-connected layers of a pre-trained CNN [Kriszhevshy *et al.* 2012]

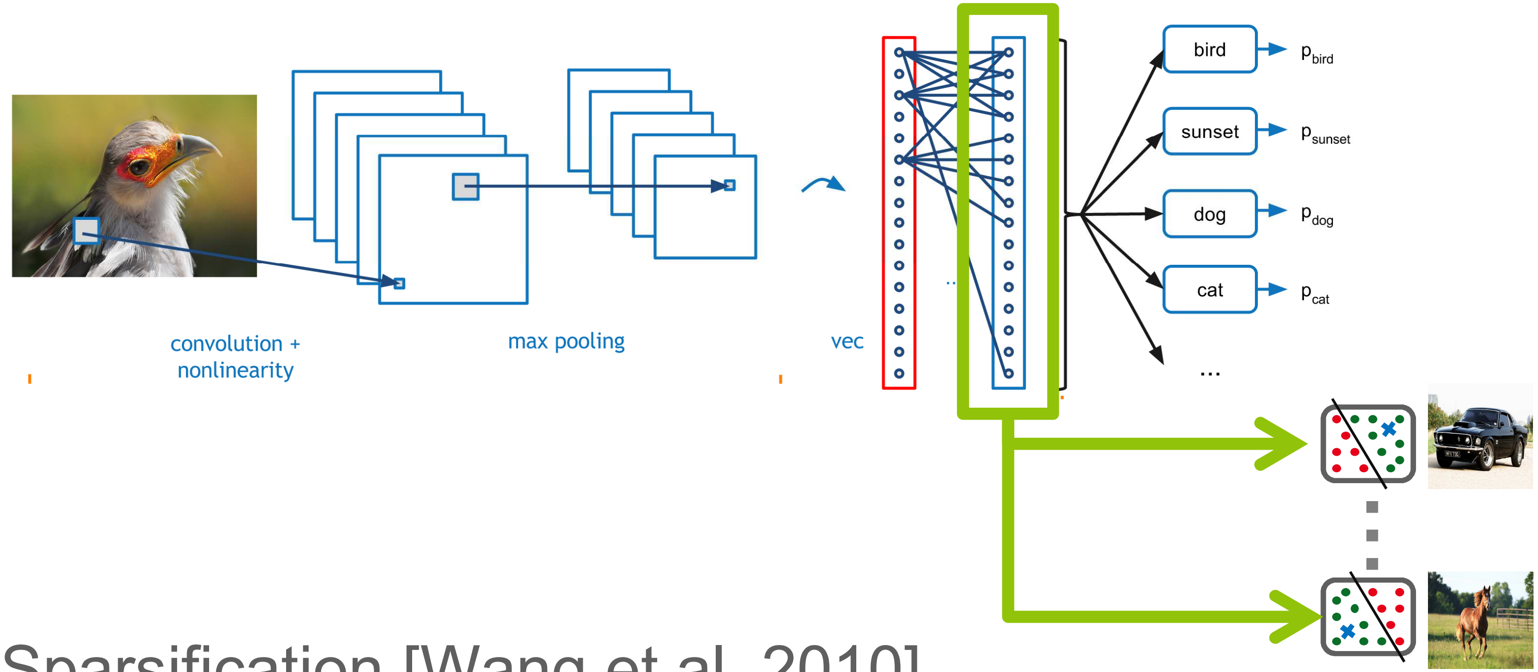
- **Semantic Features**
  - Image described in terms of semantic concepts
  - [Torresani *et al.*, Li *et al.* 2010] (200 - 1500)
  - [Bergamo *et al.* 2012] (15000)

What is a Semantic Feature ?



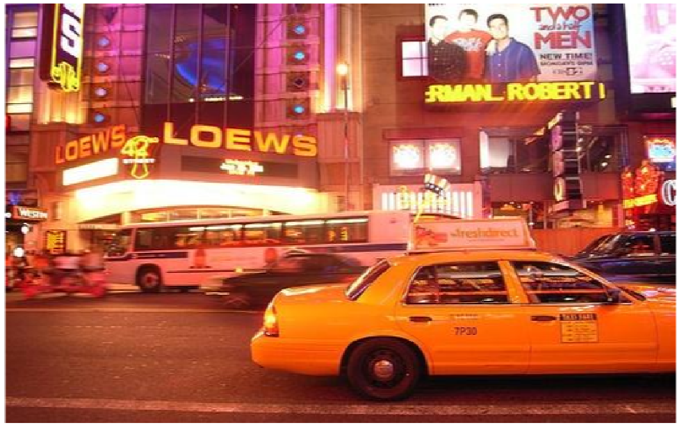
State-of-the-Art

- **Ginsca *et al.* 2015**
  - First to use CNNs as mid-level features

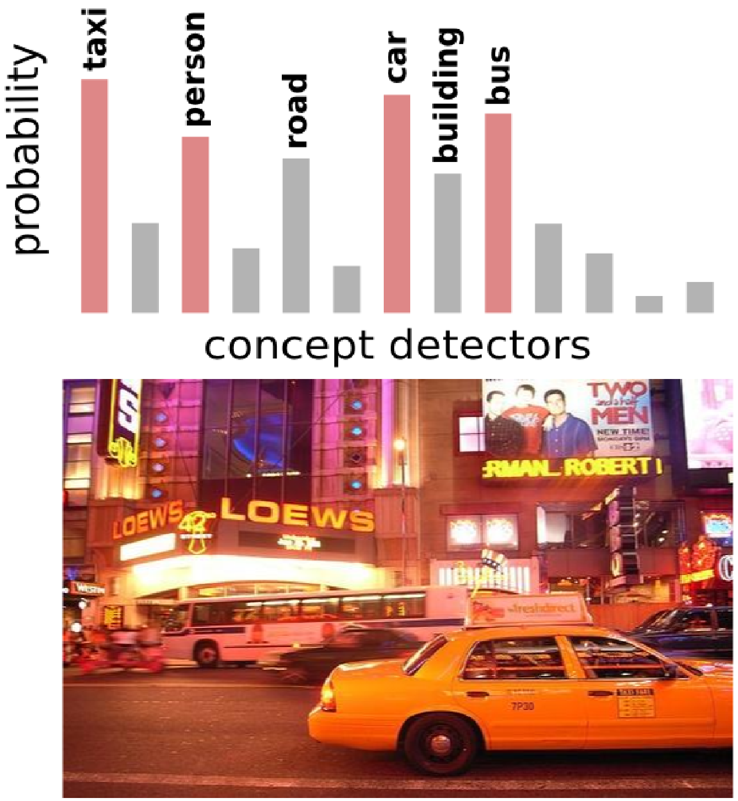


  - Sparsification [Wang et al. 2010]
    - keep only the K largest concepts and set all others to 0
    - Inverse index scheme can be used

State-of-the-Art

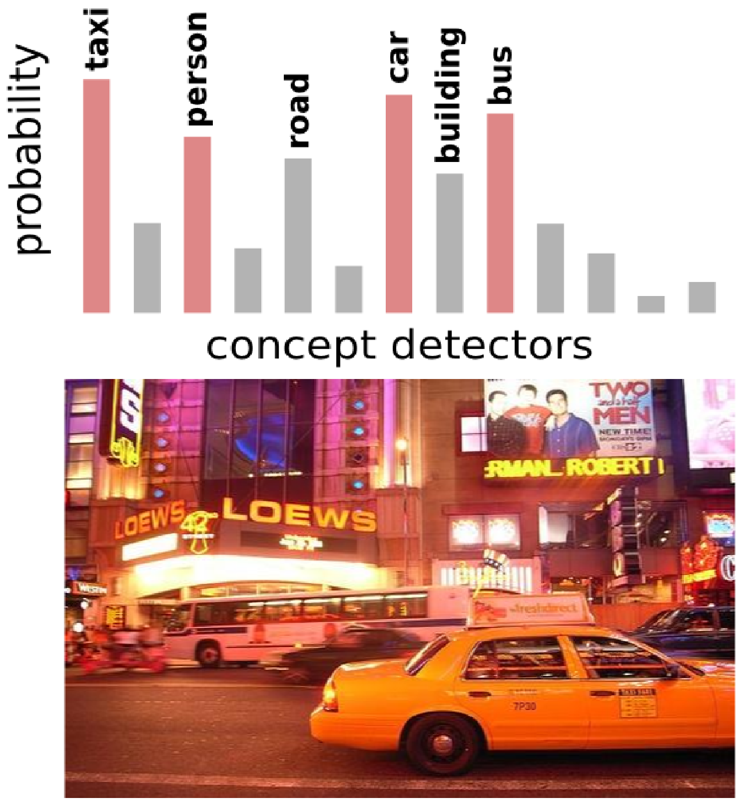


State-of-the-Art



State-of-the-Art

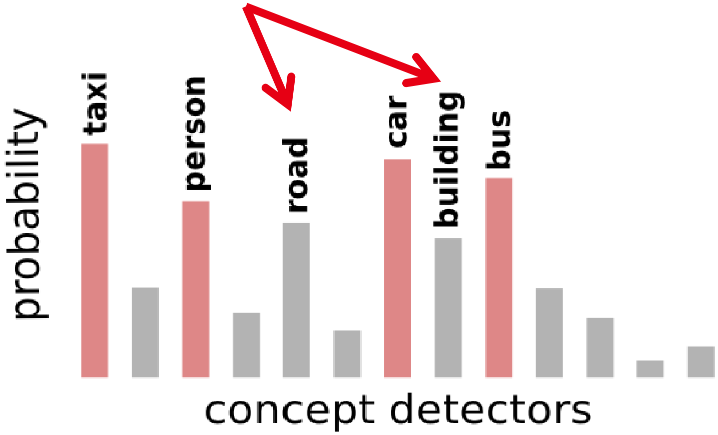
K = 4

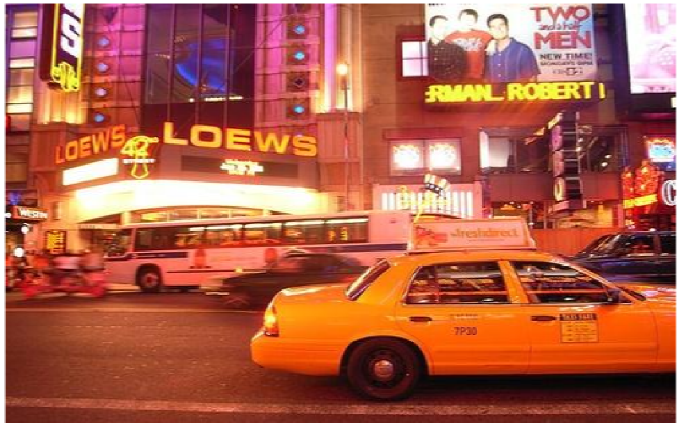


State-of-the-Art

K = 4

Missing relevant information

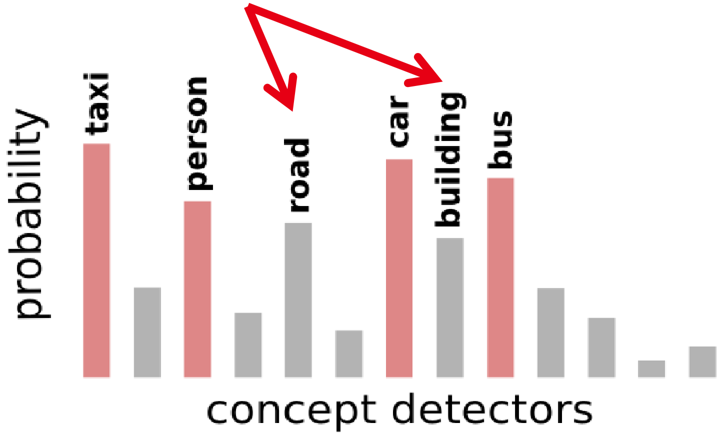


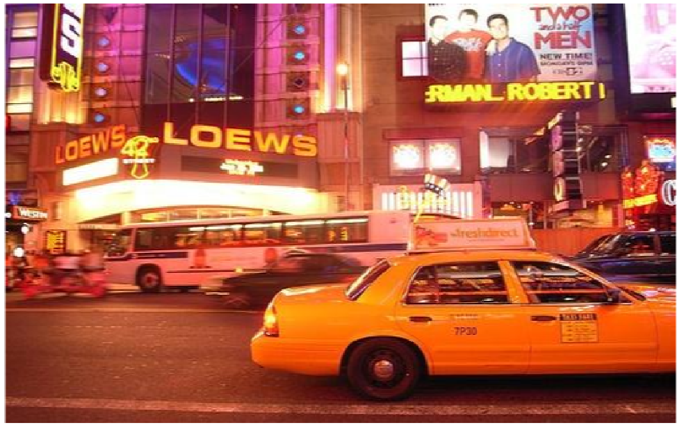


State-of-the-Art

K = 4

Missing relevant information

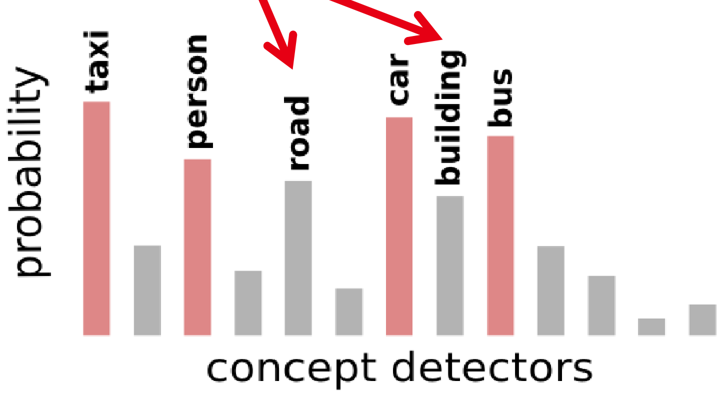


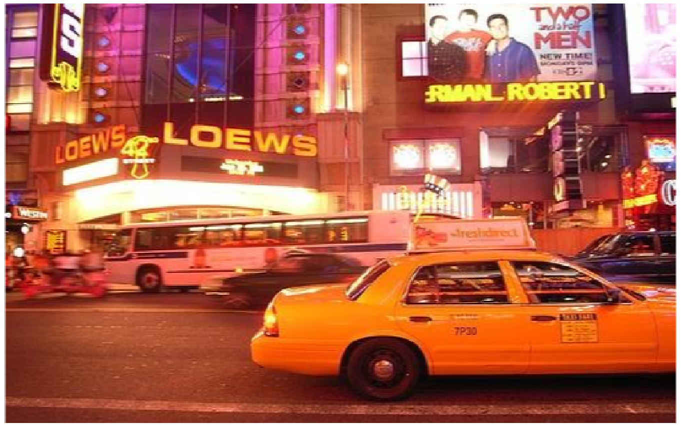


State-of-the-Art

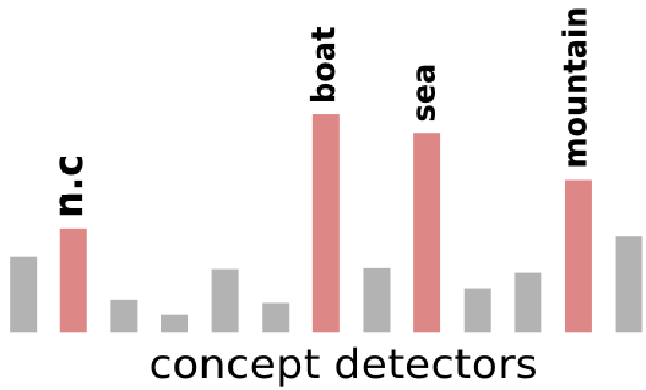
K = 4

Missing relevant information





K = 4



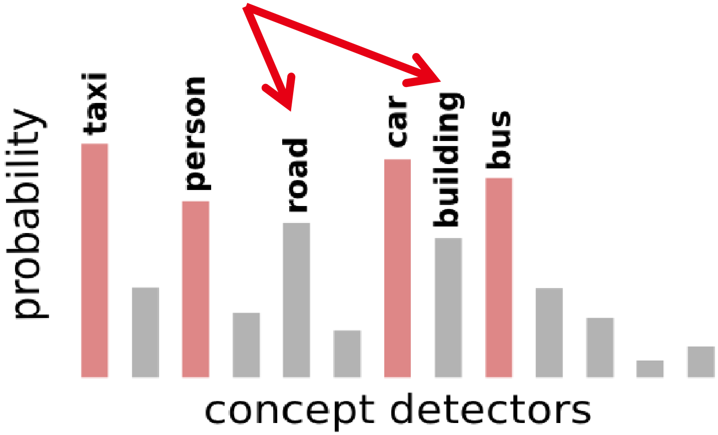
State-of-the-Art

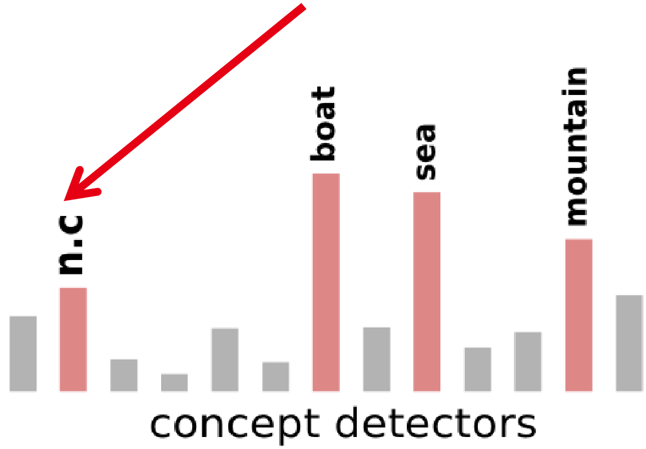
K = 4

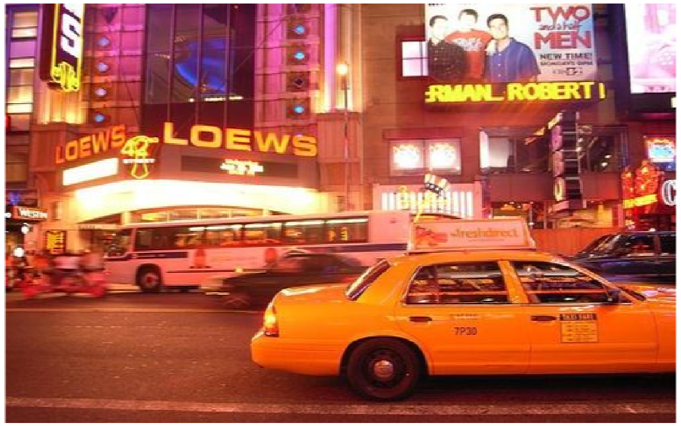
Missing relevant information

K = 4

Keep noisy information

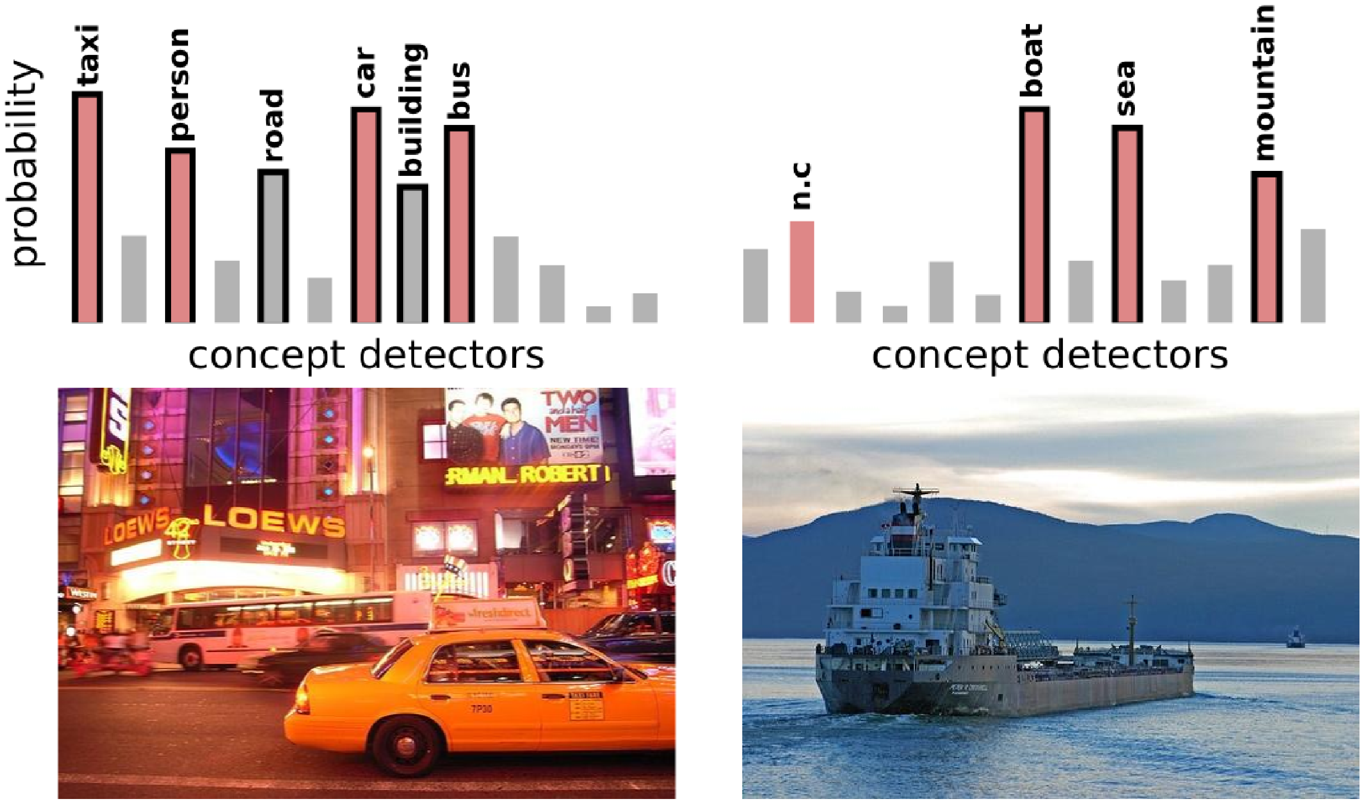






n.c = noisy concept

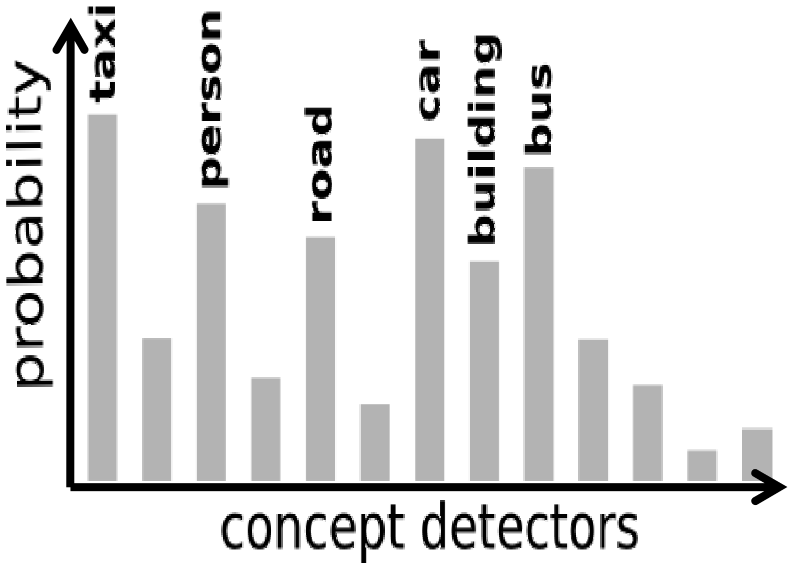
Proposed Approach: CBS



- **How many concepts should we retain for each image?**

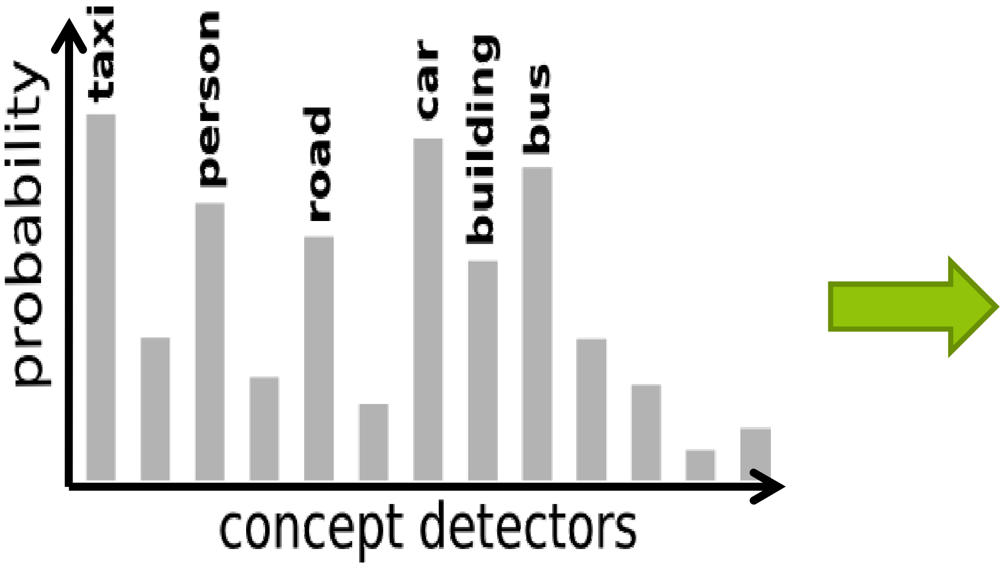
Proposed Approach: CBS

- **Sparsification process**



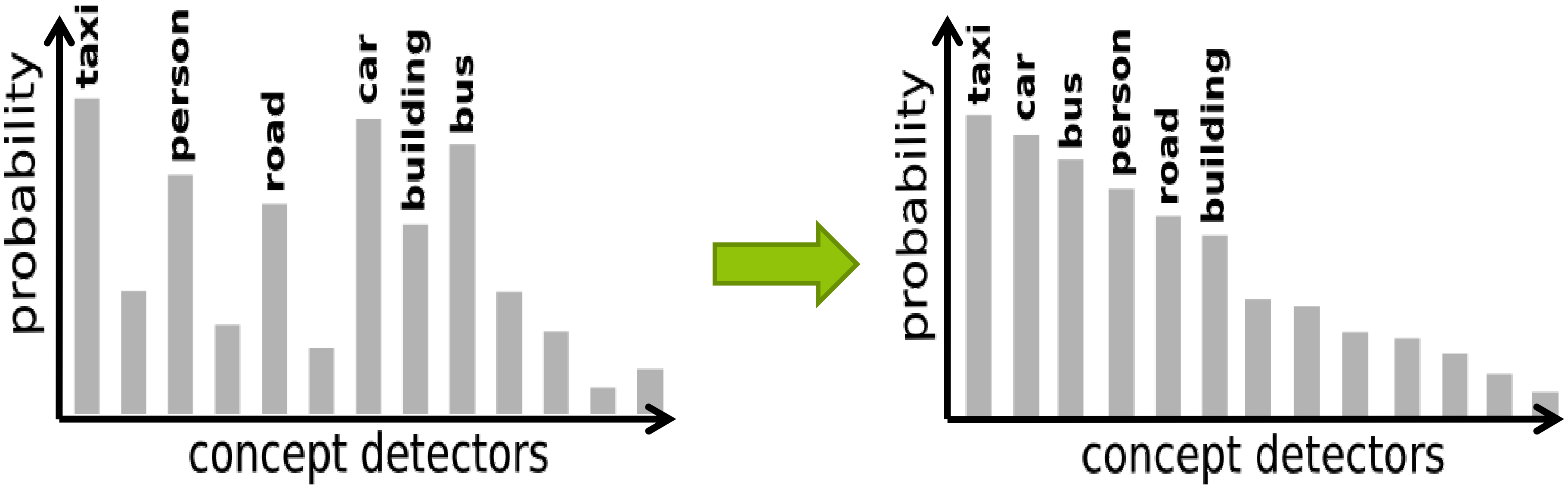
Proposed Approach: CBS

- **Sparsification process**



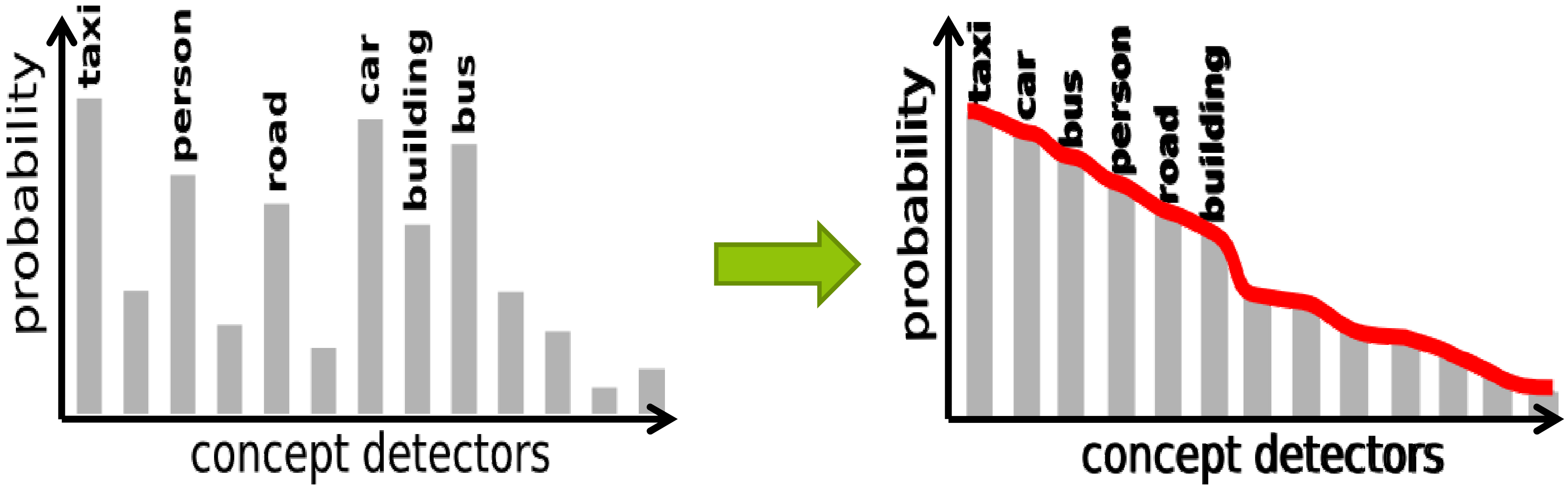
Proposed Approach: CBS

- Sparsification process



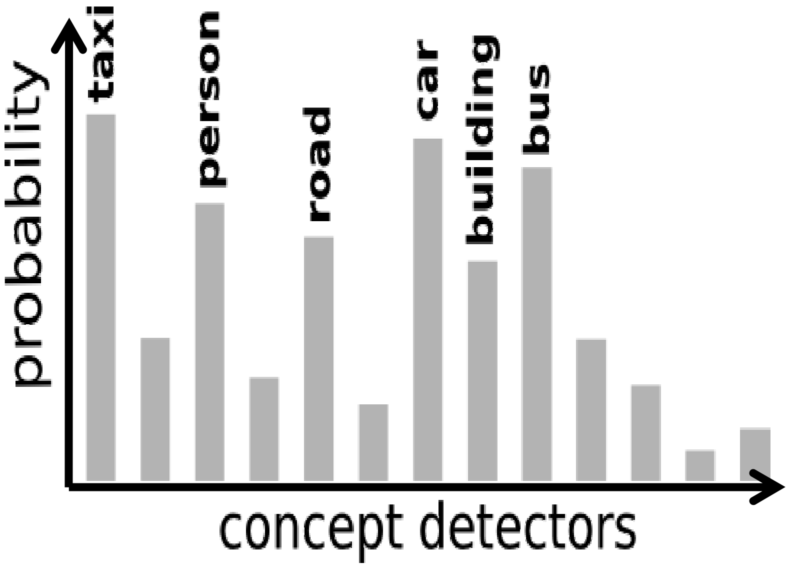
Proposed Approach: CBS

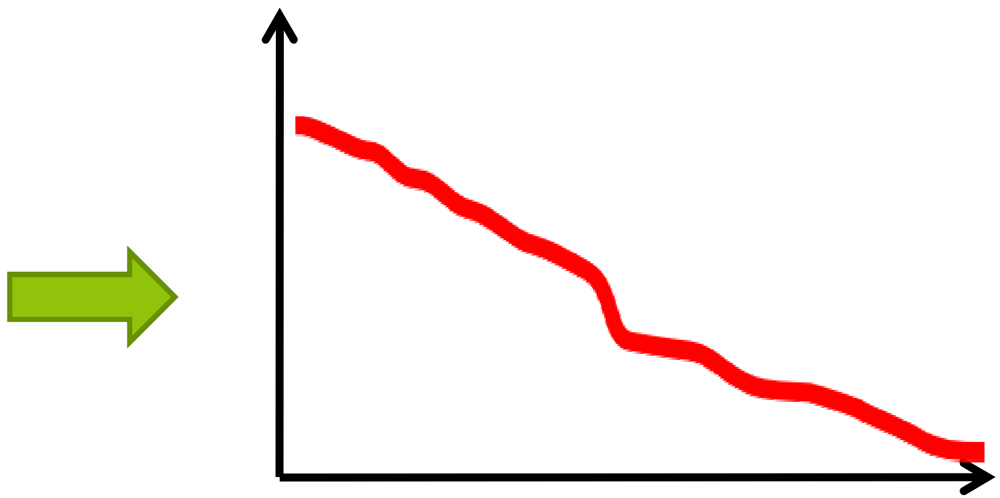
- Sparsification process



Proposed Approach: CBS

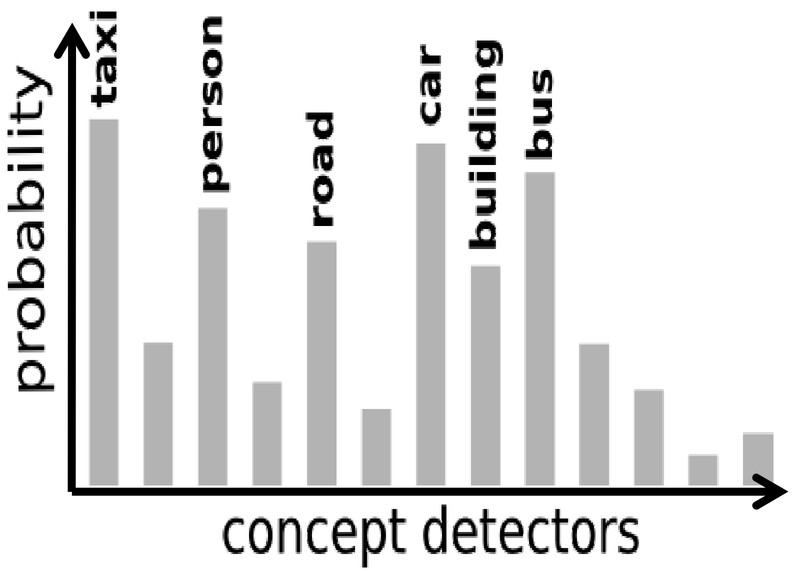
- Sparsification process

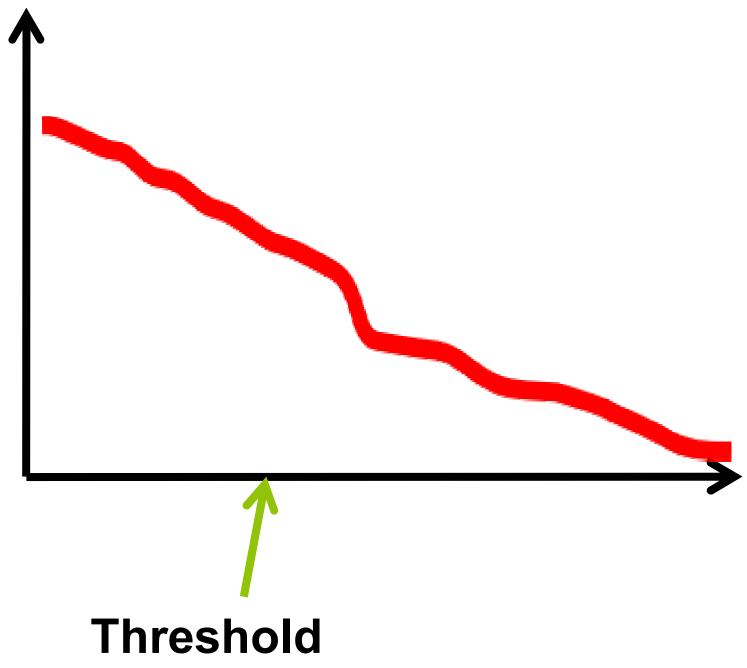




Proposed Approach: CBS

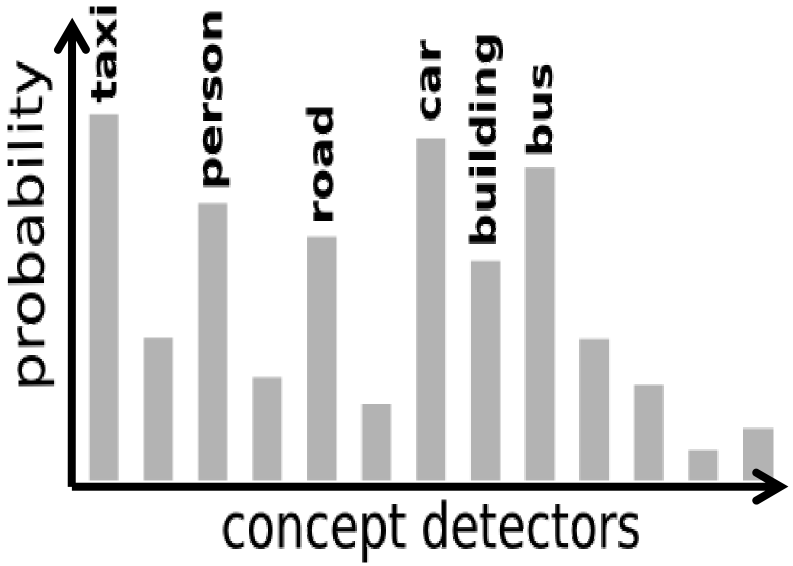
- Sparsification process

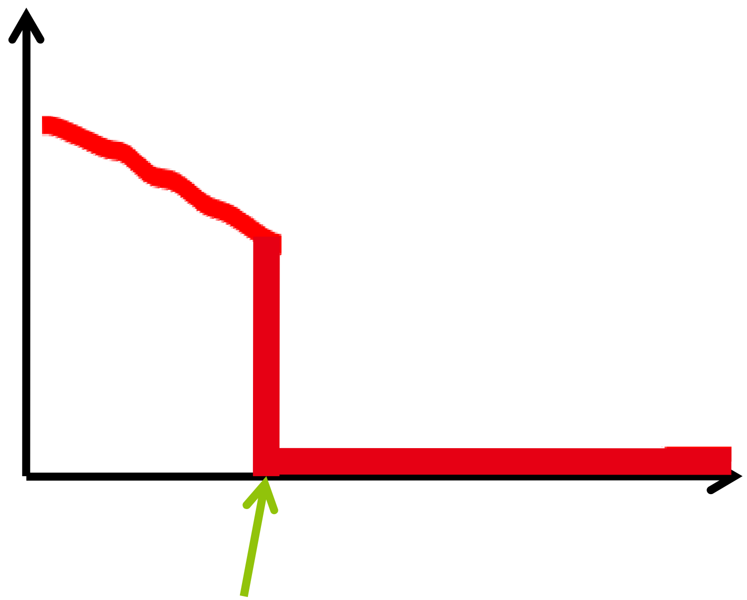




Proposed Approach: CBS

- **Sparsification process**





Proposed Approach: CBS

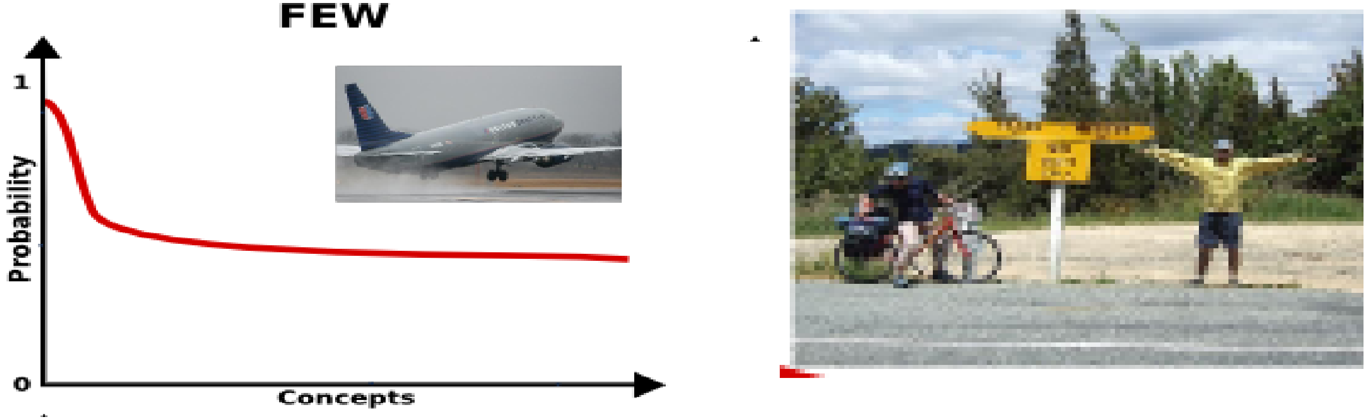
- Two things to consider:

NUMBER OF DOMINANT CONCEPTS

Proposed Approach: CBS

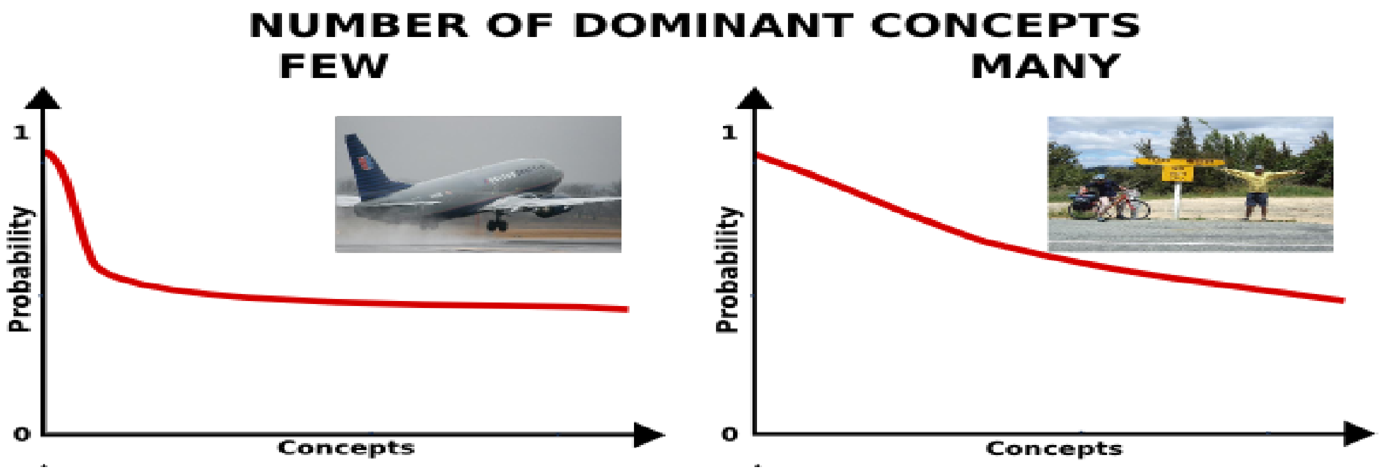
- Two things to consider:

NUMBER OF DOMINANT CONCEPTS
FEW



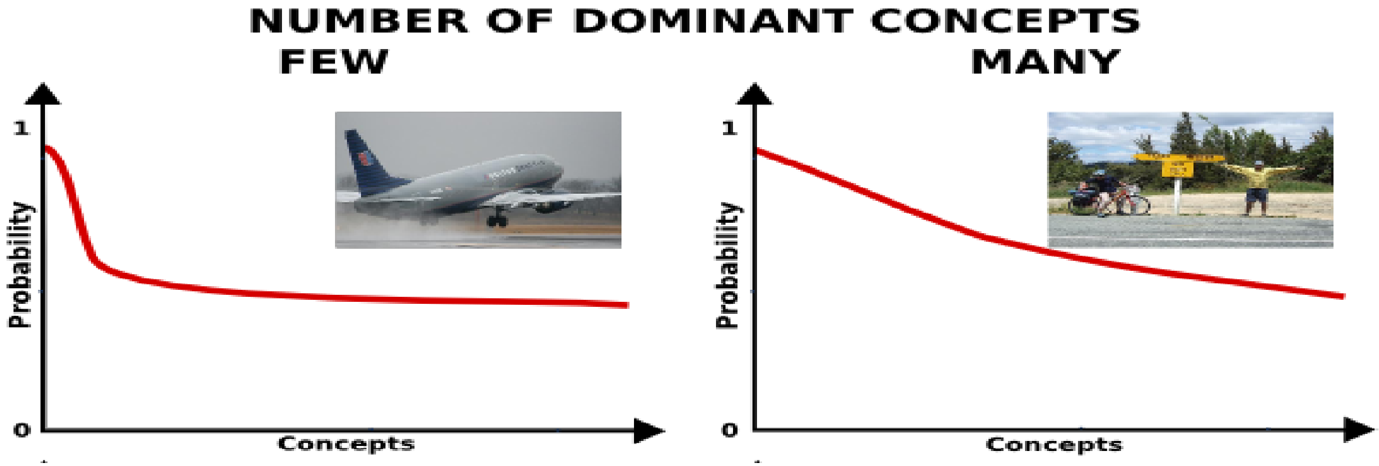
Proposed Approach: CBS

- Two things to consider:



Proposed Approach: CBS

- **Two things to consider:**



- **Modelisation**

  Amount of concepts in the image

  $\approx$

  Amount of information in the Semantic Feature

Proposed Approach: CBS

- **Estimation : Shannon Entropy**

Semantic Feature ➡ Random Variable

Random Variable ➡ **Shannon Entropy**

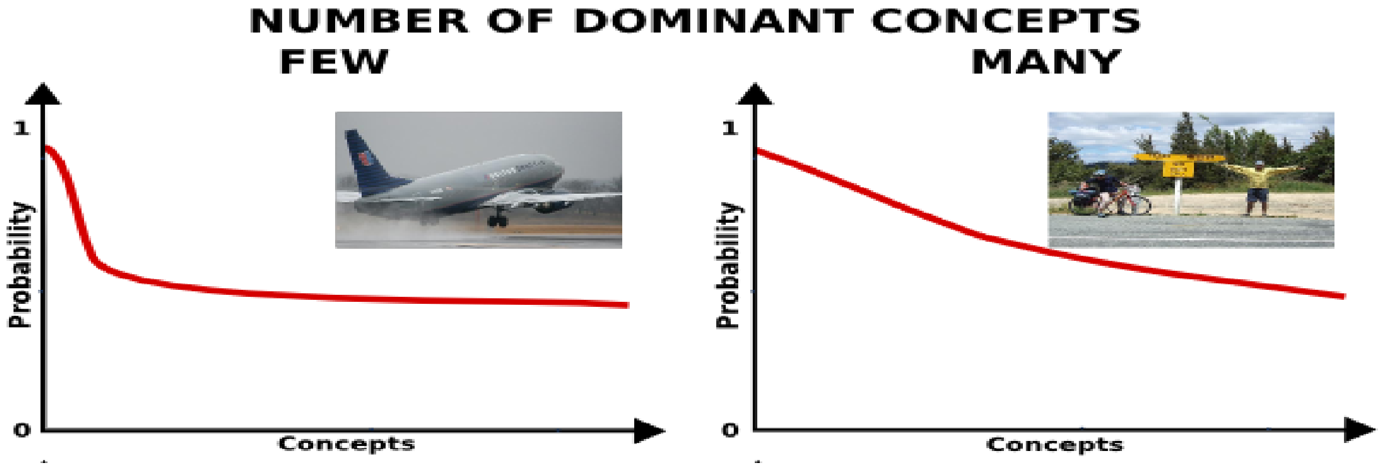
Proposed Approach: CBS

- Two things to consider:

NUMBER OF DOMINANT CONCEPTS

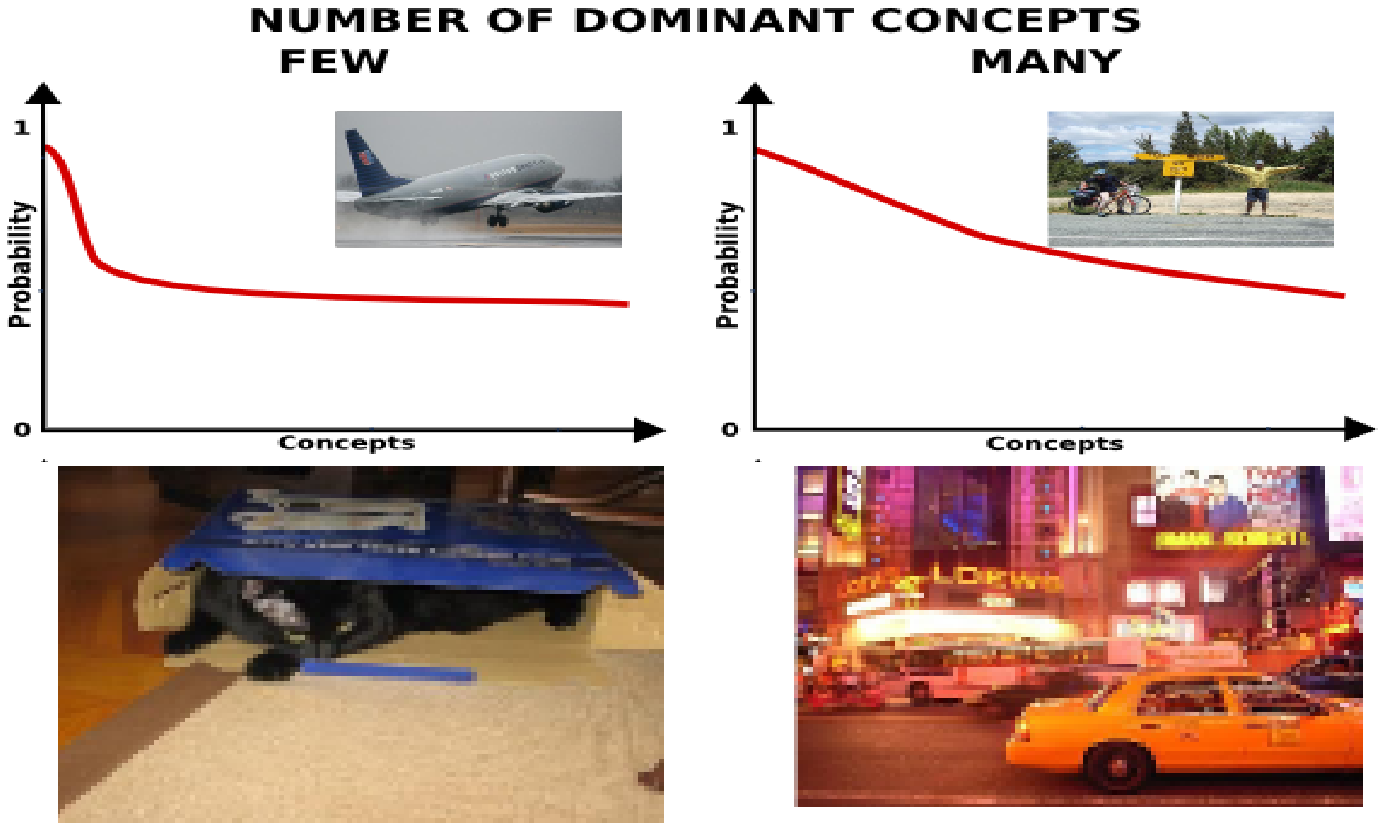
Proposed Approach: CBS

- Two things to consider:



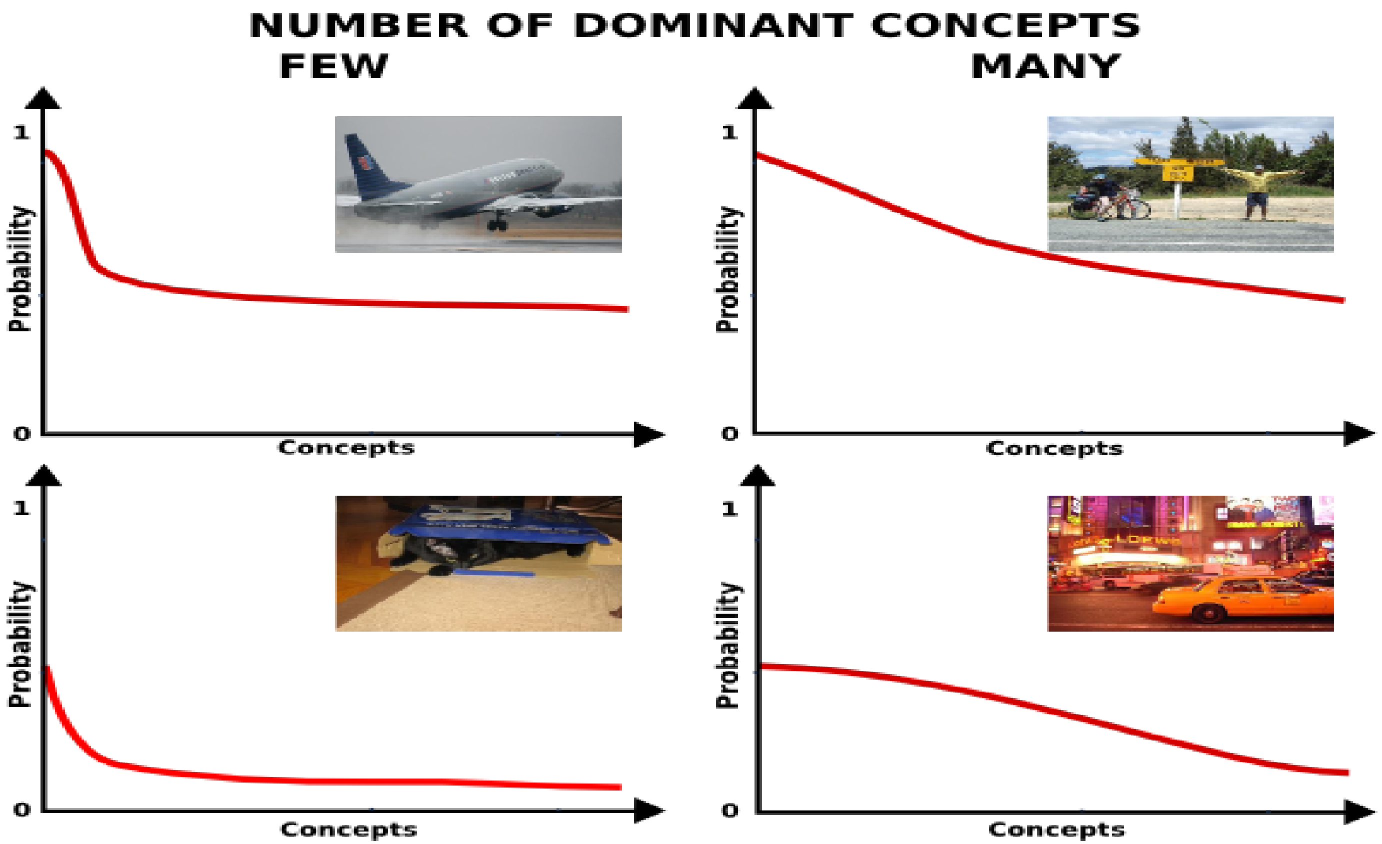
Proposed Approach: CBS

- **Two things to consider:**



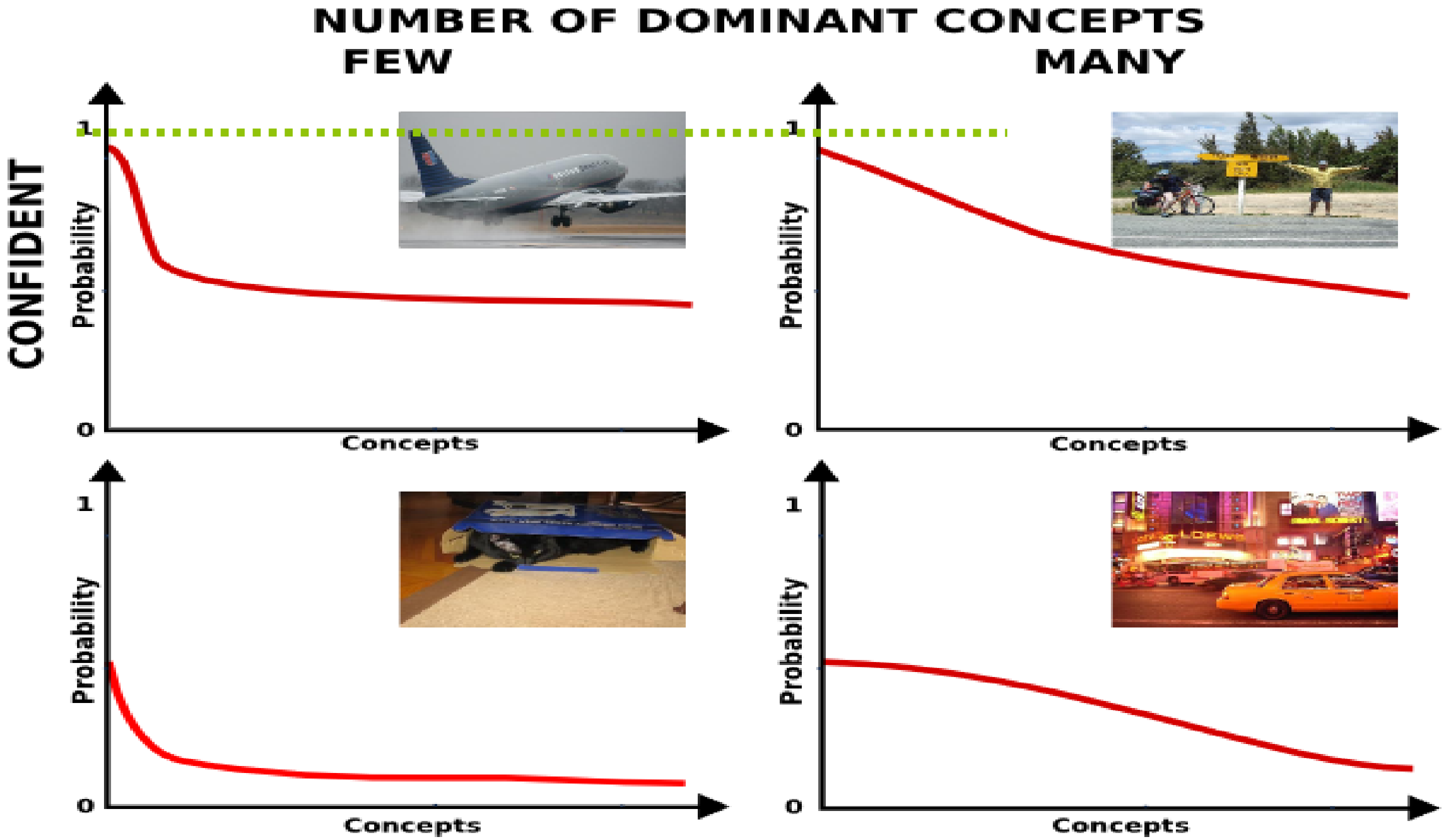
Proposed Approach: CBS

- Two things to consider:



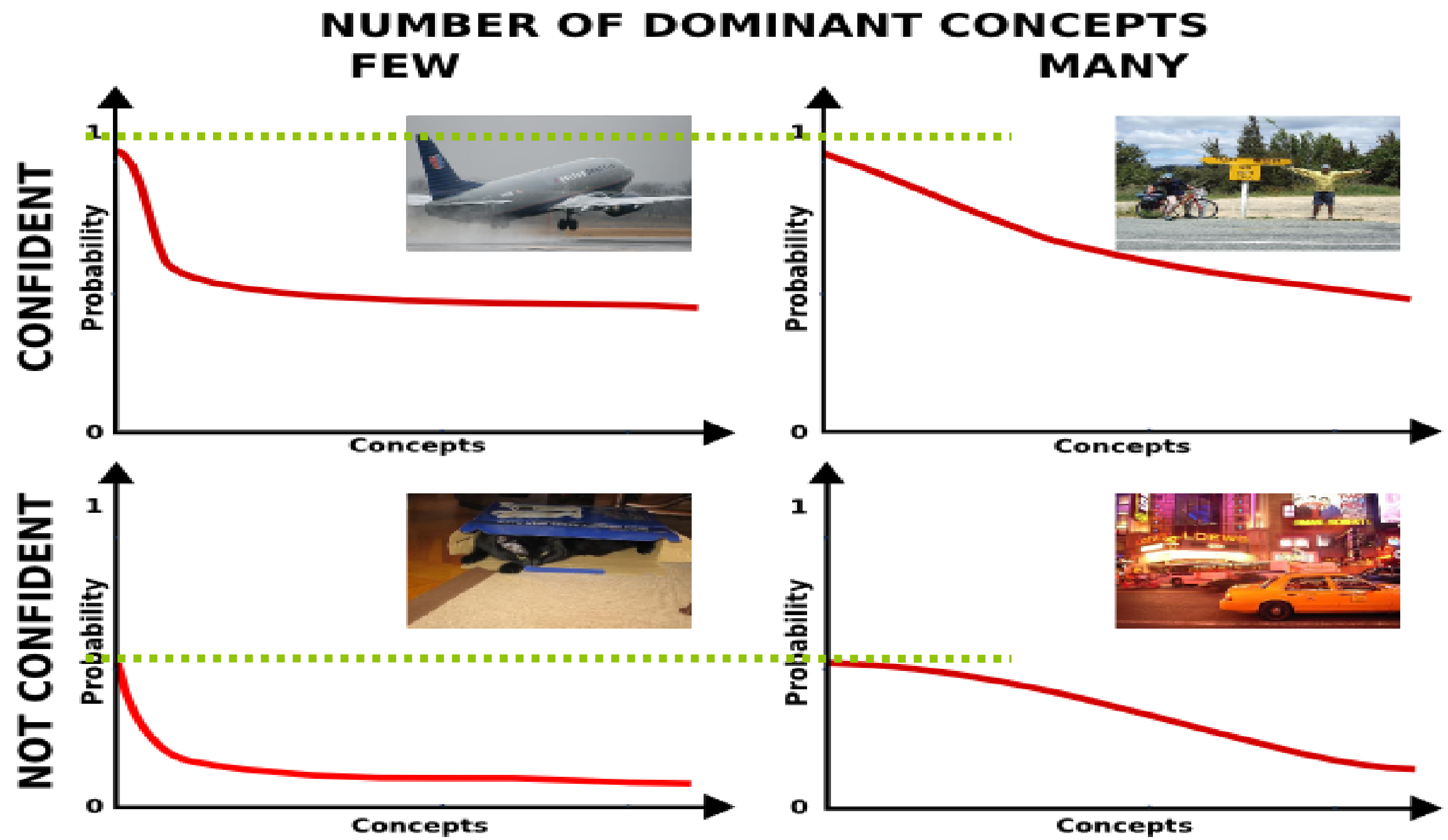
Proposed Approach: CBS

- Two things to consider:



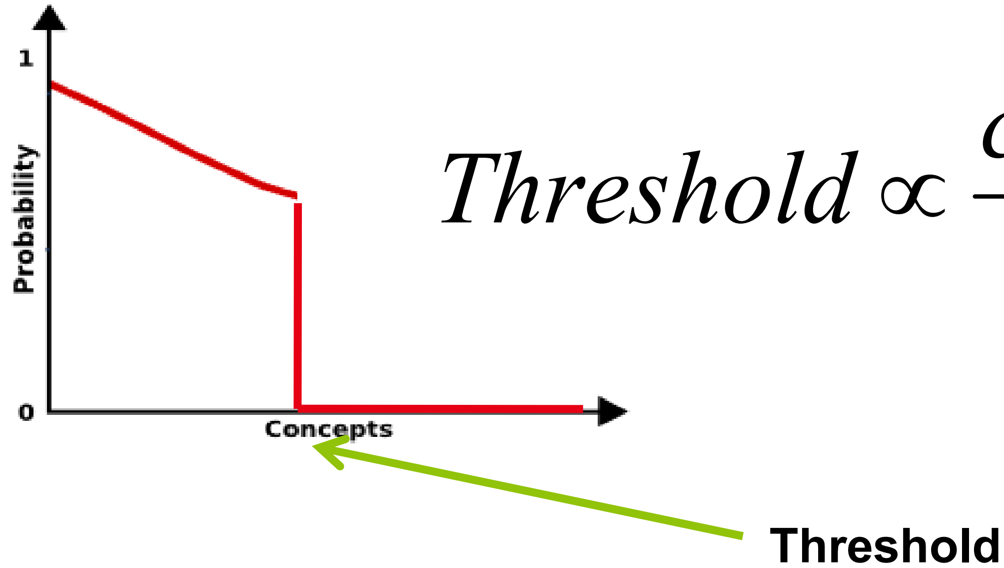
Proposed Approach: CBS

- Two things to consider:



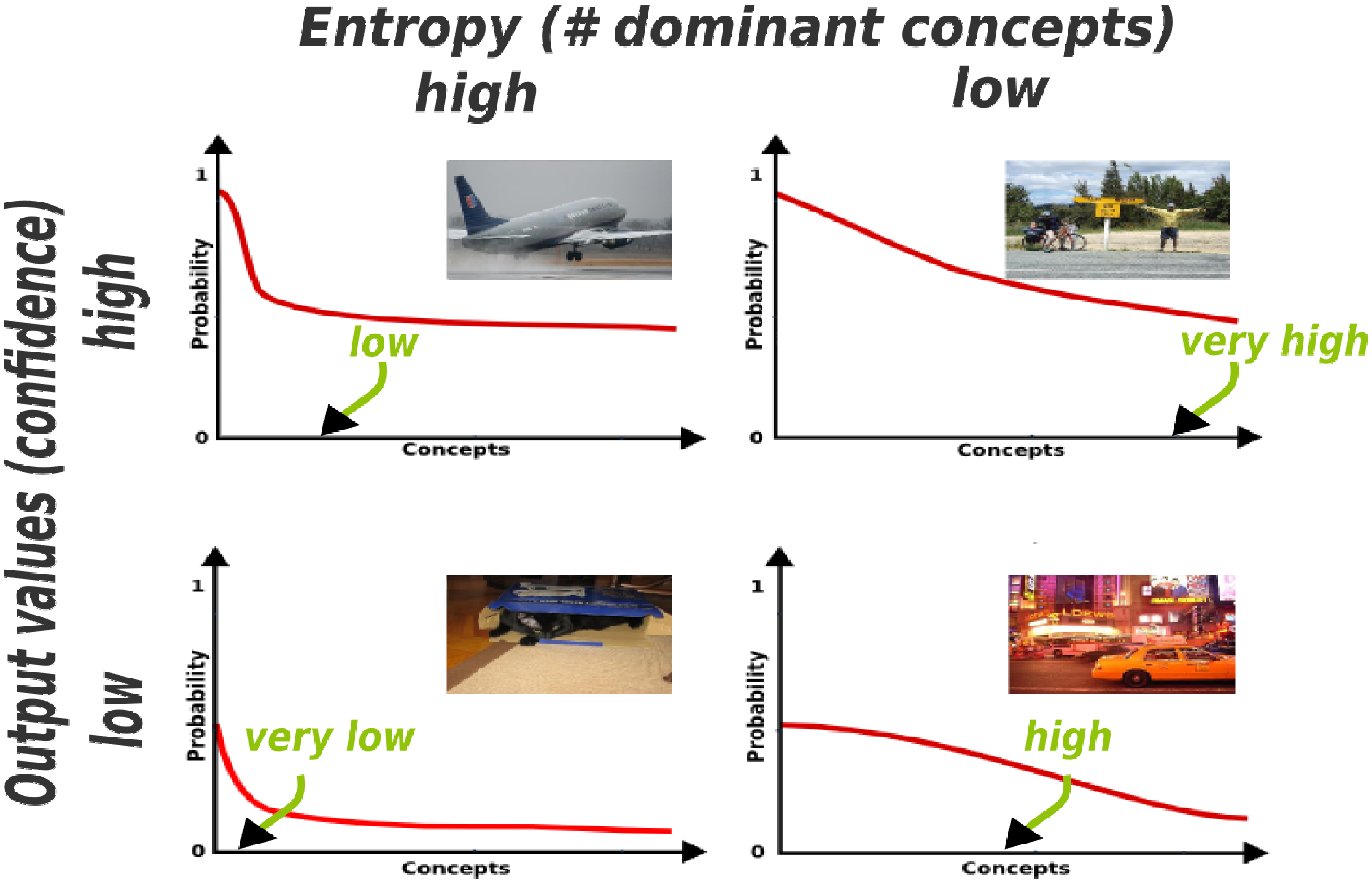
Proposed Approach: CBS

- **Modelisation**
  Confidence = value of the output detectors
- **Estimation**
  - Maximum value of the feature
- **Final Estimation of the Threshold**
  - Trade-off between entropy and confidence

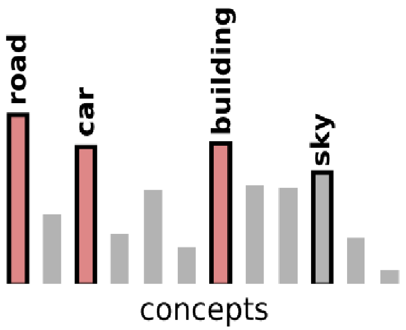


$$Threshold \propto \frac{confidence}{entropy}$$

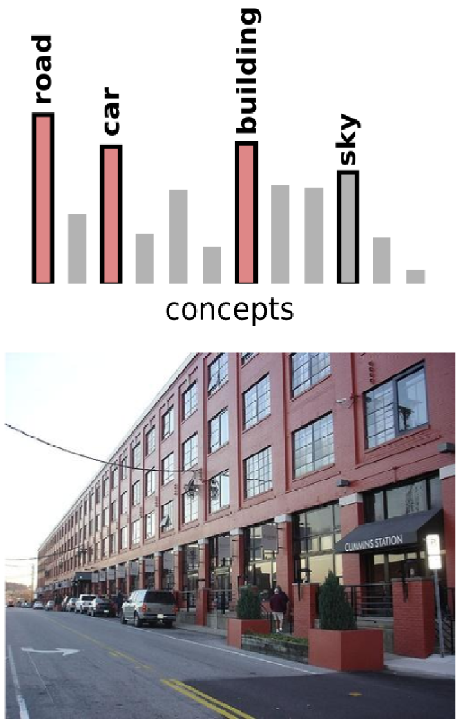
Proposed Approach: CBS



Proposed Approach: CLE

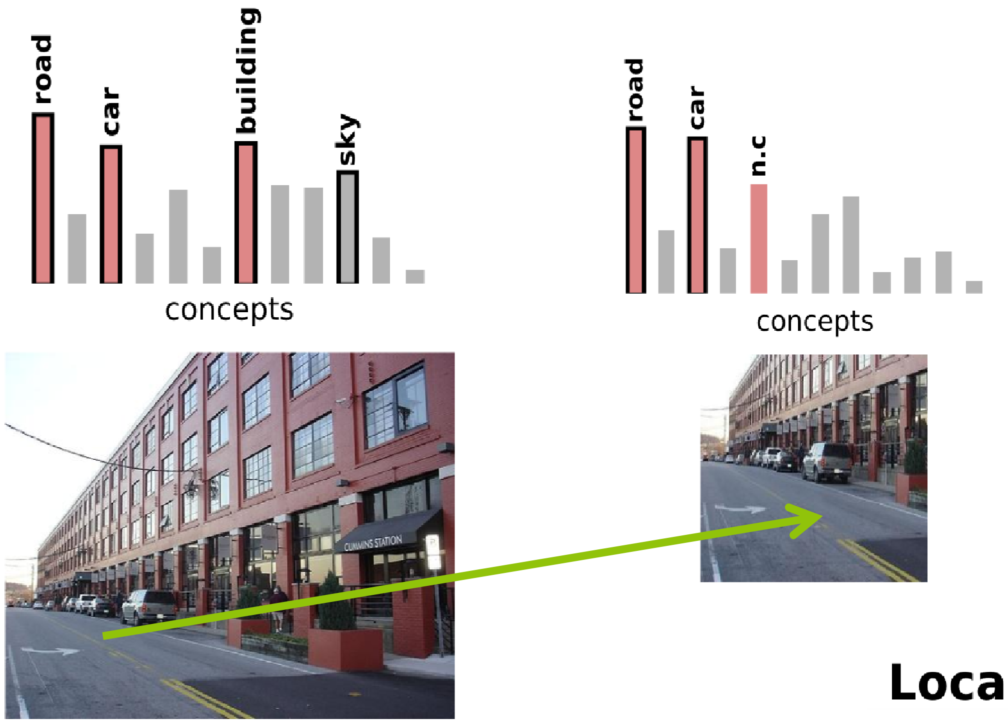


Proposed Approach: CLE



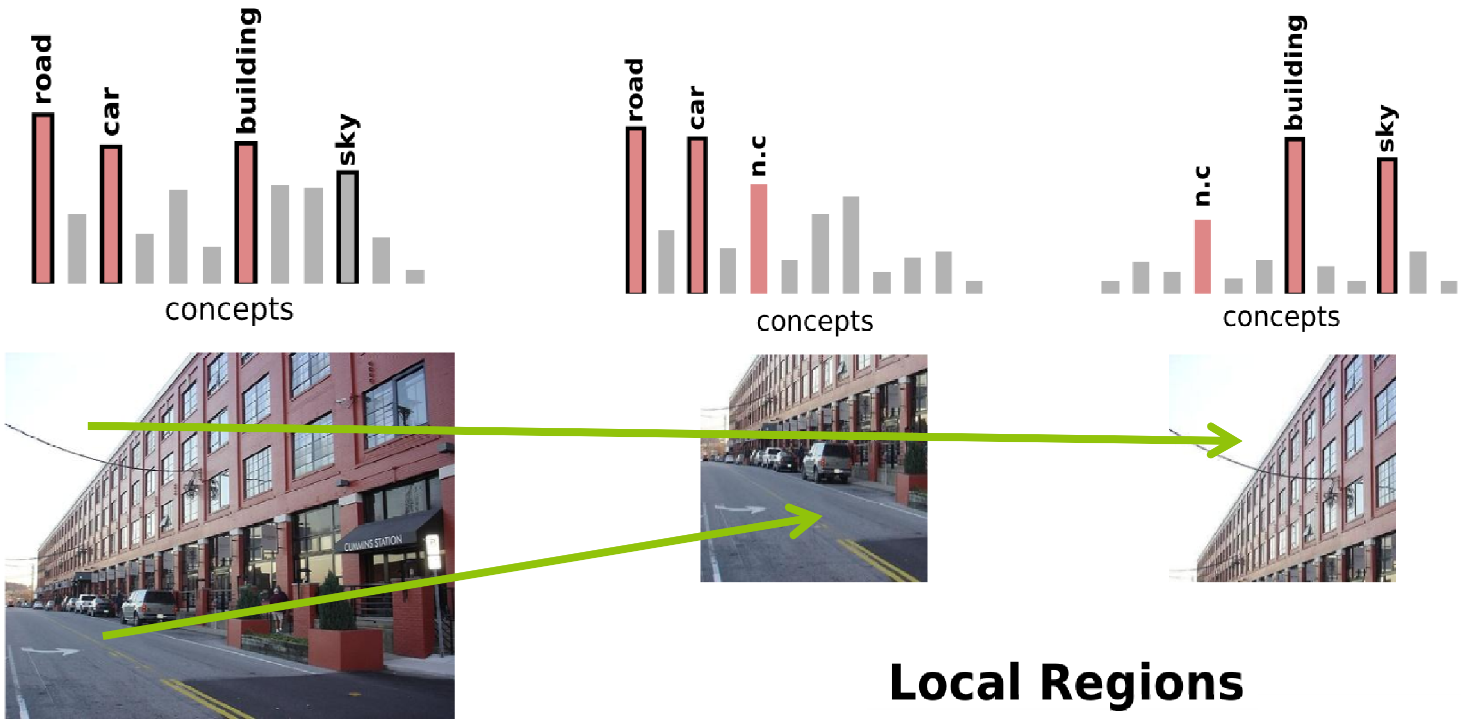
Local Regions

Proposed Approach: CLE



Local Regions

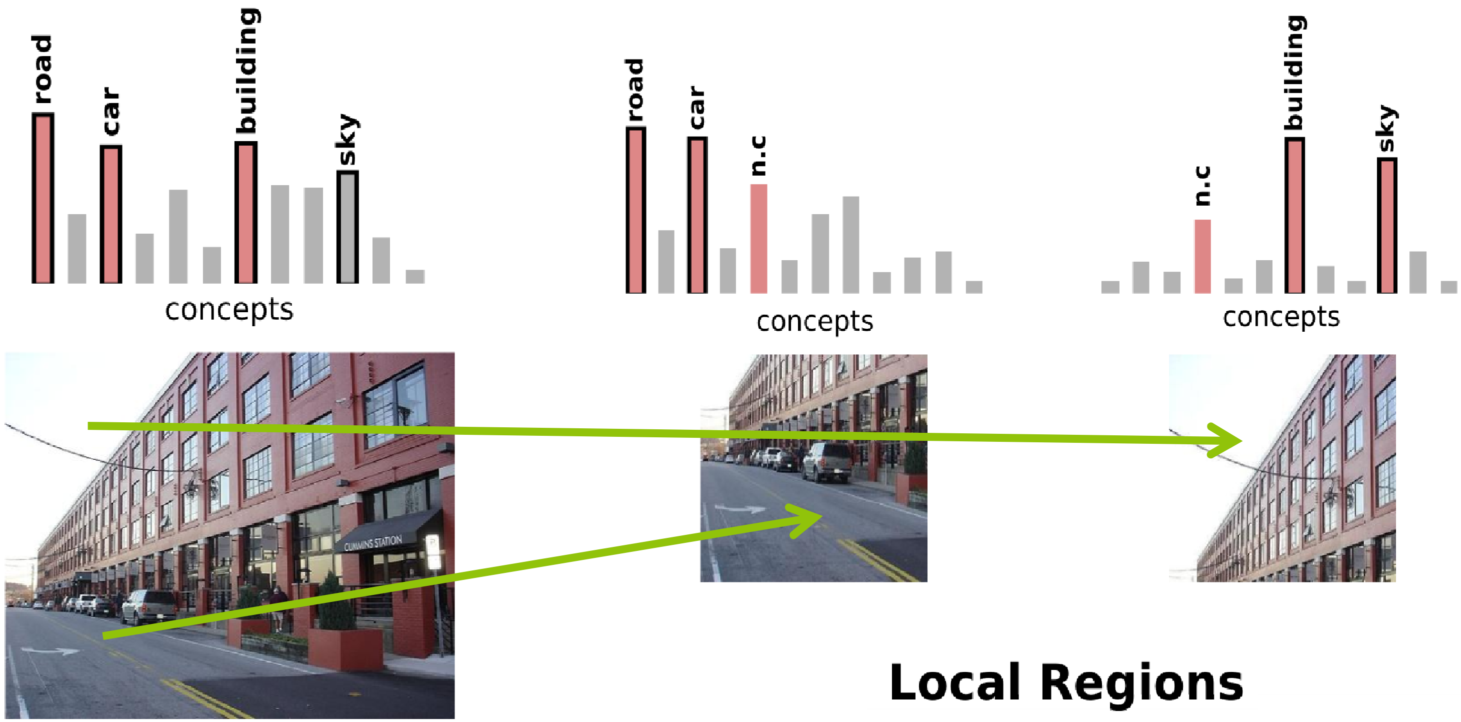
Proposed Approach: CLE



Local Regions

Proposed Approach: CLE

- Make more sense at local scale



  - consider rich-information regions
  - ignore low-information regions

Experimental Protocol

	Pascal VOC 07	Pascal VOC 12	MIT 67
Classification	✓ mAP	✓ mAP	✓ Accuracy
Retrieval	✓ mAP@K		✓ mAP@K

- **Pascal VOC 07**
  - Train/Collection: 5k – Test/Queries: 5k
- **Pascal VOC 12**
  - Train: 10k – Test: 10k
- **MIT Indoor 67**
  - Train/Collection: 5k – Test/Queries: 1k

Classification Results

	Method	VOC 2007 mAP (in%)	VOC 2012 mAP (in%)
CNN	Oquab *et al.*, 2014	77.7	n.a.
	Chatfield *et al.*, 2014	82.42	83.2
	Wei *et al.*, 2014	81.5	81.7
	Simonyan *et al.*, 2015 (VGG fc7)	86.1	84.5
Semantic	Simonyan *et al.*, 2015 (VGG fc8)	77.4	77.2
	Bergamo *et al.*, 2012	53.2	49.3
	Torresani *et al.* (reimpl.)	82.4	81.7
	Ginsca *et al.*, 2015	82.8	81.7
	CLE (ours)	**88.2**	**86.6**

Without sparsification

Fixed sparsification

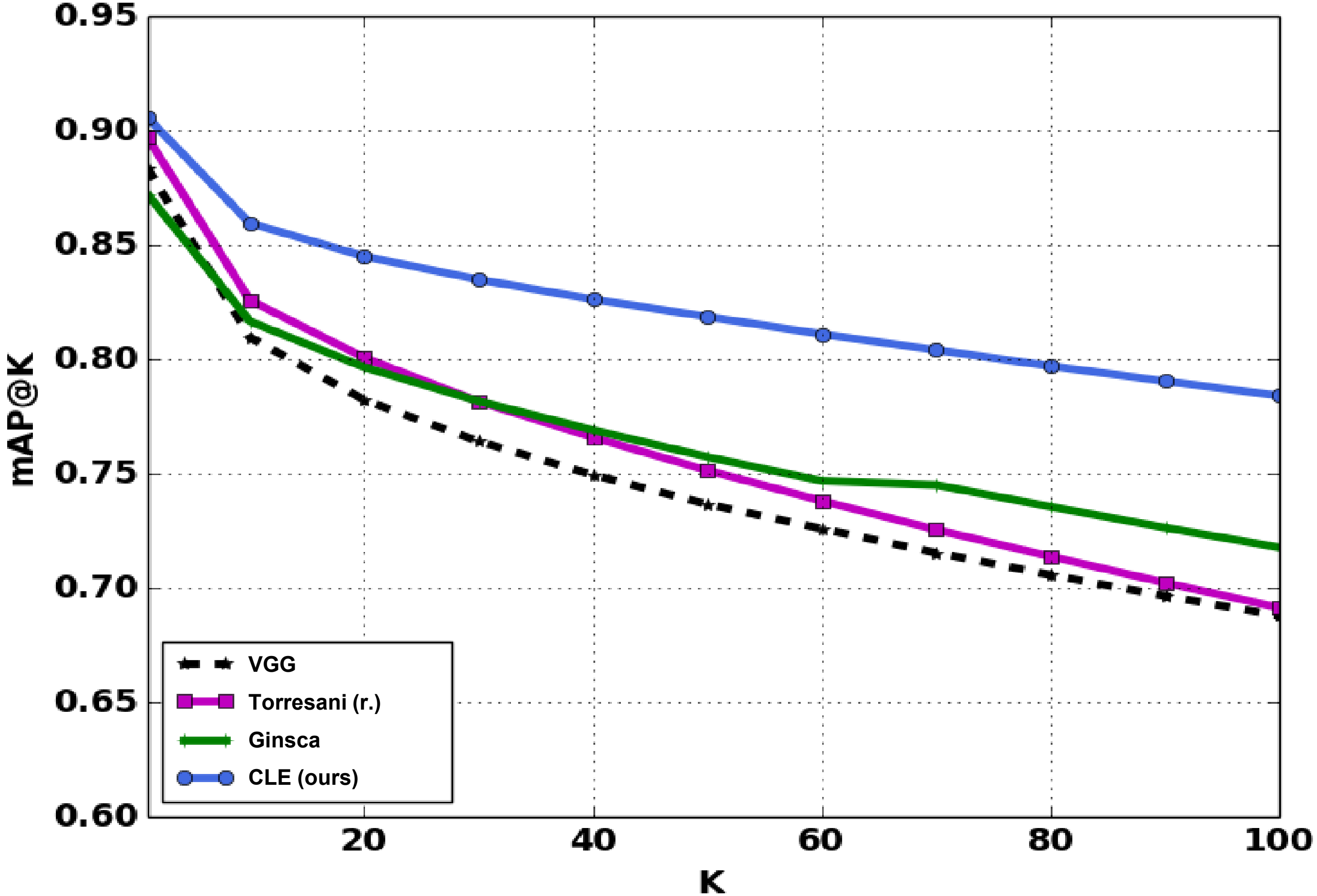
Classification Results
Scene classification

	Method	MIT Indoor 67 Classification Accuracy (in%)
	Doersch *et al.*, 2013	66.9
	Oquab *et al.*, 2014	69.0
	VGG (fc7)	65.8
	Zhou *et al.*, 2014	68.2
	Xie *et al.*, 2015 (Best paper)	70.1
Semantic	VGG (fc8)	48.7
Semantic	Bergamo *et al.*, 2012	44.6
Semantic	Torresani *et al.*, 2010 (r.)	58.9
Semantic	Ginsca *et al.*, 2015	61.5
Semantic	CLE (ours)	**71.6**

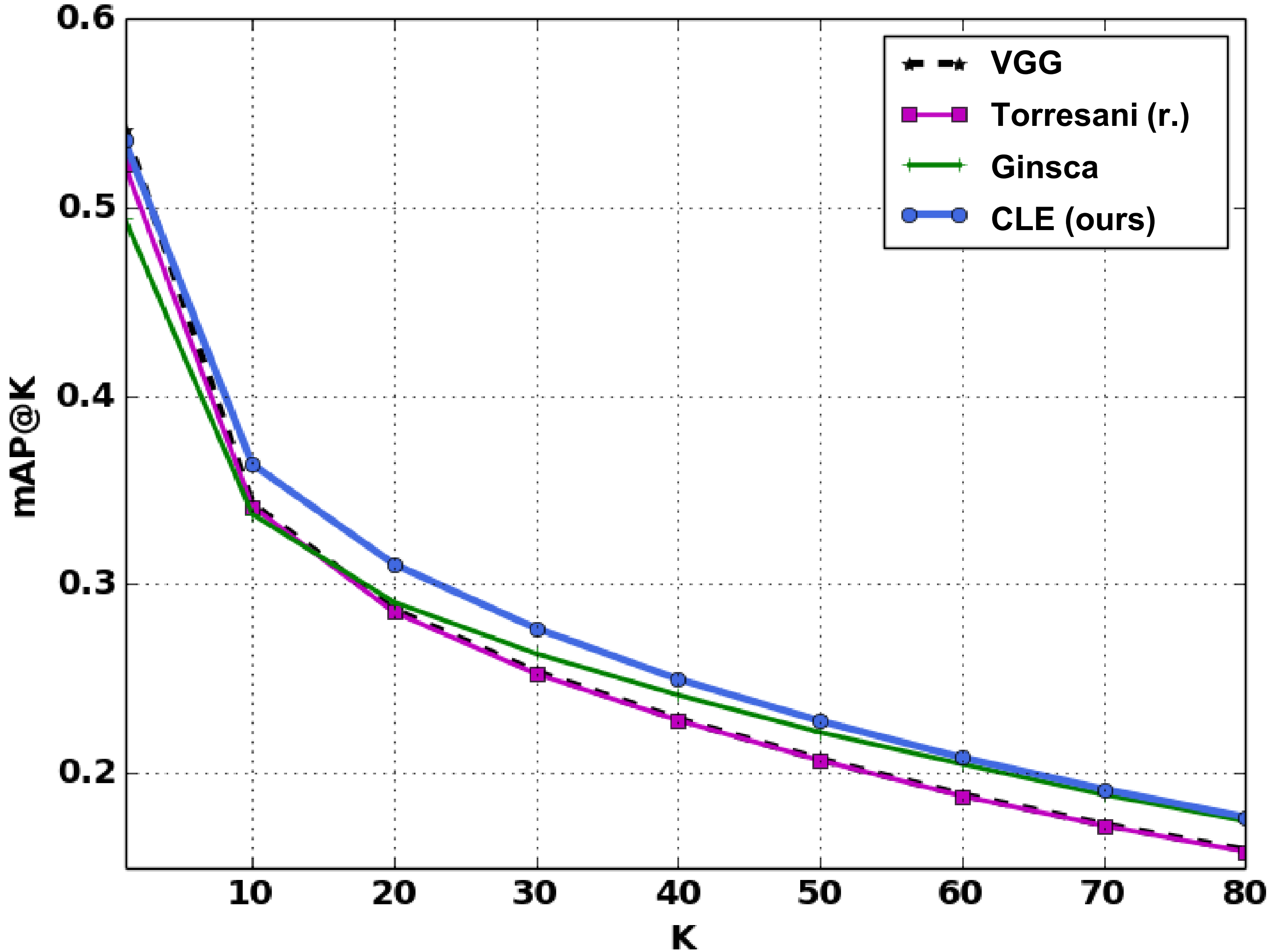
Without sparsification

Naive sparsification

Retrieval Results
Objects (VOC07)



Retrieval Results
Indoor scenes (MIT 67)



Conclusion

- **Novelty:**
  - New semantic image-representation
  - Level of sparsity adapted to the content of each image
  - Constrained local regions

- **Results:**
  - Image Classification
    - +2 points of mAP compared to the best CNN Feature
    - +5 points of mAP compared to the best Semantic Feature
  - Image Retrieval
    - +5 points of mAP compared to the best CNN Feature
    - +3 points of mAP compared to the best Semantic Feature

Thank you (questions ?)

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA SACLAY NANO-INNOV | BAT. 861 – PC142
91191 Gif-sur-Yvette Cedex - FRANCE
www-list.cea.fr