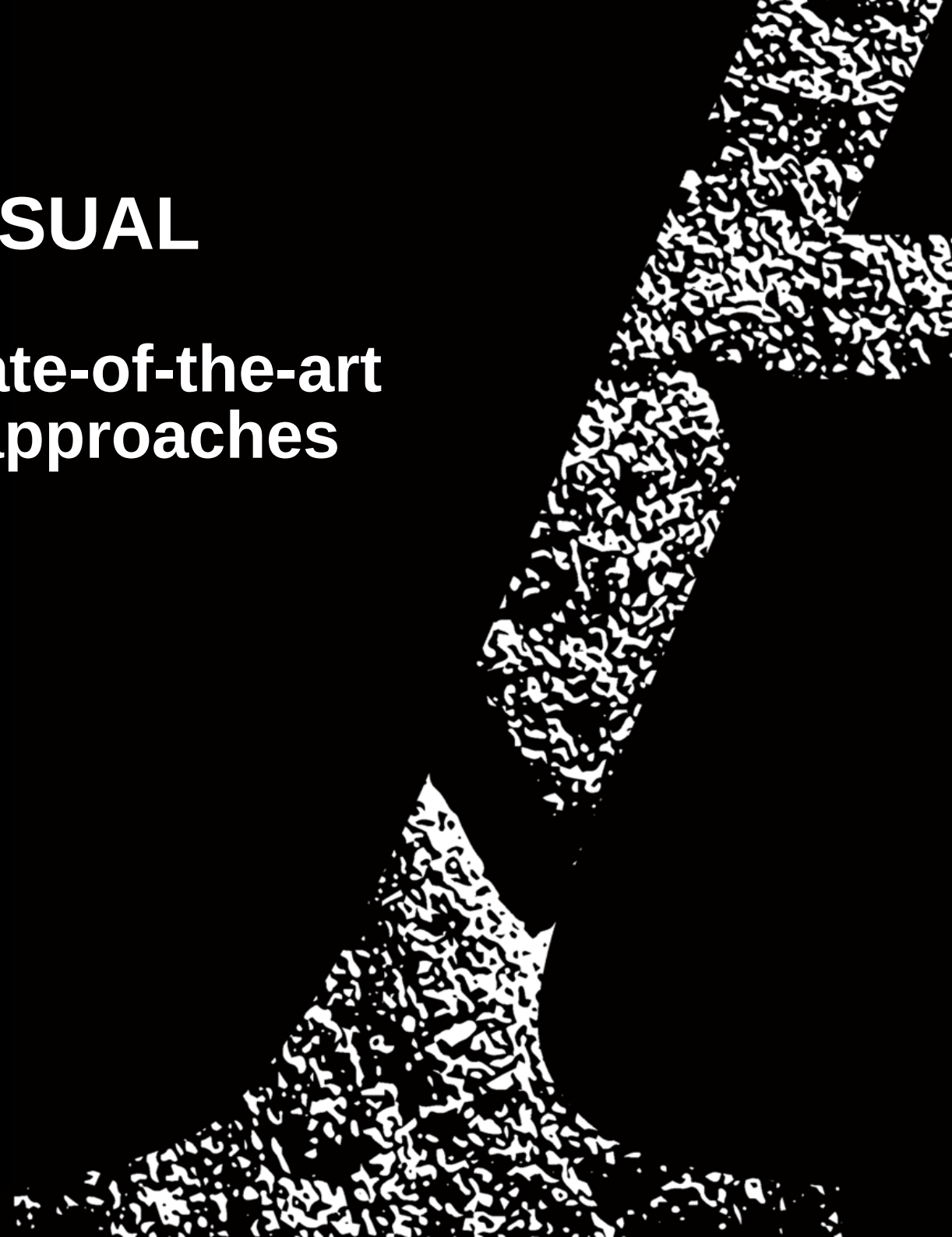
INSCRIPTIONS VISUAL RECOGNITION

A comparison of state-of-the-art object recognition approaches

Giuseppe Amato
Fabrizio Falchi
Fausto Rabitti
Lucia Vadicamo

ISTI-CNR, Pisa
lucia.vadicamo@isti.cnr.it



Why inscriptions visual recognition?

- Objective:
  - Offering the possibility of retrieving information on inscriptions from images
  - Both, using traditional Web interfaces and mobile devices

Why inscriptions visual recognition?

- Objective:
  - Offering the possibility of retrieving information on inscriptions from images
  - Both, using traditional Web interfaces and mobile devices

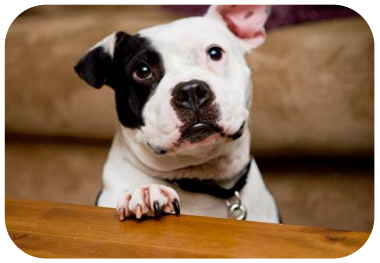
Why inscriptions visual recognition?

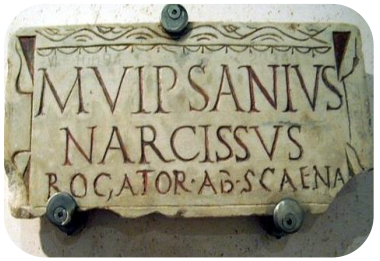
- Objective:
  - Offering the possibility of retrieving information on inscriptions from images
  - Both, using traditional Web interfaces and mobile devices

Why inscriptions visual recognition?

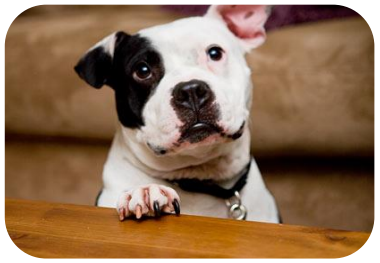
- Objective:
  - Offering the possibility of retrieving information on inscriptions from images
  - Both, using traditional Web interfaces and mobile devices

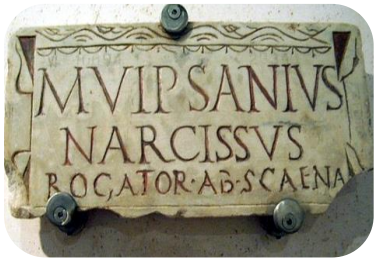
Where is the inscription  ?

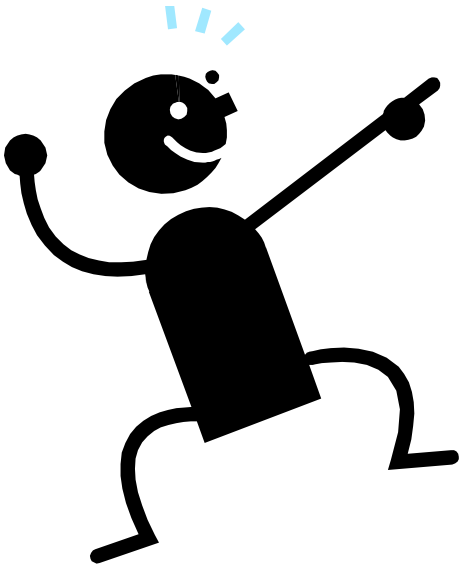




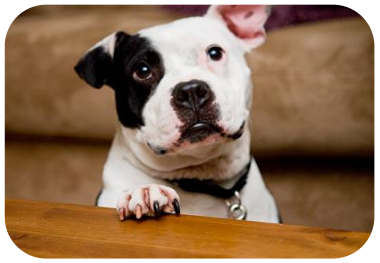
Where is the inscription  ?

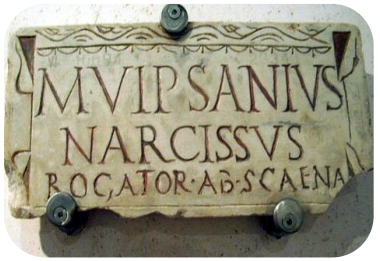


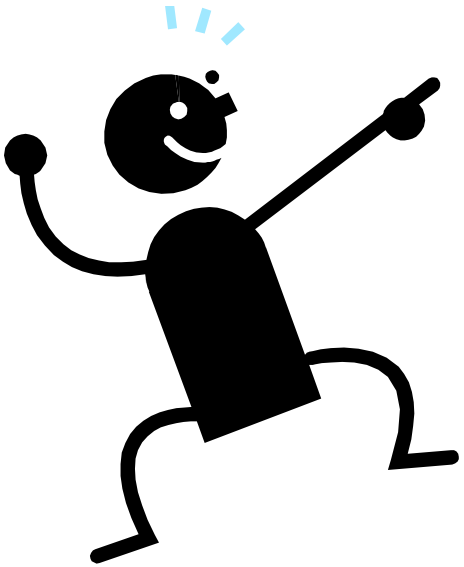




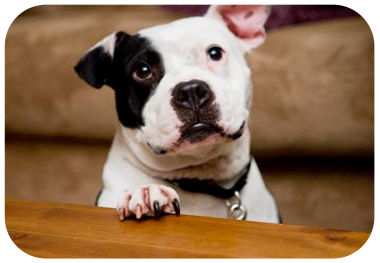
Where is the inscription  ?

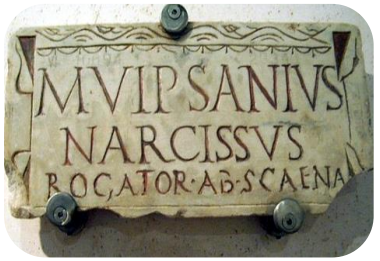


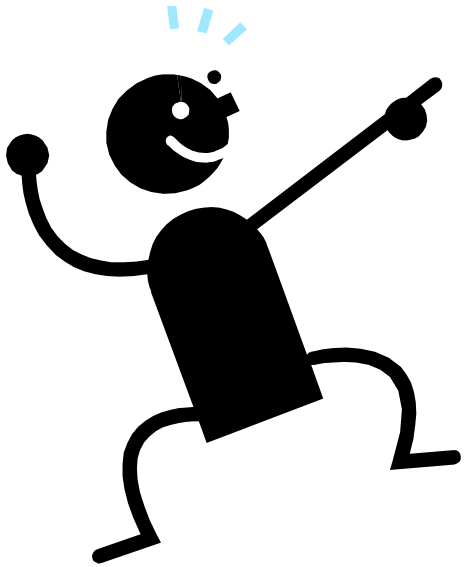




Where is the inscription  ?







Images need appropriate mathematical representation

Image recognition

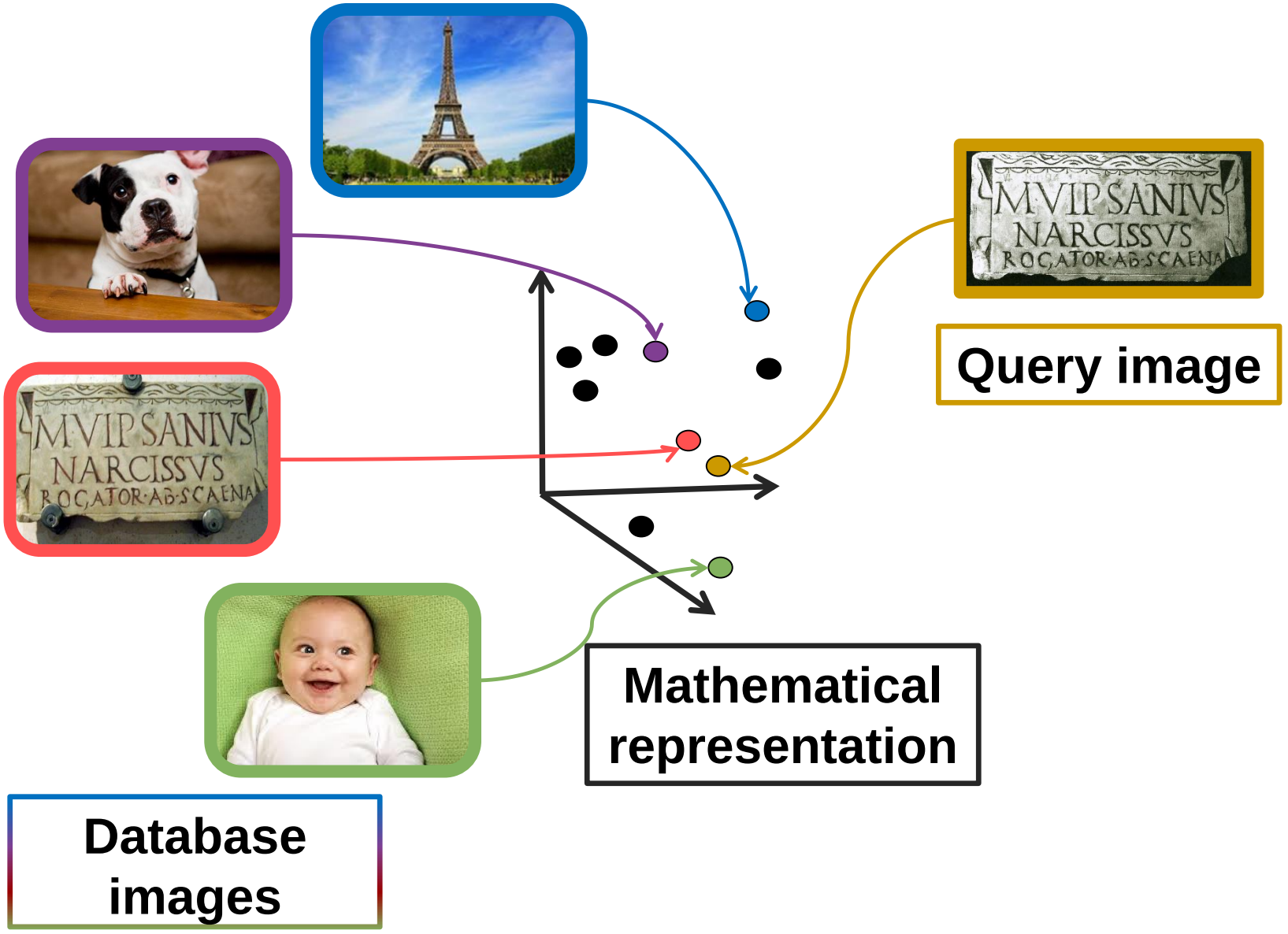


Image recognition

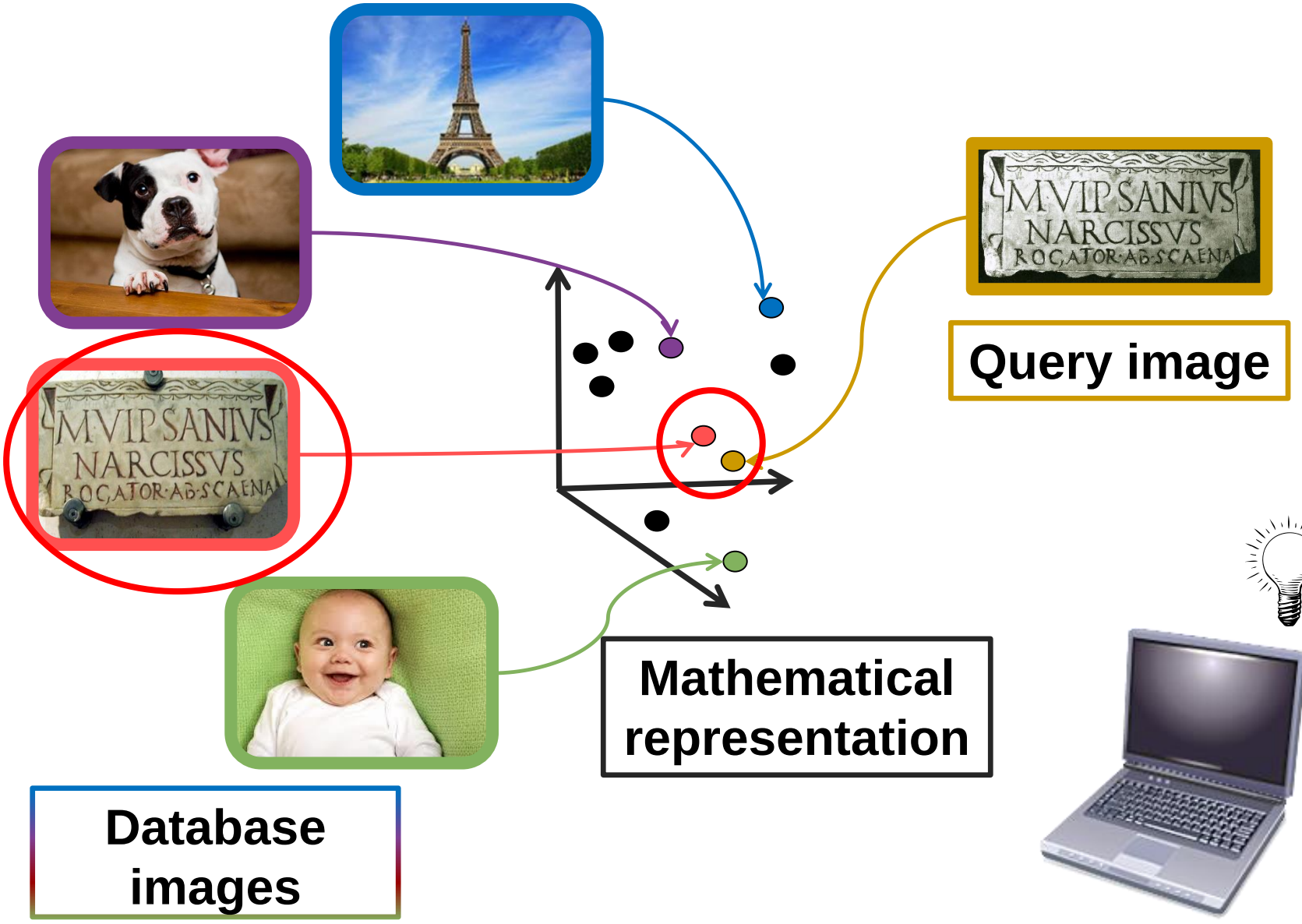
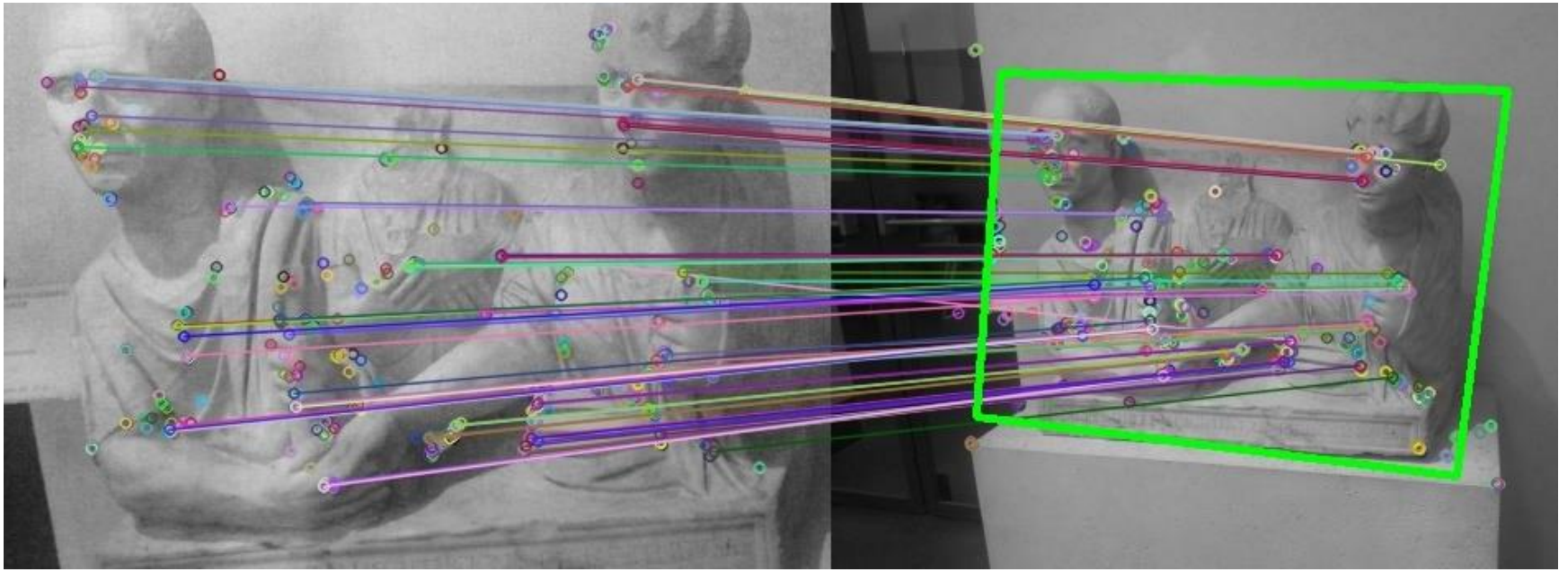


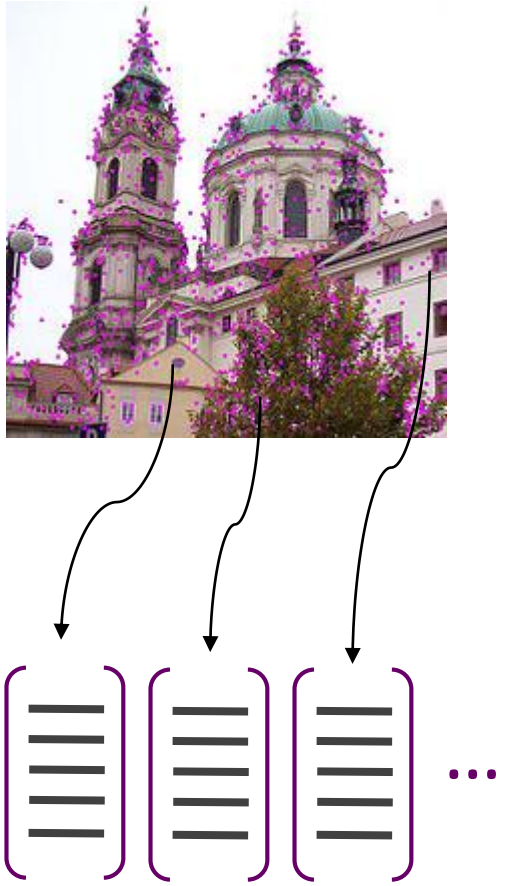
Image recognition background



- Using **Local features**: mathematical representation of visual appearance of specific areas (keypoints) of images
- Images are compared by analyzing matching local feature pairs
  - Matching pairs are selected by computing pairwise similarity among local features in different images

Local features

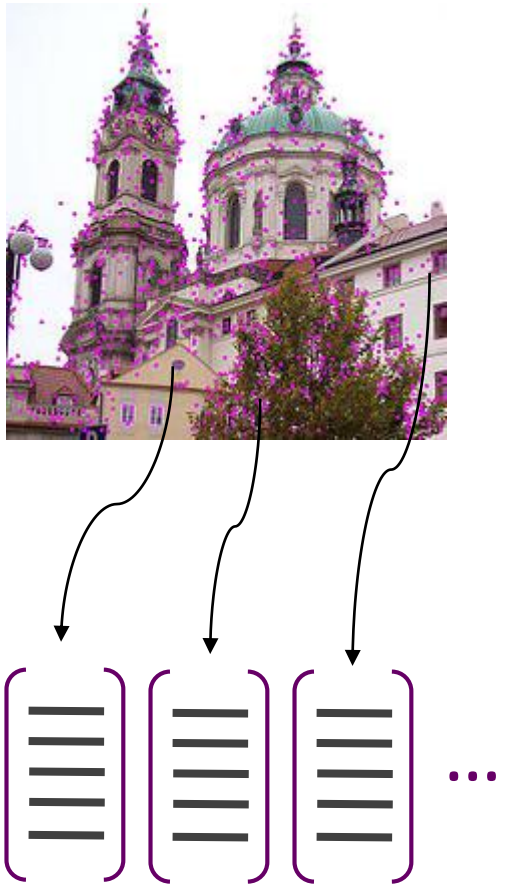
- Local features extraction:
  - **Keypoints detection**
    - Uses strategies to decide which points in the image are relevant to be represented
  - **Descriptor building**
    - Build a descriptor (e.g. vector) for each keypoint in the image



- SIFT= Scale Invariant Feature Transform
  - The most important and cited local feature
  - [Lowe 1999]

Local features

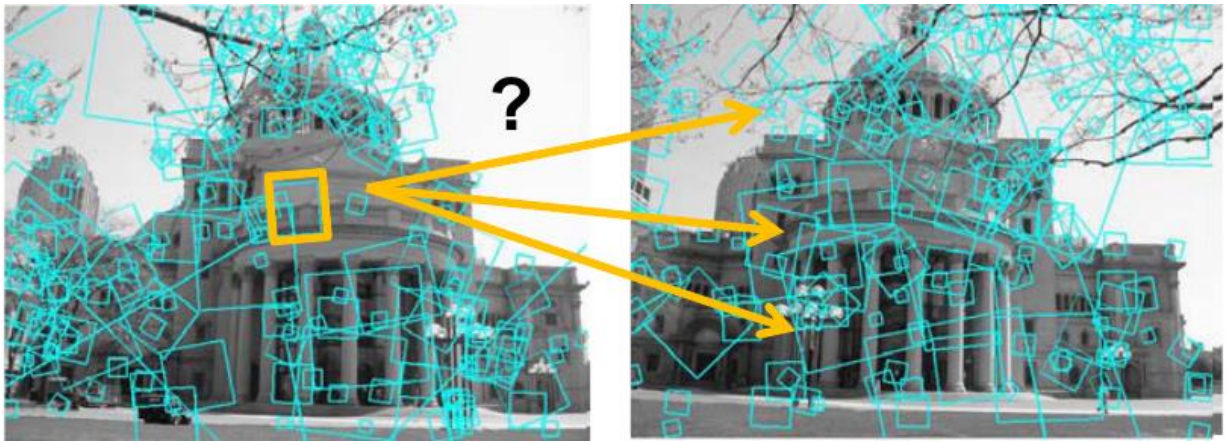
- Local features extraction:
  - **Keypoints detection**
    - Uses strategies to decide which points in the image are relevant to be represented
  - **Descriptor building**
    - Build a descriptor (e.g. vector) for each keypoint in the image



- SIFT= Scale Invariant Feature Transform
  - The most important and cited local feature
  - [Lowe 1999]

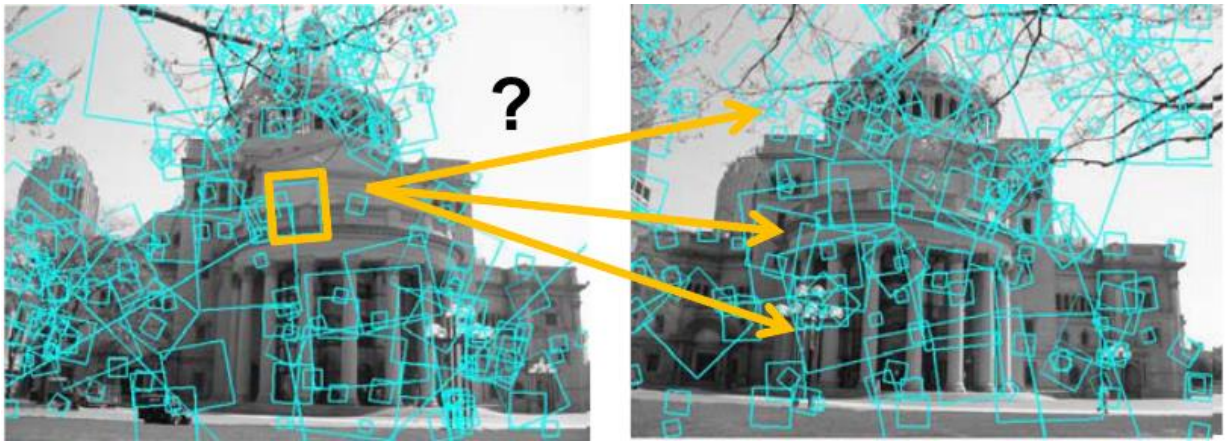
Comparing images

- Matching local features is expensive
  - Too many keypoints (thousands) per image



Comparing images

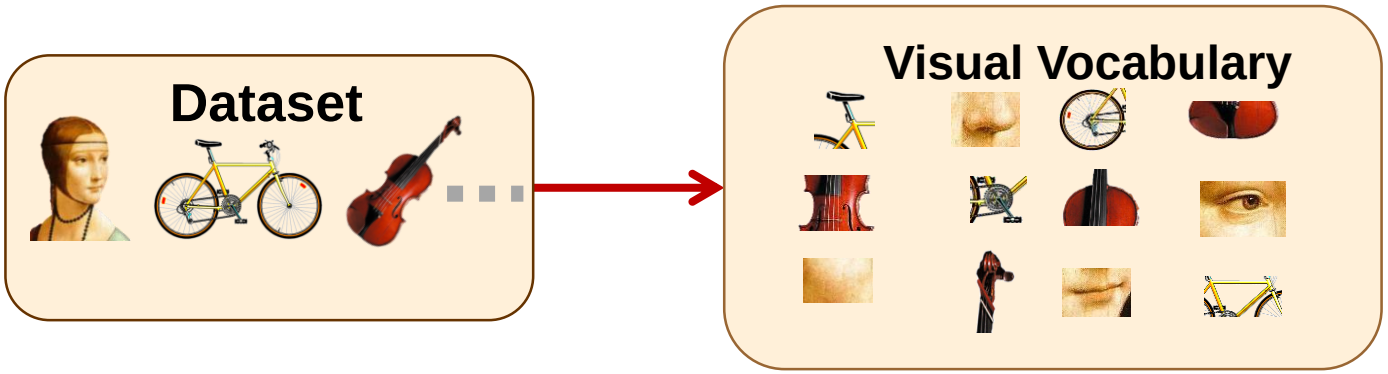
- Matching local features is expensive
  - Too many keypoints (thousands) per image



- Solution: aggregate local feature

Local Features Aggregation

- Aggregation approaches:
  - **Bag-of-Features** (BoF)
    - [Sivic & Zisserman 2003]
  - **Vector of Locally Aggregated Descriptor** (VLAD)
    - [Jégou et al. 2010]
- Both use a visual vocabulary in the aggregation phase



  - BoF vocabulary ~ 100·000 – 400·000 elements
  - VLAD vocabulary ~ 64-512 elements


Experiments

Dataset: EDR

17,155 photos

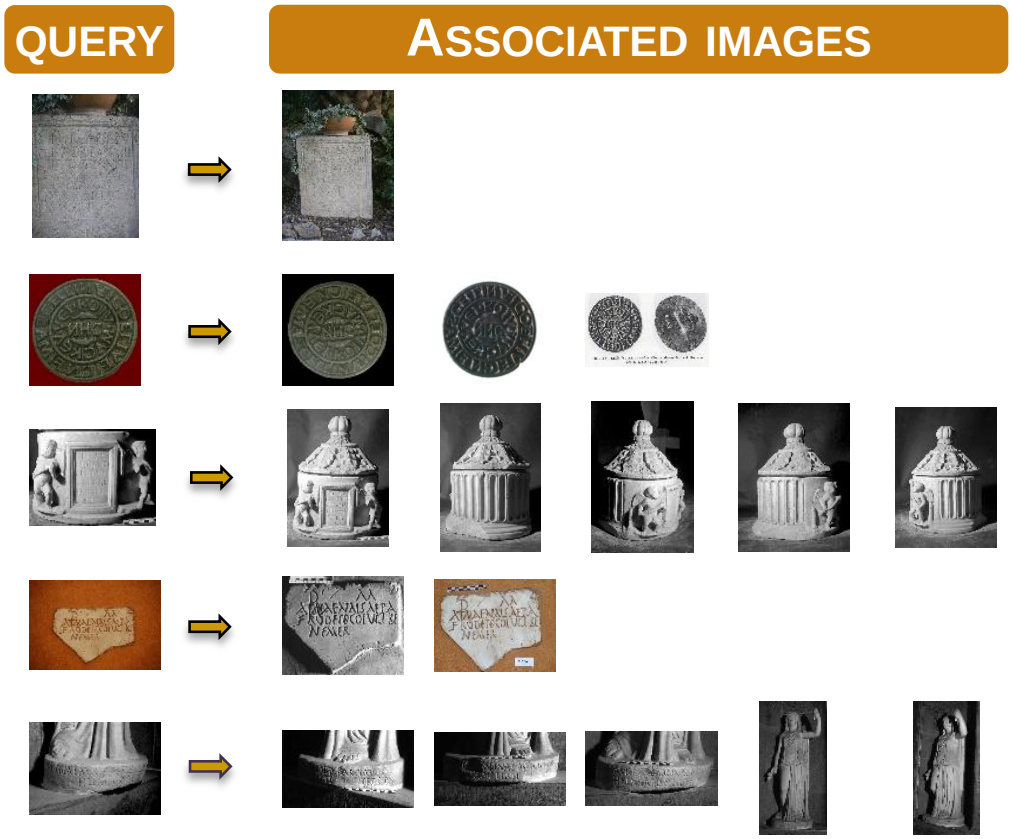
14,560 inscriptions

Epigraphic Database Roma (EDR)

EAGLE project

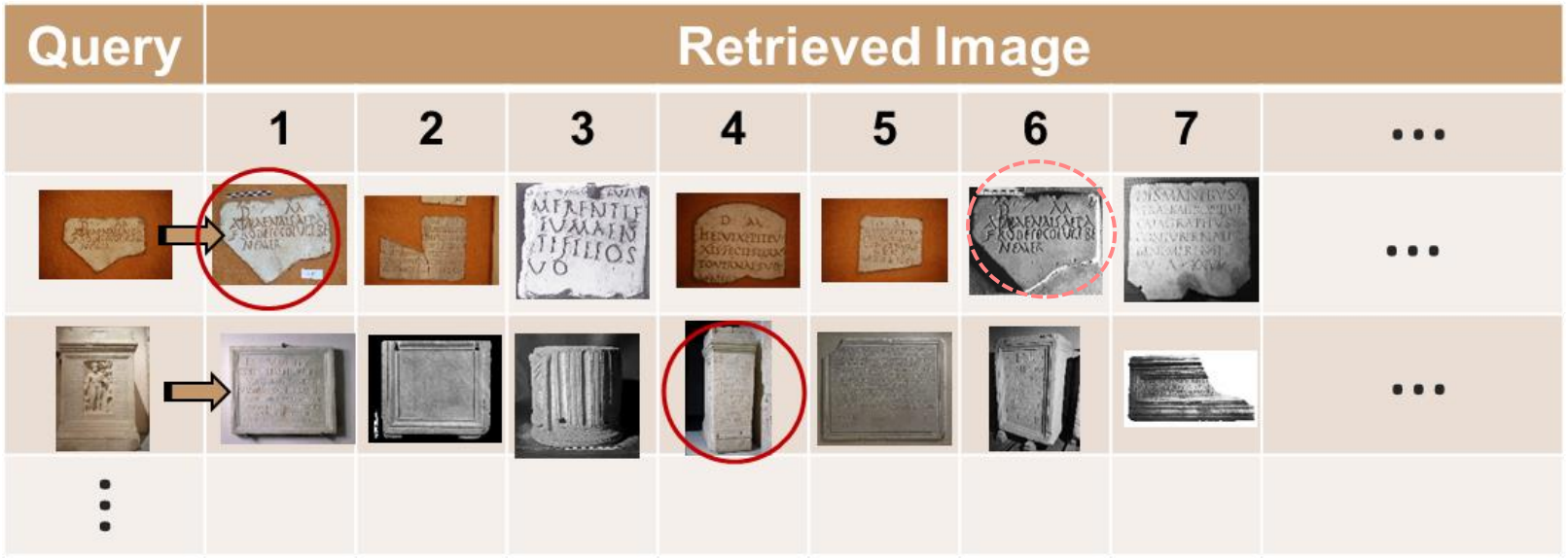
Ground Truth

- **70** queries selected from the whole dataset
  - represent various types of inscriptions
  - for each query we select the associated images using database consistency
  - only inscriptions that have more than one photo are used



Quality Measure

- To evaluate the quality of the results we use the probability ***p*** of finding an image of the same query object between the first r result ($1 \leq r \leq 100$)



Results

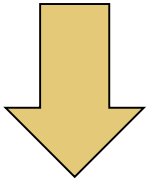
Approach	avg SIFTs	vocabulary size	Bytes	p r =1	p r =10	p r =100
VLAD	235	256	131,072	**.69**	.74	.84
BoF / cos TF-IDF	235	400,000	940	.64	**.76**	.87
VLAD	235	128	65,536	.64	.73	.87
BoF / cos TF-IDF	235	200,000	940	.60	.71	.81
VLAD	1591	256	131,072	.56	.71	**.90**
VLAD	1591	128	65,536	.56	.69	.87
BoF / cos TF-IDF	235	100,000	940	.56	.69	.79
VLAD	235	64	32,768	.53	.70	.86
VLAD	1591	64	32,768	.50	.61	.79
VLAD-PCA (d'=512)	1591	128	2,048	.44	.59	.79

- BoF approach obtained good results when a large vocabulary (400,000) was used
- The best accuracy was obtained by using the VLAD approach with a vocabulary size of 256

Conclusions

- We tested the state-of-the-art object recognition techniques on Epigraphic Database Roma
  - 17,155 photos related to 14,560 inscriptions

- The best accuracy was obtained by using the Vector of Locally Aggregated Descriptor (VLAD)
  - The obtained accuracy was of .69, which is good considering the difficulties of the task and the few images available for each inscription in the dataset



We plan to use VLAD approach in the official EAGLE inscription recognition engine

Thanks

Lucia Vadicamo
lucia.vadicamo@isti.cnr.it