

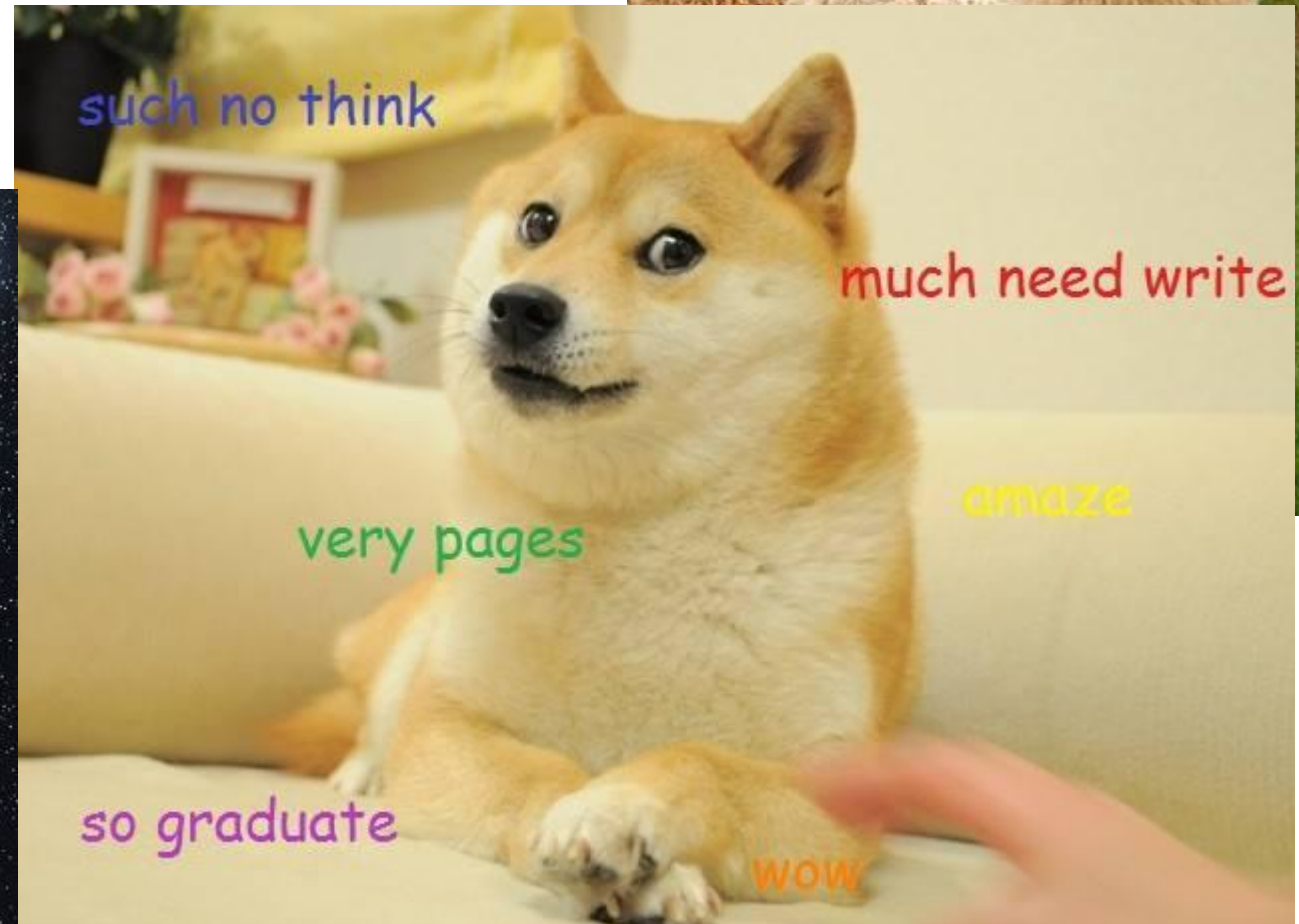
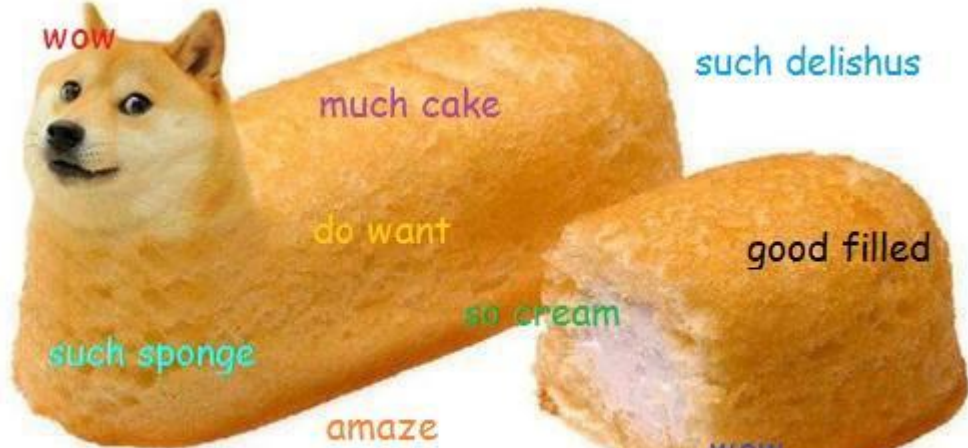
# PIMMI

a command line interface to study image propagation

Nicolas Hervé, research department, Institut National de l'Audiovisuel, France

Béatrice Mazoyer, médialab, Sciences Po, France

# How to study meme propagation?



## What PIMMI does

- create clusters of identical images (total or partial copies) on millions of files
- deal with image transformations (crop, zoom)
- adapt to corpus characteristics (number and nature of images)

## What PIMMI does not

- clustering of semantically similar images (e.g. separate cats from dogs)
- face recognition (e.g. find all images of Elizabeth II)

# Use cases

Pimmi was designed to conduct studies on the use and re-use of images:

- Propagation of memes on social networks
- Usage of press agency photos in a press corpus
- Dissemination of fake news based on image montages
- Comparison of editorial choices between different media
- ...



Cluster 0

20 img / 6



copydays\_img|peg|pegqual|10/200400.jpg

Cluster 7

42 img / 4



copydays\_img|peg|pegqual|10/201400.jpg

Cluster 14

20 img / 7



copydays\_img|peg|pegqual|10/202200.jpg

Cluster 21

19 img / 9

Cluster 1

19 img / 6



copydays\_img|peg|pegqual|10/200100.jpg

Cluster 8

19 img / 12



copydays\_img|peg|pegqual|10/200800.jpg

Cluster 15

18 img / 7



copydays\_img|peg|pegqual|10/202400.jpg

Cluster 22

15 img / 8

Cluster 2

18 img / 10



copydays\_img|peg|pegqual|10/200700.jpg

Cluster 9

20 img / 10



copydays\_img|peg|pegqual|10/201500.jpg

Cluster 16

20 img / 7



copydays\_img|peg|pegqual|10/202300.jpg

Cluster 23

20 img / 10

Cluster 3

20 img / 9



copydays\_img|peg|pegqual|10/201300.jpg

Cluster 10

19 img / 6



copydays\_img|peg|pegqual|10/201800.jpg

Cluster 17

20 img / 9



copydays\_img|peg|pegqual|10/201600.jpg

Cluster 24

19 img / 7

Cluster 4

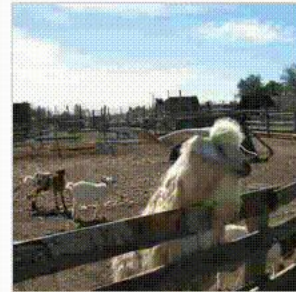
18 img / 6



copydays\_img|peg|pegqual|10/201100.jpg

Cluster 11

20 img / 8



copydays\_img|peg|pegqual|10/202100.jpg

Cluster 18

18 img / 11



copydays\_img|peg|pegqual|10/201700.jpg

Cluster 25

20 img / 7

Cluster 5

18 img / 7



copydays\_img|peg|pegqual|10/201200.jpg

Cluster 12

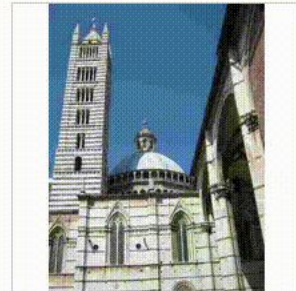
20 img / 10



copydays\_img|peg|pegqual|10/200200.jpg

Cluster 19

495 img / 1



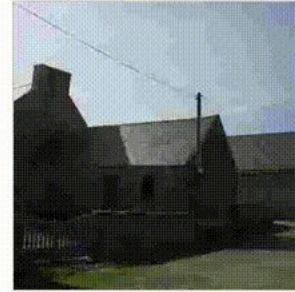
copydays\_img|peg|pegqual|10/202000.jpg

Cluster 26

19 img / 10

Cluster 6

17 img / 8



copydays\_img|peg|pegqual|10/201000.jpg

Cluster 13

7 img / 11



copydays\_img|peg|pegqual|10/201900.jpg

Cluster 20

21 img / 8



copydays\_img|peg|pegqual|10/202000.jpg

Cluster 27

13 img / 9



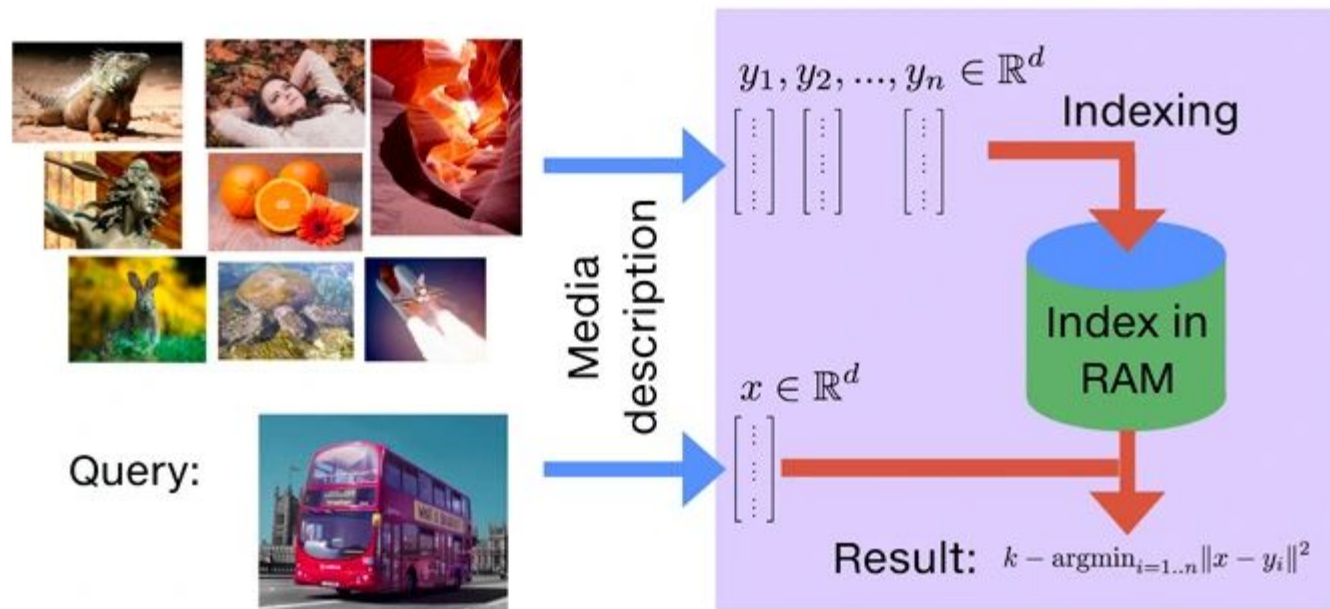
# How it works

Each image is represented as a set of keypoints using SIFT (Scale Invariant Feature Transform) algorithm. Local keypoints descriptors are vectors.



# How it works

Local descriptors are indexed in a database (FAISS) optimized for similarity search. Different index structures are available, depending on corpus size.





# How it works

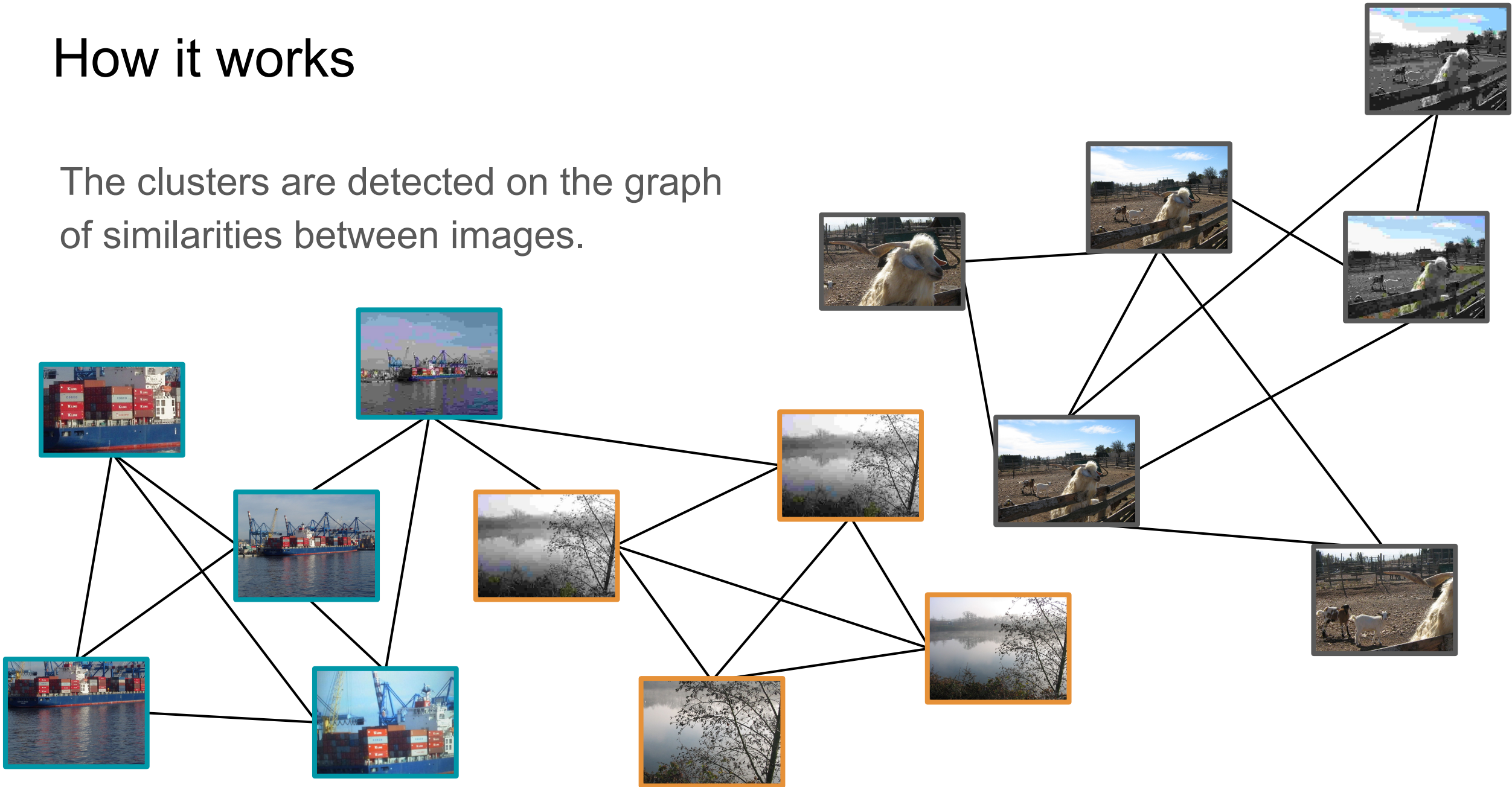
PIMMI searches for matching local descriptors in the database.

For each pair of images who have matches, it filters only pairs compatible with some expected geometric transformations.



# How it works

The clusters are detected on the graph of similarities between images.





# How to use PIMMI

What you need:

- Python  $\geq$  3.7
- Images in .jpg or .png format saved on your computer
- A command-line shell

```
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$ python --version
```

```
Python 3.8.2
```

```
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$ pими fill small_dataset/ demo
```

```
2023-02-03 19:07:51,216 : INFO : pими - listing images recursively from : small_dataset/
```

```
2023-02-03 19:07:51,217 : INFO : pими - Using opencv : 4.7.0
```

```
2023-02-03 19:07:51,231 : INFO : pими - nb tasks launched : 11 / 25000
```

```
2023-02-03 19:07:51,342 : INFO : pими - ~ [0] (1000, 1, 0.1, 10, 1.6) extracting 375 sift points for (512 x 341, False)
small_dataset/000001.jpg
```

```
2023-02-03 19:07:51,344 : INFO : pими - ~ [0] result retrieved from queue : 375
```

```
2023-02-03 19:07:51,398 : INFO : pими - ----- Training IVF1024,Flat index on 1598 features
```

```
Training level-1 quantizer
```

```
Training level-1 quantizer on 1598 vectors in 128D
```

```
WARNING clustering 1598 points to 1024 centroids: please provide at least 39936 training points
```

```
Training IVF residual
```

```
IndexIVF: no residual training
```

```
2023-02-03 19:07:51,664 : INFO : pими - ----- Adding 1598 features
```

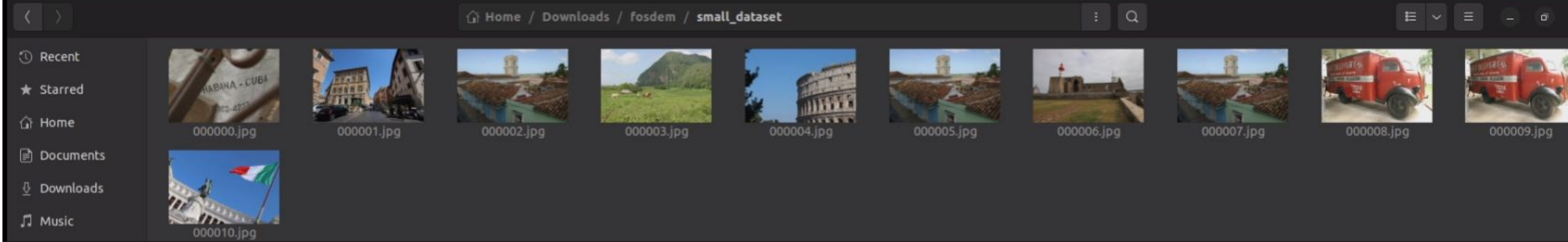
```
IndexIVFFlat::add_core: added 1598 / 1598 vectors
```

```
2023-02-03 19:07:51,695 : INFO : pими - index has 11 images with 1598 feature points
```

```
2023-02-03 19:07:51,696 : INFO : pими - index saved ./index/demo.IVF1024,Flat.faiss / ./index/demo.IVF1024,Flat.meta
```

```
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$
```

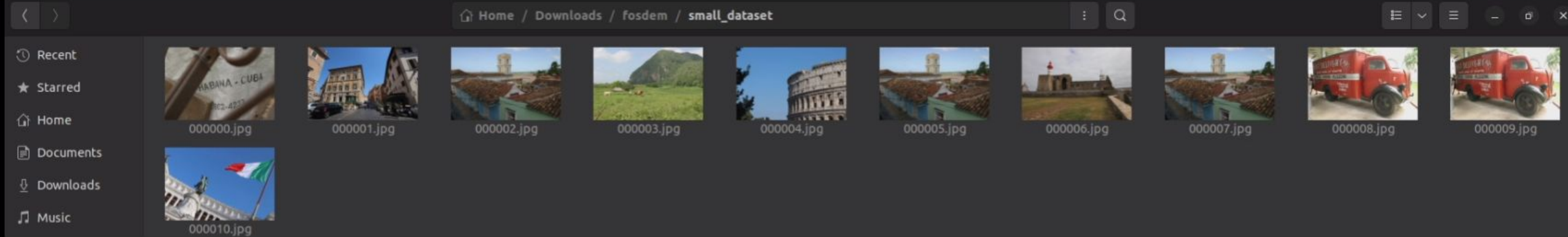
# Demo





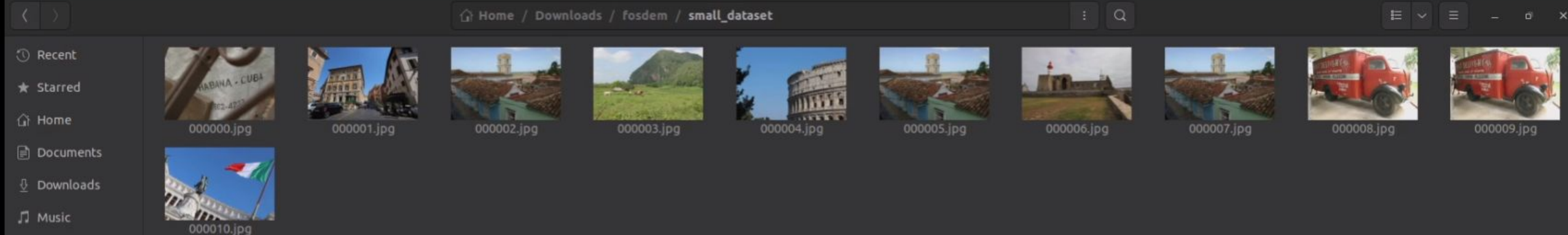
```
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$ pimmi query --simple small_dataset/ demo
2023-02-03 19:08:02,095 : INFO : pimmi - index loaded ./index/demo.IVF1024,Flat.faiss
2023-02-03 19:08:02,096 : INFO : pimmi - meta loaded ./index/demo.IVF1024,Flat.meta
2023-02-03 19:08:02,096 : INFO : pimmi - index has 11/11 images with 1598 feature points
2023-02-03 19:08:02,096 : INFO : pimmi - - type : IVF1024,Flat
2023-02-03 19:08:02,096 : INFO : pimmi - found 11 images in index
2023-02-03 19:08:02,096 : INFO : pimmi - total number of queries 11
2023-02-03 19:08:02,096 : INFO : pimmi - query 11 files from pack 0 -> ./index/demo.IVF1024,Flat.mining_000000.csv
2023-02-03 19:08:02,097 : INFO : pimmi - Using opencv : 4.7.0
2023-02-03 19:08:02,231 : INFO : pimmi - ~ [0 :: 0] (1000, 1, 0.1, 10, 1.6) extracting 375 sift points for (512 x 341, False) small_dataset//000001.jpg
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$
```

# Demo



```
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$ pimmi clusters demo > demo_clusters.csv
2023-02-03 19:08:23,522 : INFO : pimmi - Loading query results
2023-02-03 19:08:23,523 : INFO : pimmi - Number of vertices in the graph: 5
2023-02-03 19:08:23,523 : INFO : pimmi - Number of edges in the graph: 8
2023-02-03 19:08:23,523 : INFO : pimmi - Is the graph directed: 1
2023-02-03 19:08:23,523 : INFO : pimmi - Maximum degree in the graph: 4
2023-02-03 19:08:23,523 : INFO : pimmi - Connected components in the graph: 2
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$ xsv table demo_clusters.csv
path          image_id  nb_points  degree  cluster_id  quality
000002.jpg    7         66         4        0           0.5
000005.jpg    1         66         4        0           0.5
000007.jpg    2         66         4        0           0.5
000008.jpg    10        133        2        1           0.16296296296296298
000009.jpg    5         139        2        1           0.16296296296296298
(fosdem) bmazoyer@ptl11206:~/Downloads/fosdem$
```

# Demo





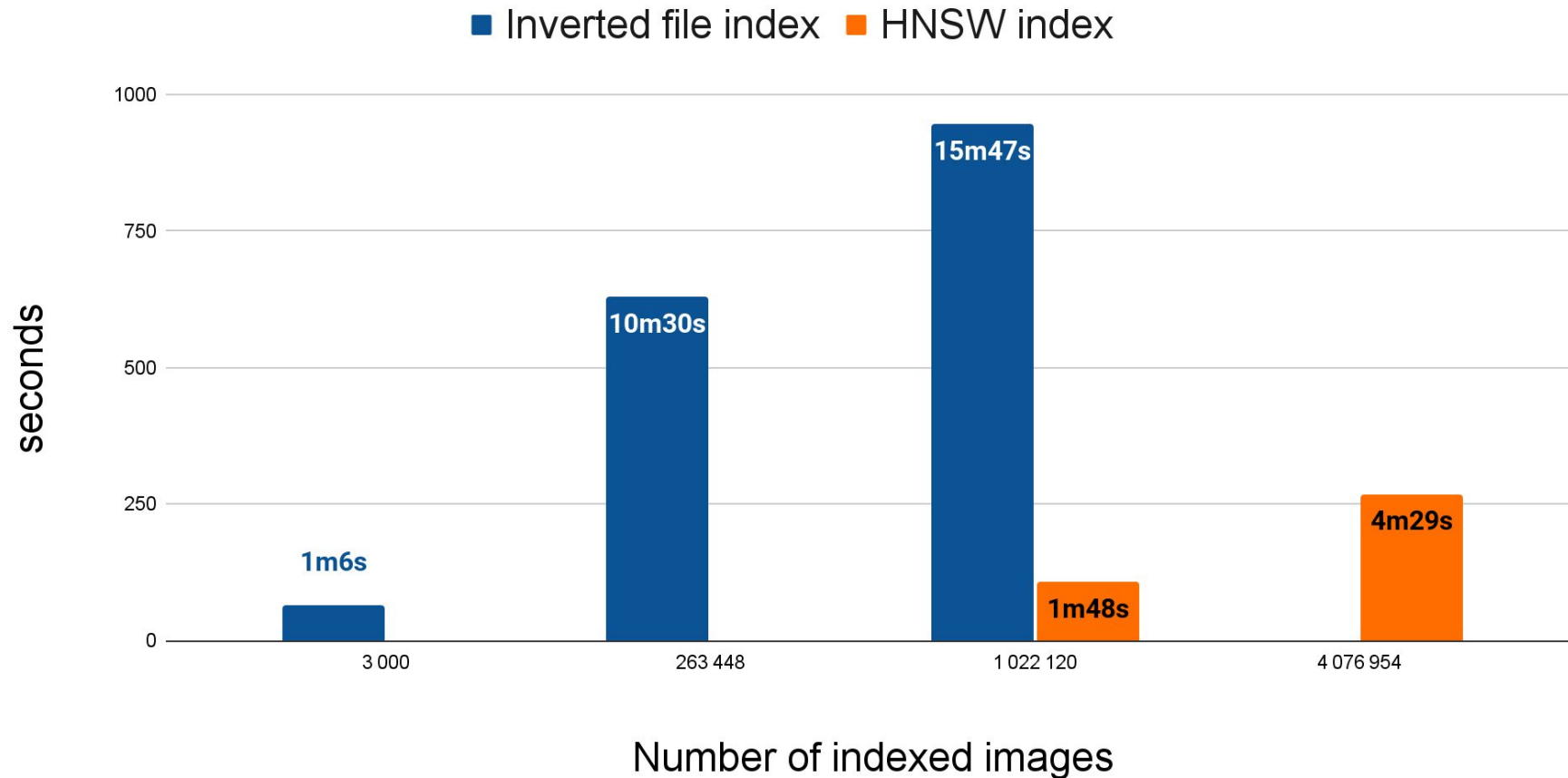
# How long does it take?

It depends on:

- the number of indexed images
- the chosen index
- the number of cores

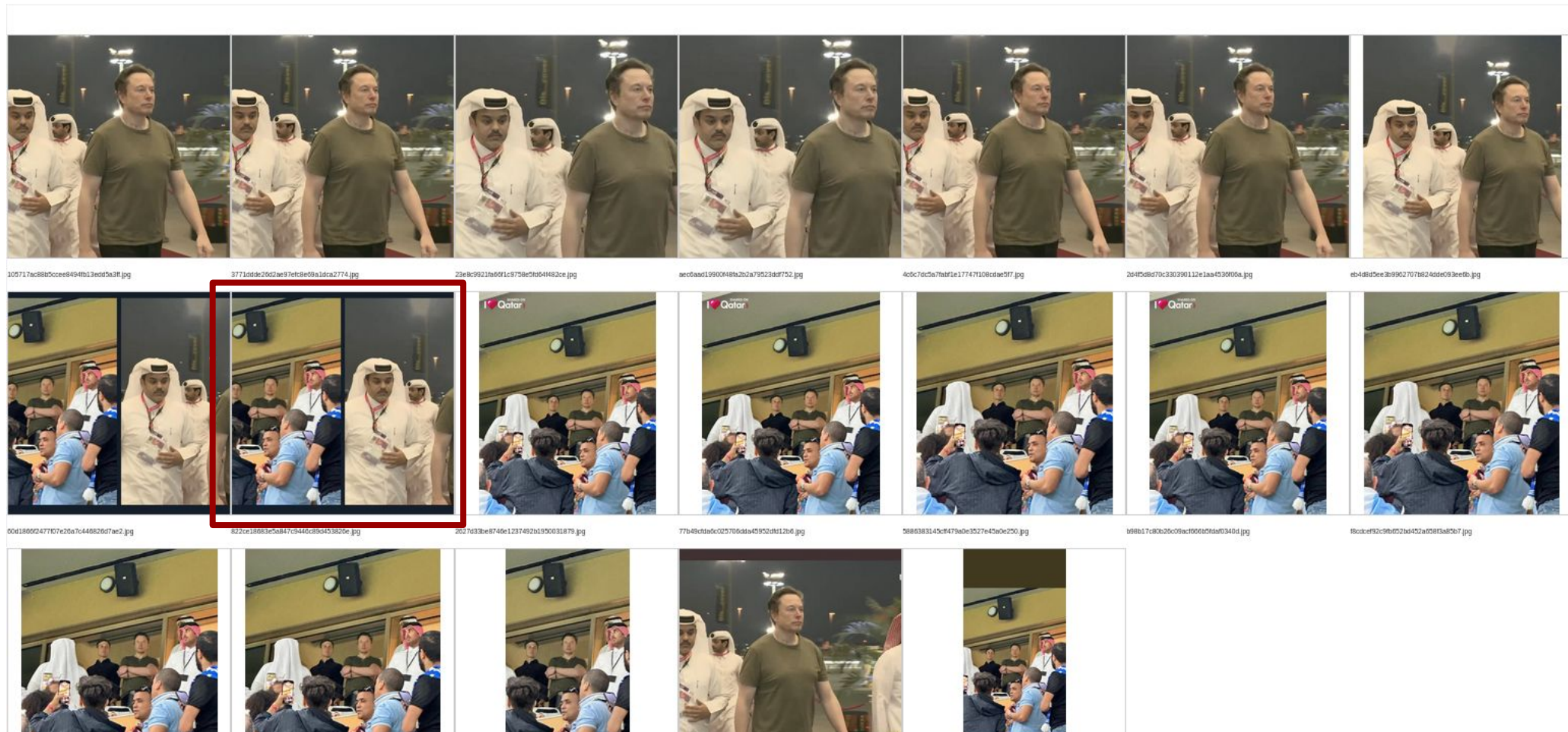
## Query time for 3000 queries

24 cores, 256 Go RAM



# Future improvements

Detect parts of images - should solve text in images issues





# Future improvements

Detect parts of images

Show images in their context (tweets, instagram posts, etc.) or with additional metadata

Display the graph of image similarities

# We need your use cases!

The development of PIMMI is still in progress

We need new use cases to improve the tool



# Sources and references



Images

doge: <https://knowyourmeme.com/memes/doge>



small dataset: [https://github.com/nrv/pimmi/tree/main/demo\\_dataset/small\\_dataset](https://github.com/nrv/pimmi/tree/main/demo_dataset/small_dataset)

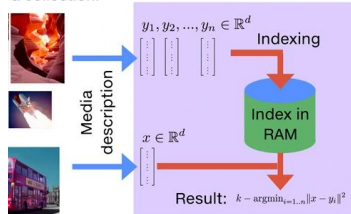


copydays (INRIA):

<https://lear.inrialpes.fr/~jegou/data.php#copydays>

[http://web.archive.org/web/20181015092553if\\_/http://pascal.inrialpes.fr/data/holidays/](http://web.archive.org/web/20181015092553if_/http://pascal.inrialpes.fr/data/holidays/)

a collection:



faiss:

<https://engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/>

# Sources and references

## Papers

SIFT: Lowe, David G. "Object recognition from local scale-invariant features." Proceedings of the seventh IEEE international conference on computer vision. Vol. 2. Ieee, 1999.

FAISS: Jégou, Hervé, et al. "Faiss: Similarity search and clustering of dense vectors library." Astrophysics Source Code Library (2022): ascl-2210.