Visual Embedding and Visual Search

Liangliang Cao

https://columbia6894.github.io/

Why Visual Search Is Important?

Applications demanded by the increasing amount of images/videos:

- Diving into personal albums
- Recommending Youtube/news/TV shows
- Searching clothes and fashion products
- Organizing social media

Visual search goes beyond the limits of visual classification:

- Unlimited amount of categories
- Easily integrated with other systems.

Bear in mind that searching is complicated… we will discuss one example built from scratch.

Outline

Visual search based on local descriptors (SIFT)

Visual search based on deep embedding

- FaceNet
- Pinterest Visual Search

Visual search based on local descriptors

Key reference:

Scalable recognition with a **vocabulary tree**

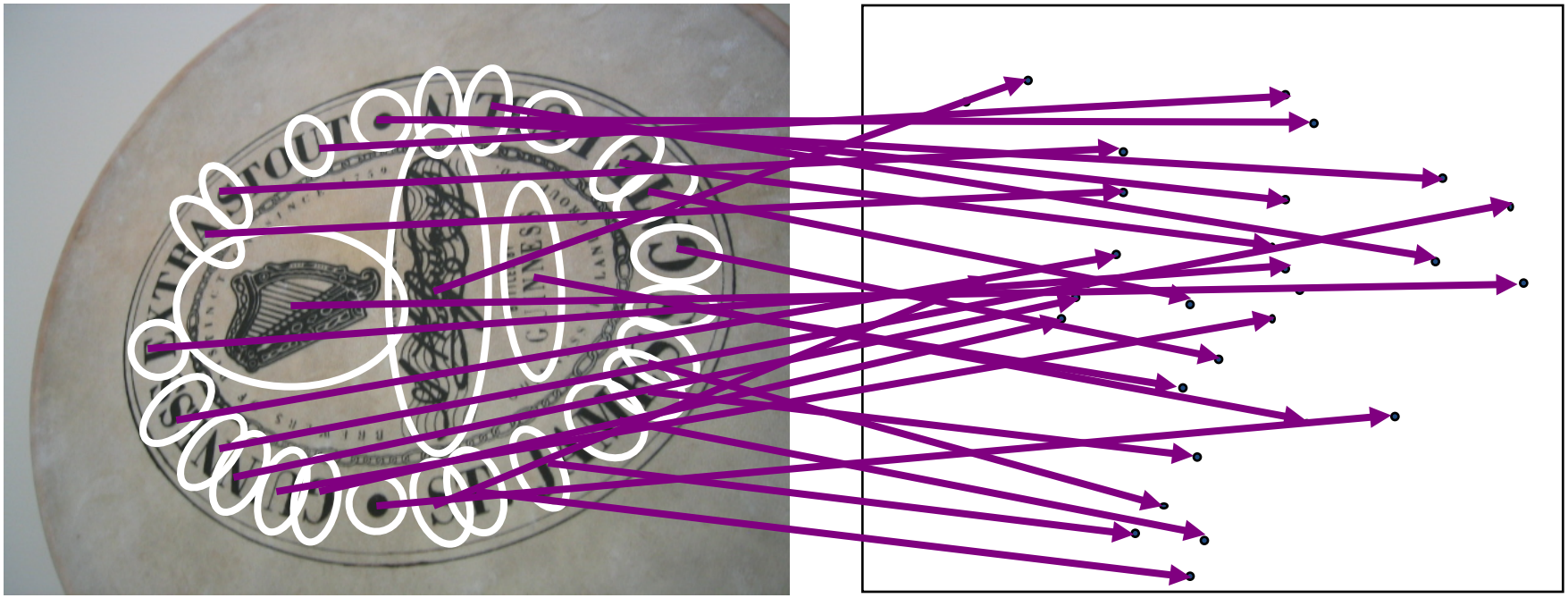
D Nister, H Stewenius - … vision and pattern **recognition**, 2006 …, 2006 - ieeexplore.ieee.org

A recognition scheme that scales efficiently to a large number of objects is presented. The efficiency and quality is exhibited in a live demonstration that recognizes CD-covers from a database of 40000 images of popular music CD's. The scheme builds upon popular …

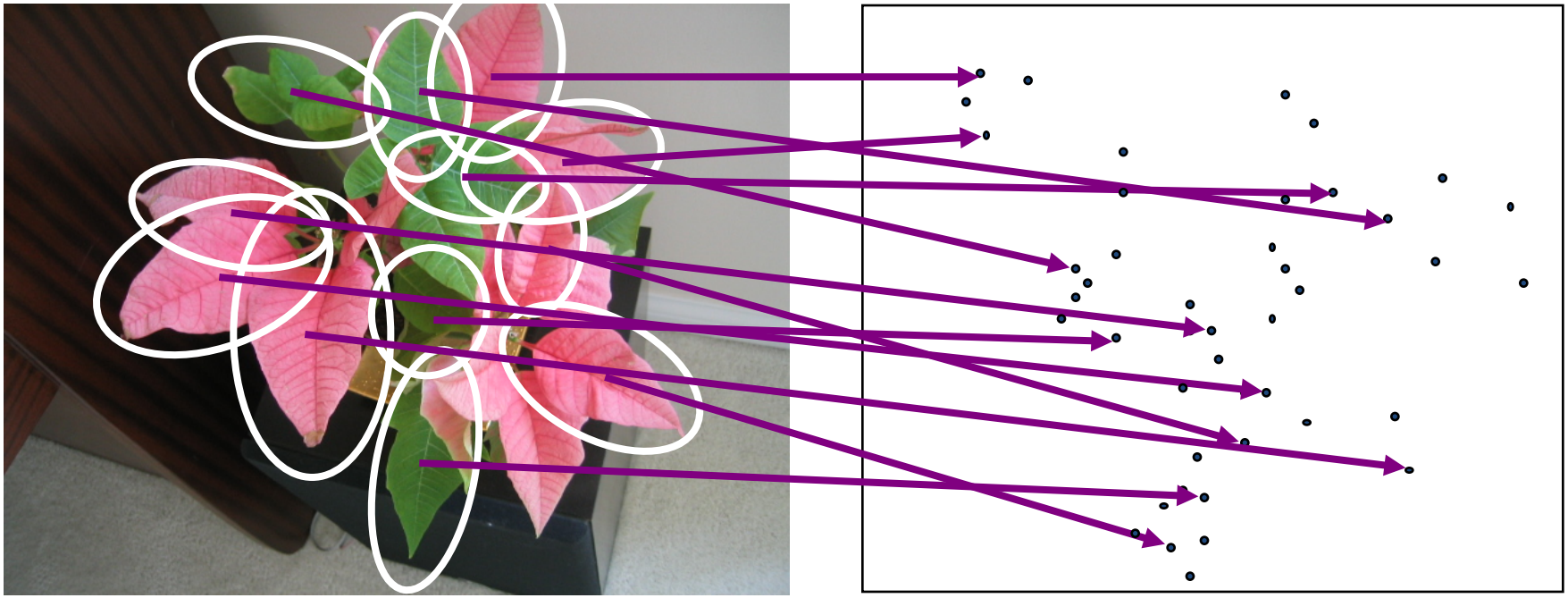
★ ❞ Cited by 3926 Related articles All 36 versions

Most of the following slides to David Nistér and Henrik Stewénius

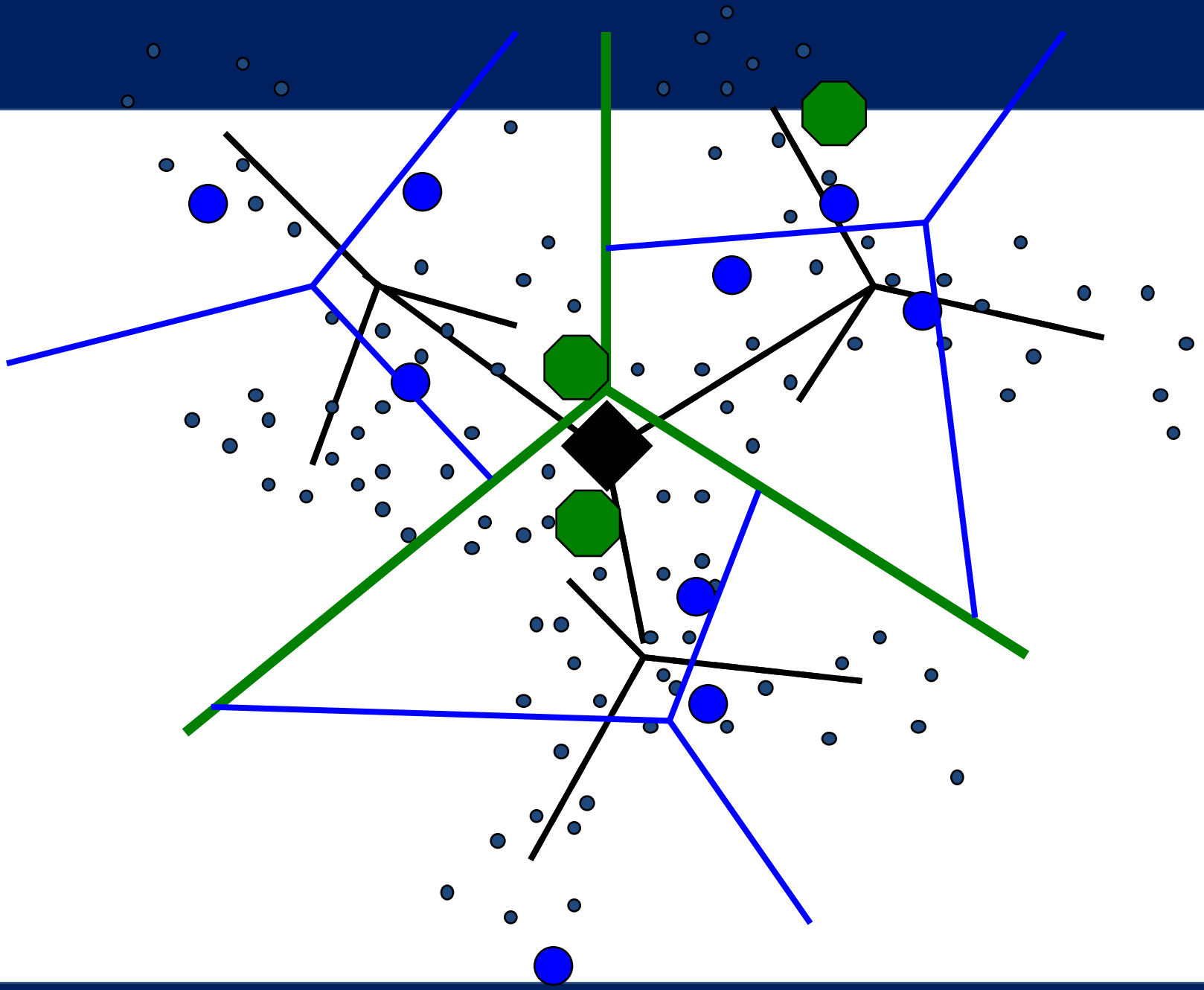
Visual search based on local descriptors



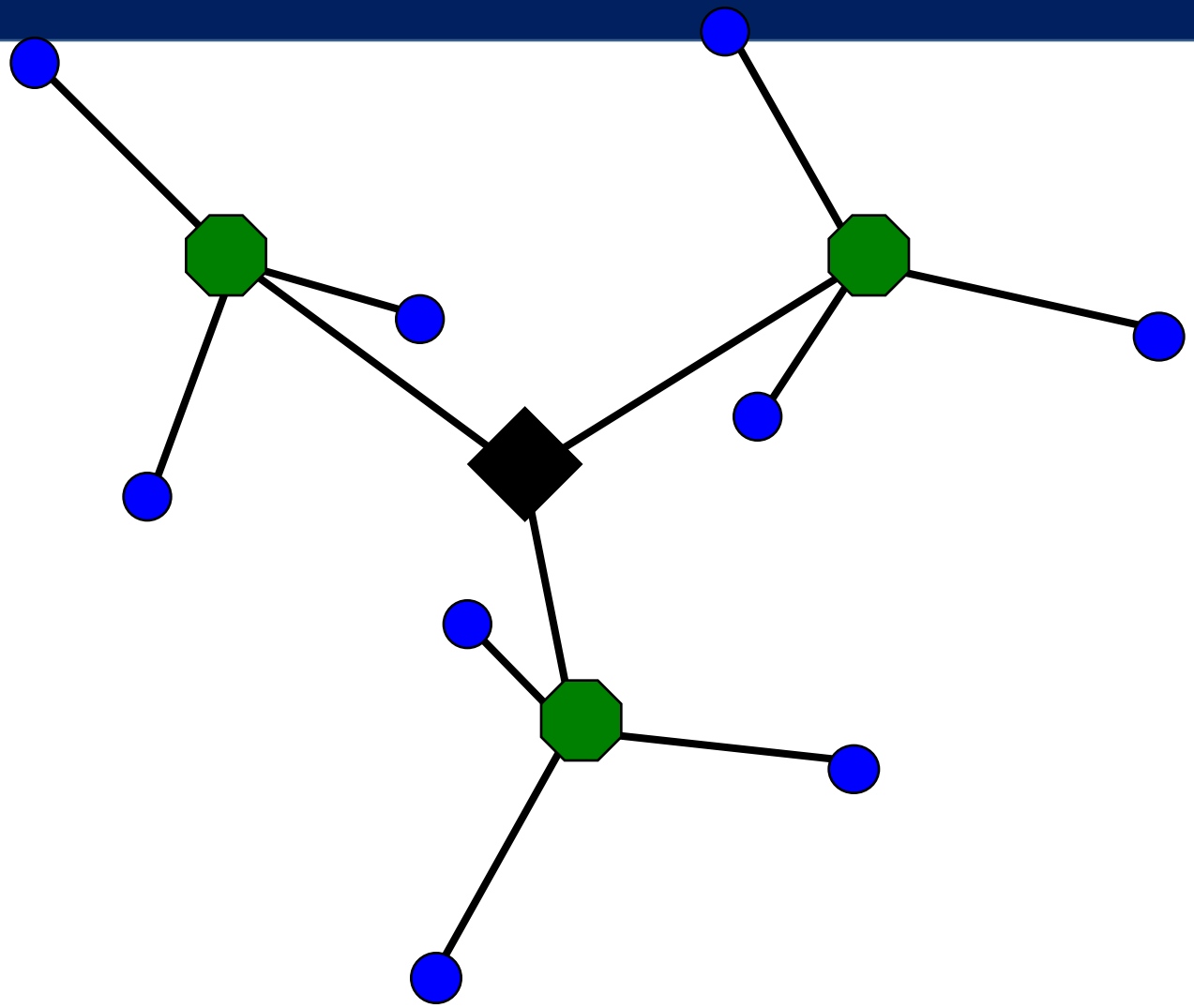
Visual search based on local descriptors



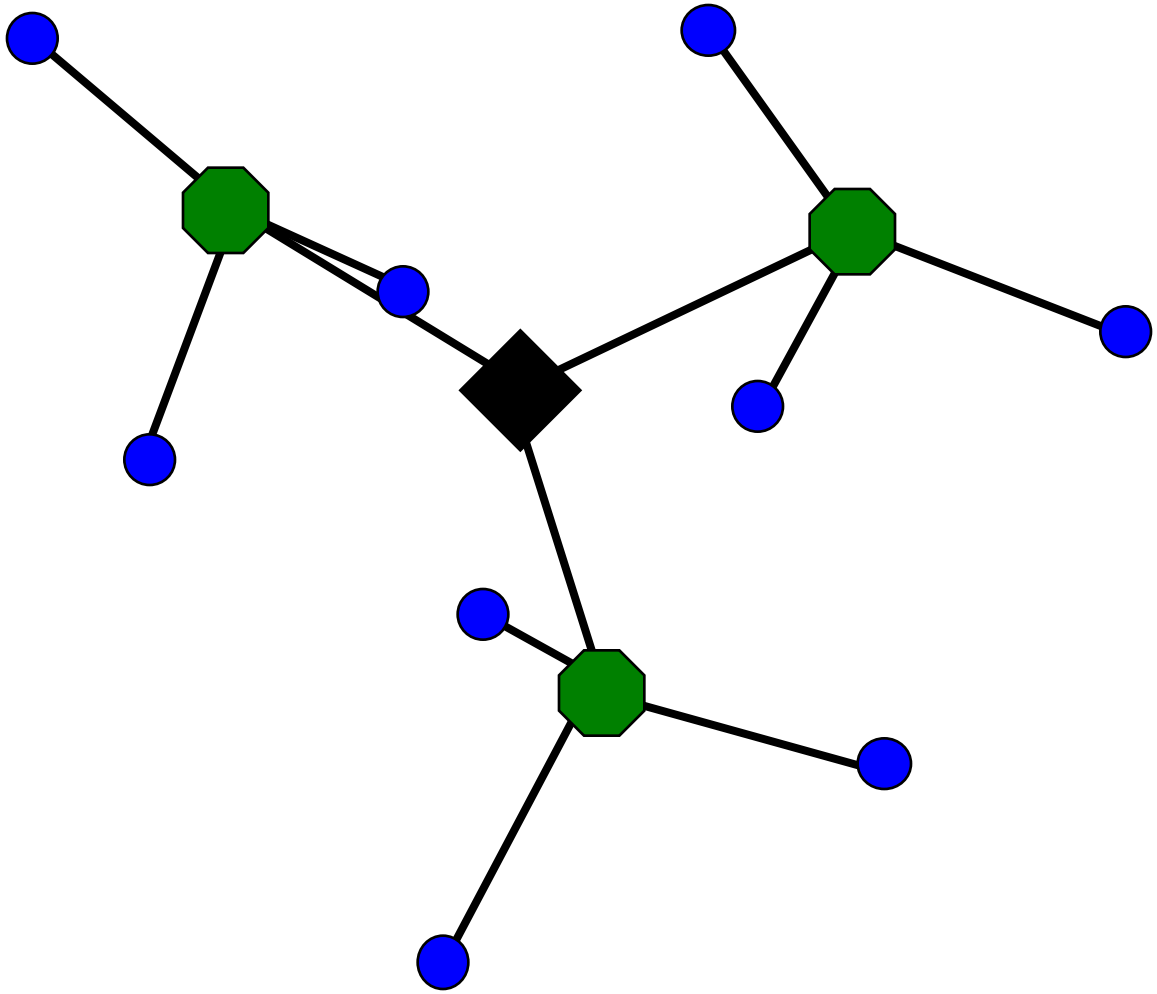
Visual search based on local descriptors



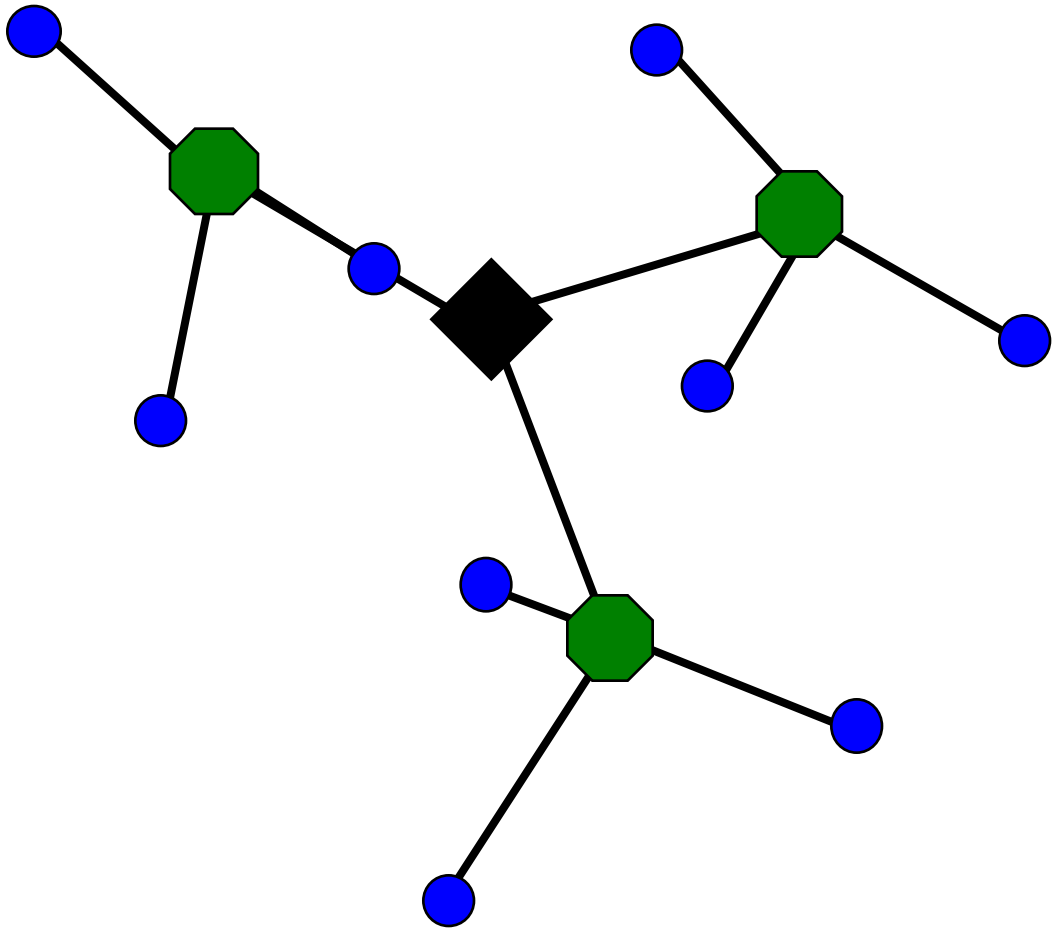
Visual search based on local descriptors



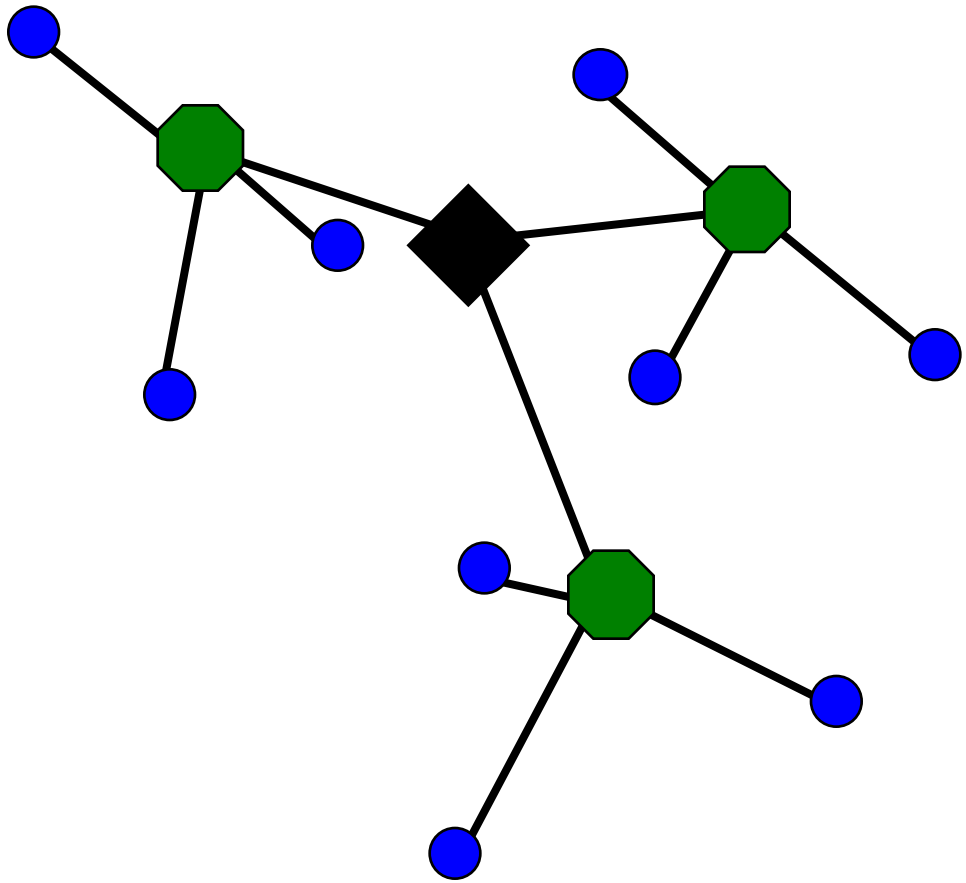
Visual search based on local descriptors



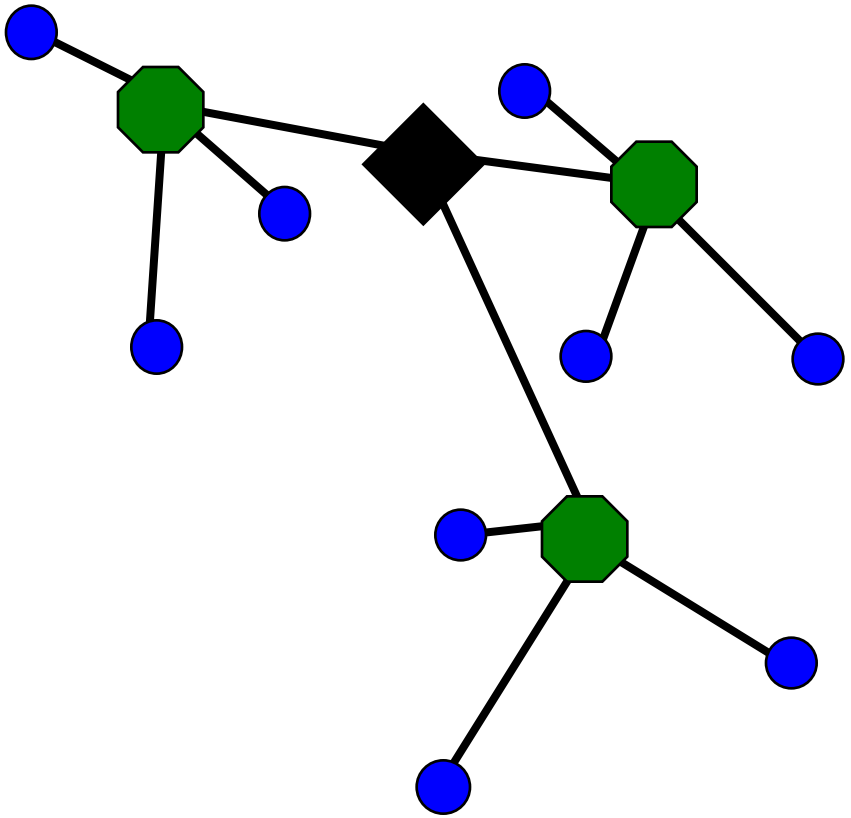
Visual search based on local descriptors



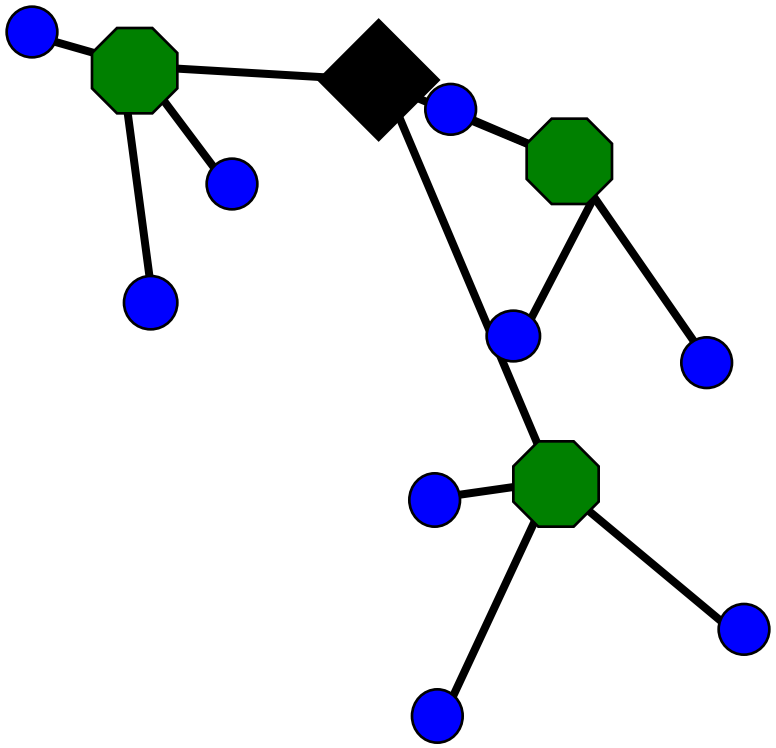
Visual search based on local descriptors



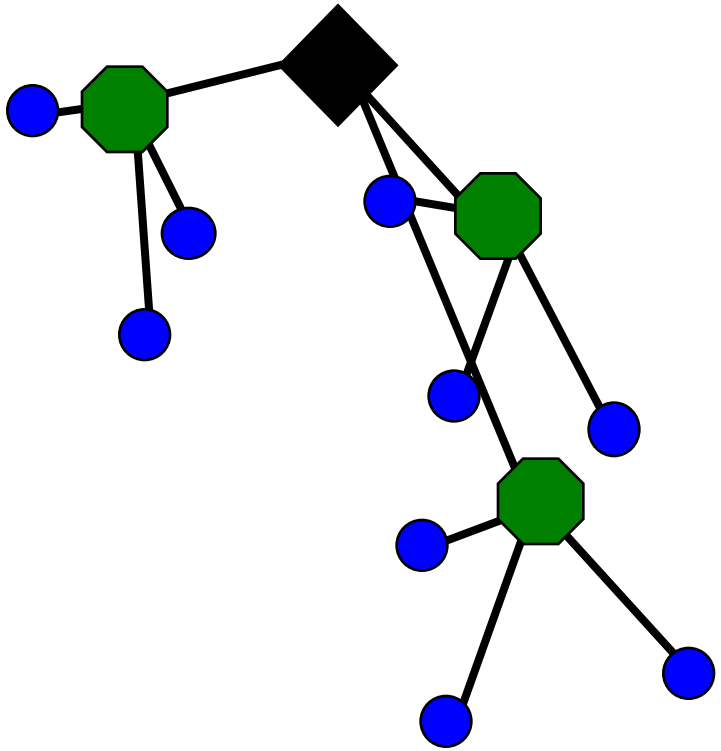
Visual search based on local descriptors



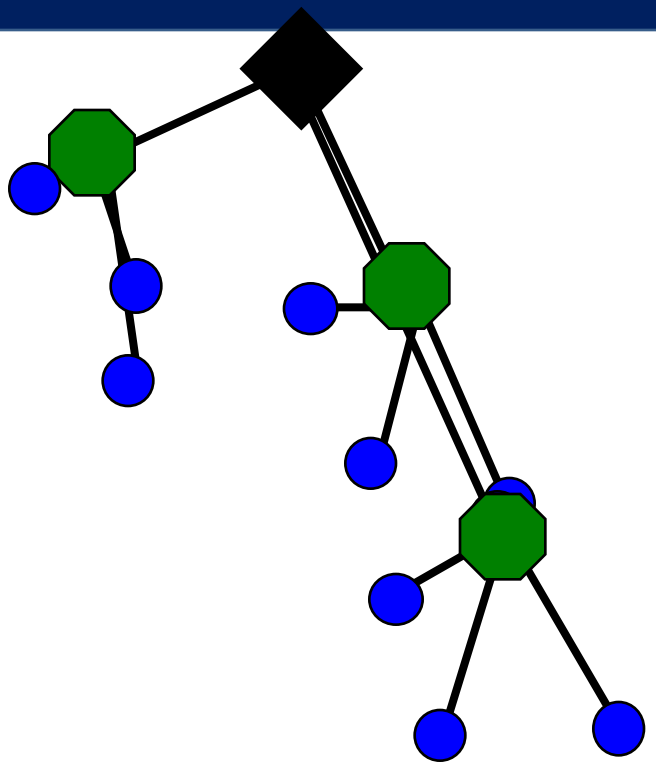
Visual search based on local descriptors



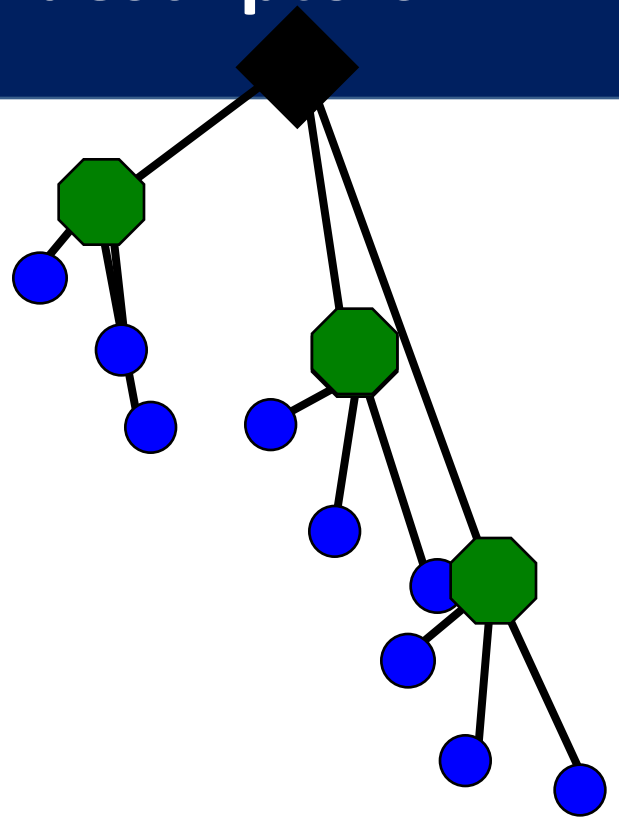
Visual search based on local descriptors

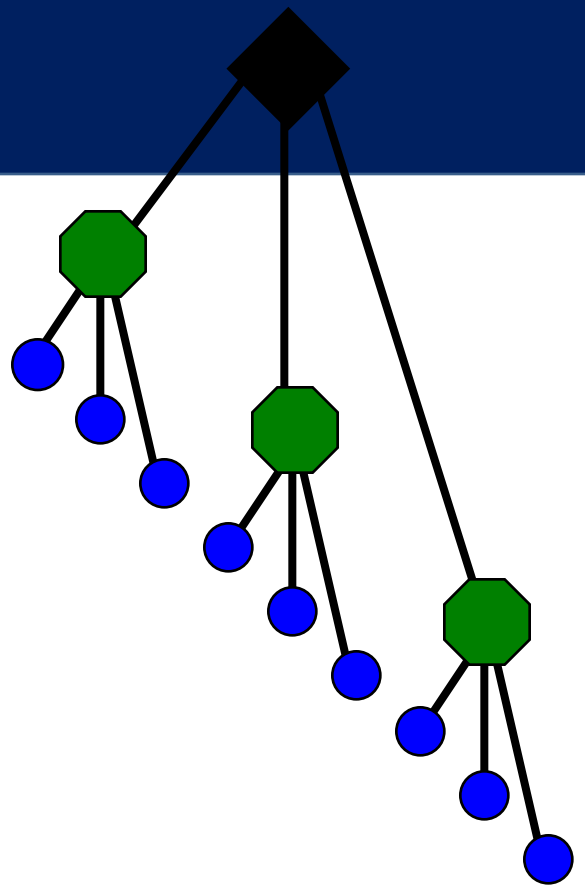


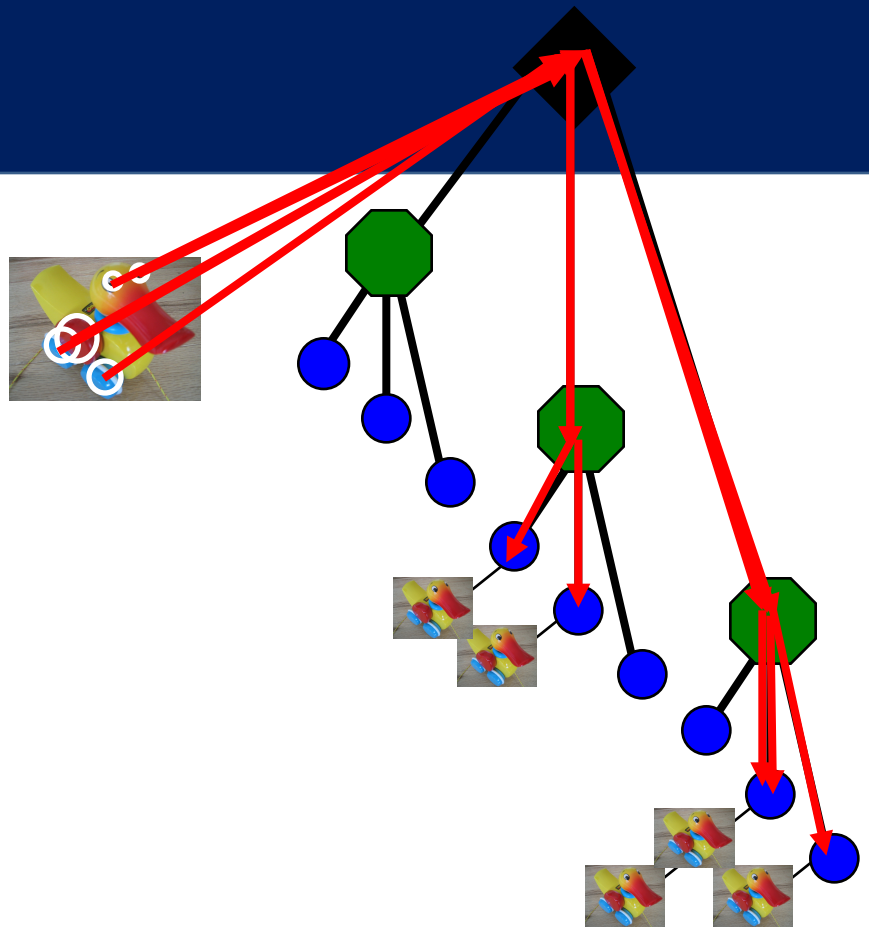
Visual search based on local descriptors

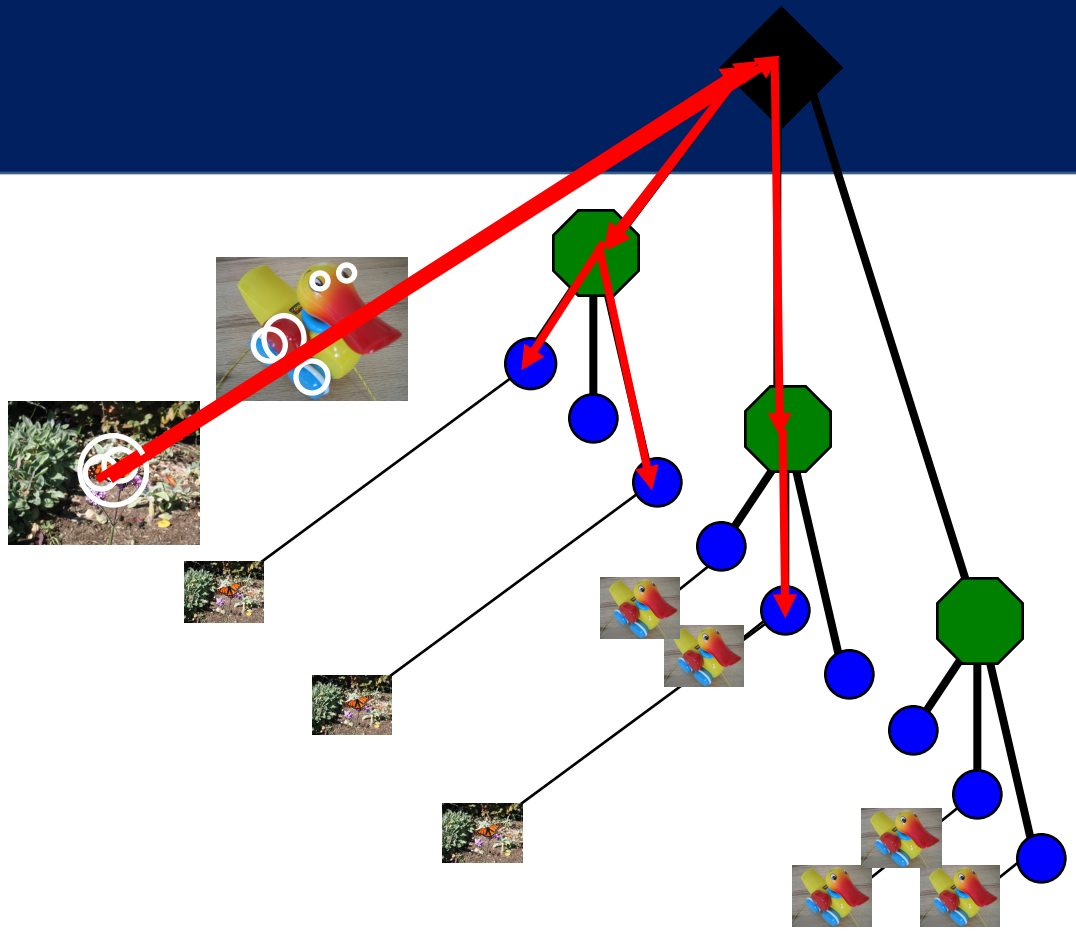


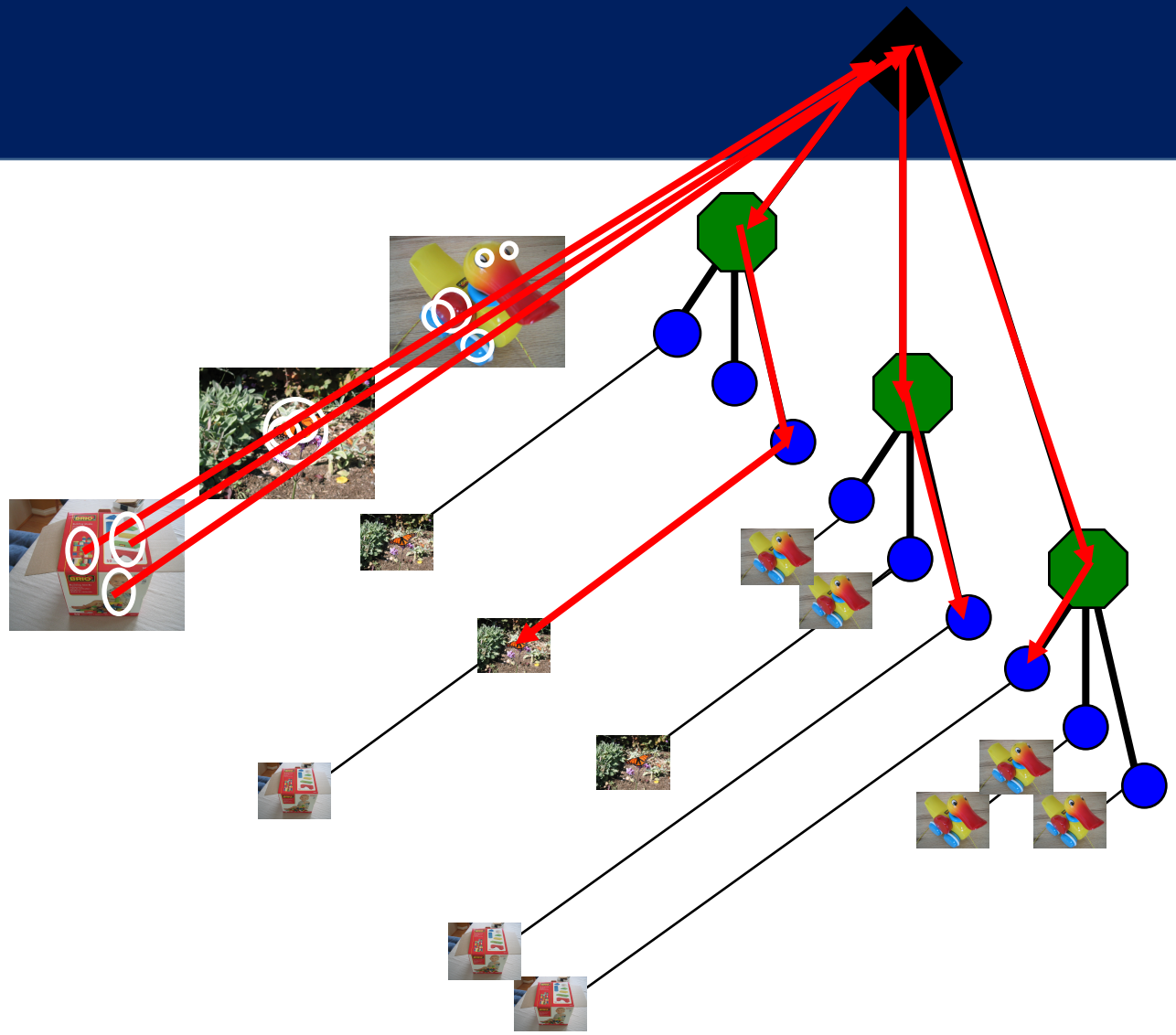
Visual search based on local descriptors

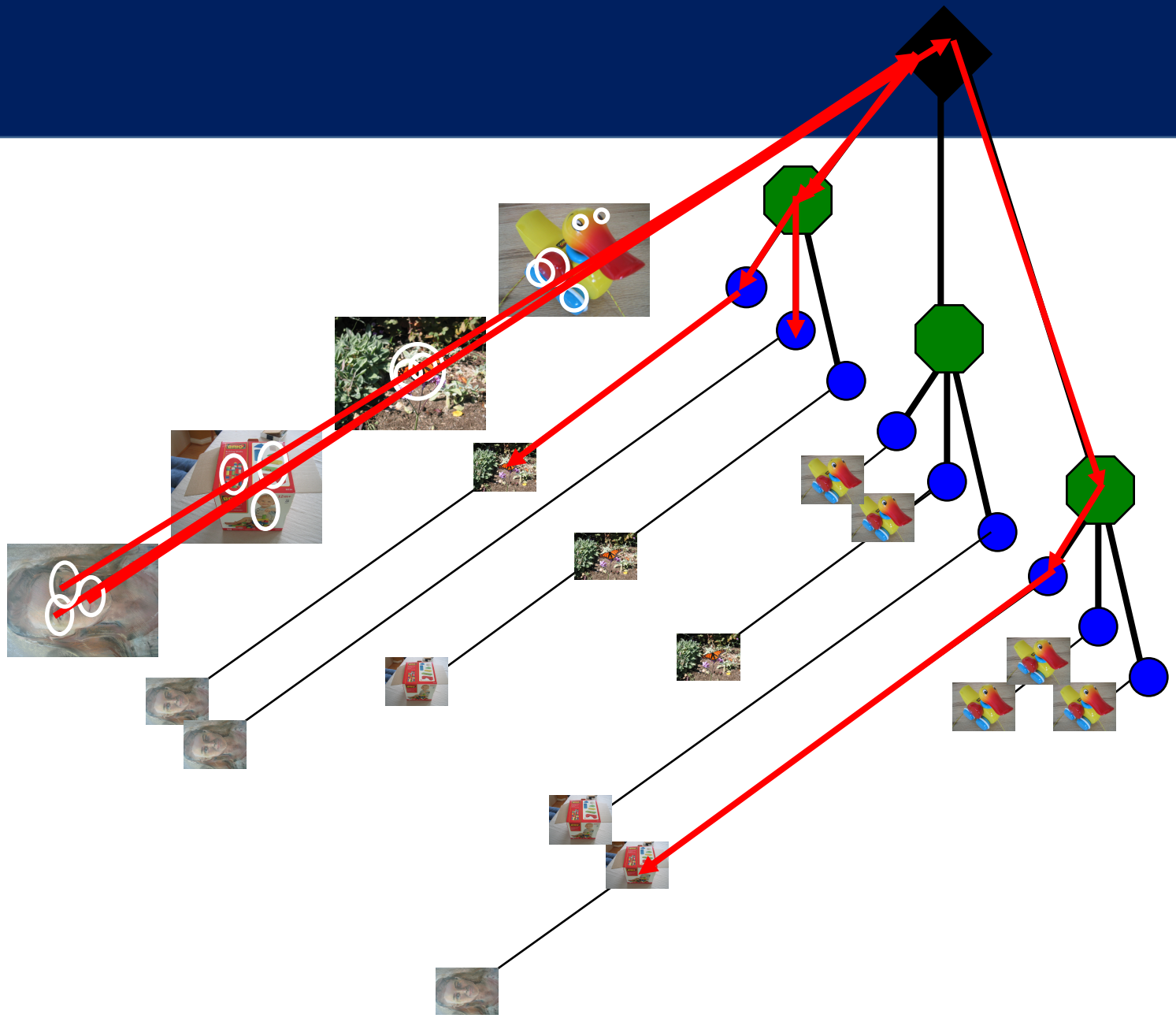


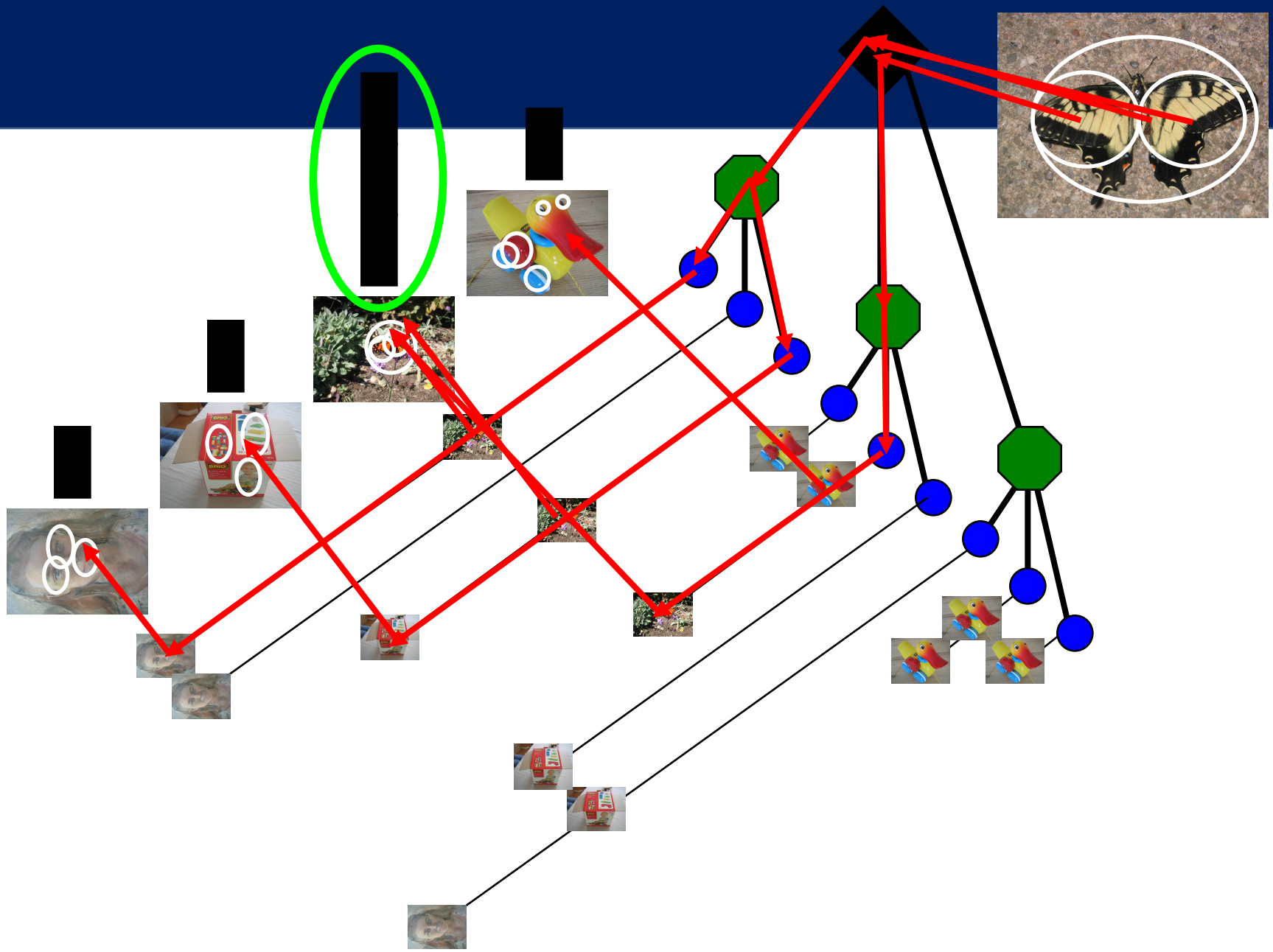




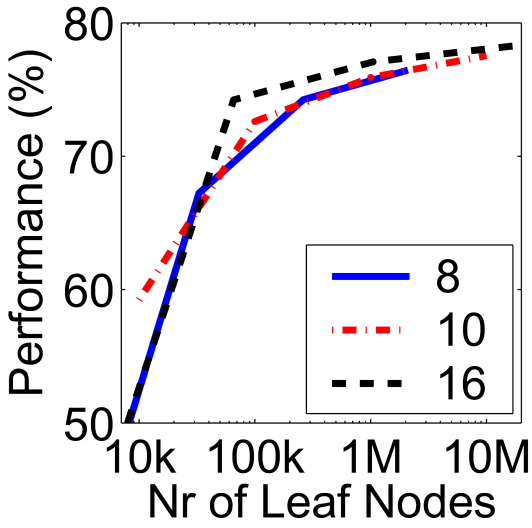


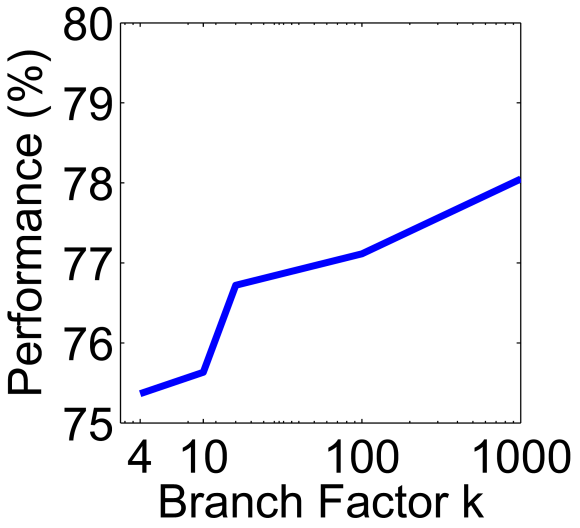






Search Results





Further Improvement

- Learning better rank functions
- Add spatial verification

Lessons

Search using local descriptors is good at finding EXACT objects, but not SIMILAR ones.

- No end to end learning
- No semantic similarities

But we have seen the power of

- Hierarchical tress (relevant open source toolkit: flann)
- Inverted index (relevant open source toolkit: lucene)

We shall use more powerful feature representation than local descriptors

Visual Search based on Deep Embedding

We want to show you the basic steps to build a visual search systems from scratch by reviewing two papers:

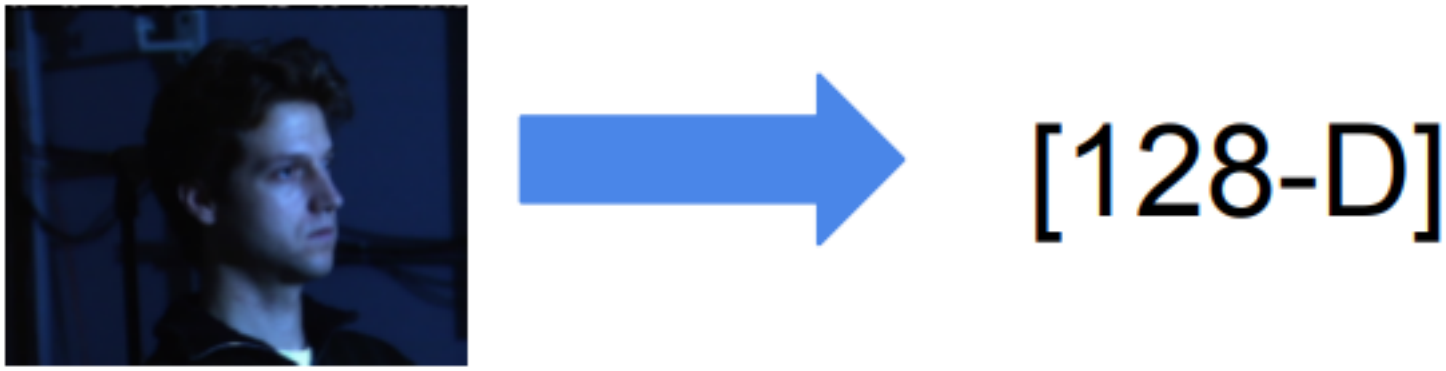
- FaceNet

- Visual Search at Pinterest

Face Embedding

Idea:

Map images to a compact Euclidean space, where distances correspond to face similarity



FaceNet: A Unified Embedding for Face Recognition and Clustering by F. Schroff et al, CVPR'15

Face Embedding

Learn embedding:

- Learned from triple loss (by origin FaceNet paper, harder to tune)
- Or learned from classification tasks (simpler in practice)

Experiment:

1. Detect 100M-200M faces of 8M identities
2. Learn embedding of 128d vectors
3. A simple nearest neighbor search gets 98.87% on LFW face dataset

Key takeaways

- We can train neural embedding for visual search!

Face Embedding

Learn embedding:

- Learned from triple loss (by origin FaceNet paper, harder to tune)
- Or learned from classification tasks (simpler in practice)

What if we do not have so many samples?

Experiment:

1. Detect 100M-200M faces of 8M identities
2. Learn embedding of 128d vectors
3. A simple nearest neighbor search gets 98.87% on LFW face dataset

Key takeaways

- We can train neural embedding for visual search!

Face Embedding

Learn embedding:

- Learned from triple loss (by origin FaceNet paper, harder to tune)
- Or learned from classification tasks (simpler in practice)

Experiment:

1. Detect 100M-200M faces of 8M identities
2. Learn embedding of 128d vectors
3. A simple nearest neighbor search gets 98.87% on LFW face dataset

What if we do not have so many samples?

Try pre-trained models from ImageNet/Celeb1M

Key takeaways

- We can train neural embedding for visual search!

Extend Face Search to General Image Search

Represent an image with multiple clues

- Embedding of the whole image
- Embedding of detected regions*
- Annotations, categories, etc.

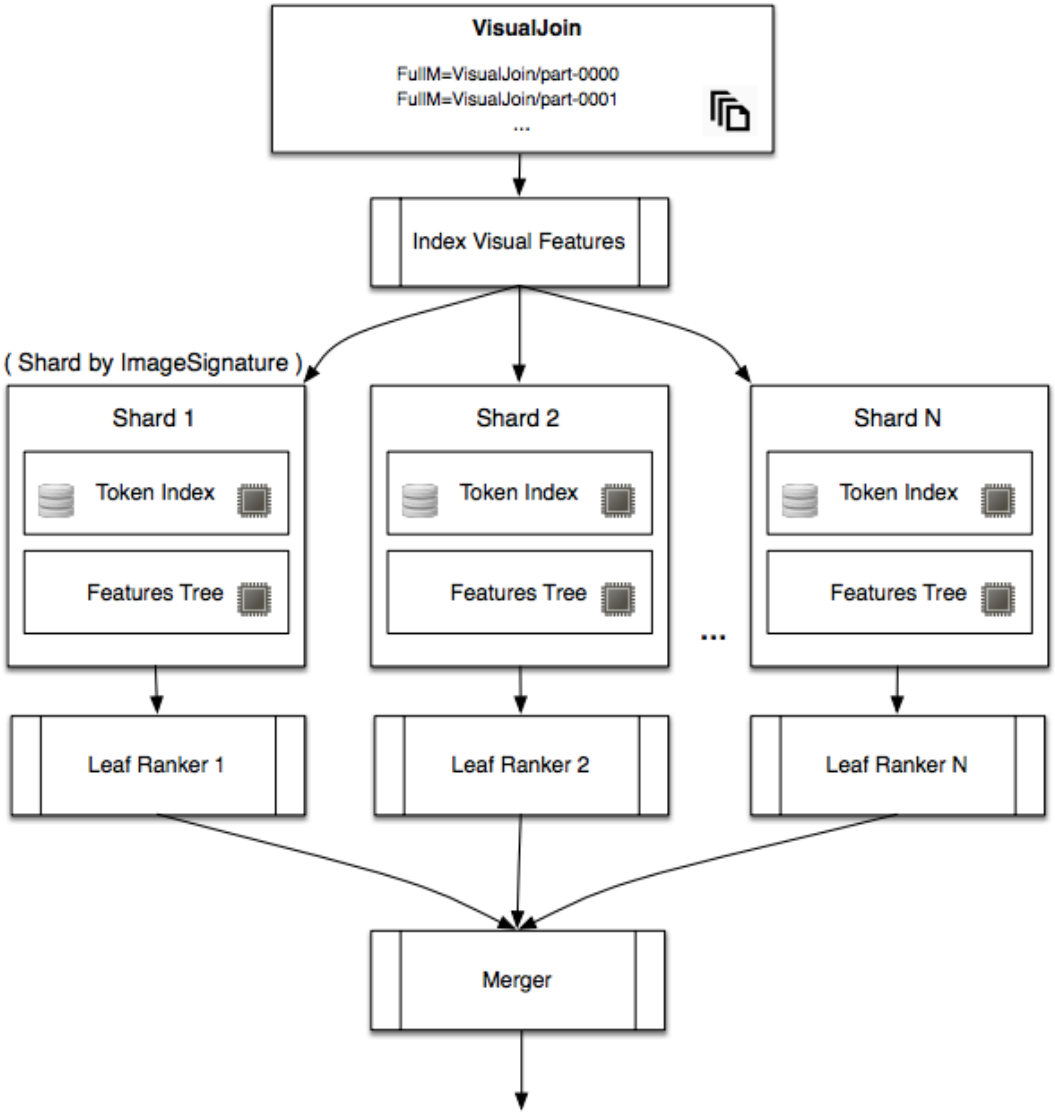
* Detection will be discussed during the guest lecture on Nov 27.

Visual Search at Pinterest,
by Y. Jing et al, KDD'15

Pinterest Image Search

Visual Search at Pinterest, by Y. Jing et al, KDD'15

Search via Distributed Computing



Visual Search at Pinterest, by Y. Jing et al, KDD'15

Lessons from Pinterest Visual Search

- Search systems are complicated but very useful
- ImageNet models give pretty good baselines for visual embedding
- Integrate visual embedding with the text search
- Continuously testing and measuring success, and improving the system