Large-scale visual recognition
Introduction

Florent Perronnin, XRCE
Hervé Jégou, INRIA

CVPR tutorial
June 16, 2012
The tasks

Query-by-example retrieval:

Classification / annotation:

PASCAL VOC 2007

INRIA Holidays
Image retrieval datasets – Oxford5k/Paris6k

Oxford5k dataset: find images of the same famous building
55 queries (11*5 buildings), varying number of relevant results (6-221)
Oxford105K = Oxford5k + a image set of 100k “distractors” for large scale tests

Philbin, Chum, Isard, Sivic and Zisserman,
« Object retrieval with large vocabularies and fast spatial matching », ICCV’07
Datasets for image retrieval

INRIA Holidays dataset: 1491 shots of personal Holiday snapshot
500 queries, each associated with a small number of results 1-11 results
1 million distracting images (with some “false false” positives)

Hervé Jégou, Matthijs Douze and Cordelia Schmid
Hamming Embedding and Weak Geometric consistency for large-scale image search, ECCV’08
Datasets for image retrieval

Stanford Mobile Visual dataset:
- 1200 reference images
- 3000 queries: images shot by mobile devices (queries) – of lower quality
## Scalability for the image search problem

### Scalable systems with Global descriptors (image-level)

- **QBIC’95**: 7.5K (but in 1995!)
- **Cortina**: Quack et al. – ACMM’04
  3 million images (10M)
- **Torralba et al.** – CVPR’08
  12.9M – 74ms with 30bit codes
- **Douze et al.’2009** – CIVR’09
  110 million images – 180ms

### Scalable systems local descriptors (object instance)

- **Sivic et al.** – CVPR’03: “Video-Google”
  5k images
- **Joly et al.** – CIVR’03
  6M video keyframes – 120M descriptors
- **Nister et al.** – CVPR’06
  50k images (then 1M images)
- **Jegou et al.** – CVPR’10
  10M images (then 100M)
Classification datasets

Traditional approach to collect classification datasets:
• query search engine → fast
• manual post-processing result → slow

small training sets
Classification datasets

Traditional approach to collect classification datasets:
- query search engine → fast
- manual post-processing result → slow

![Graph showing the relationship between classes and images]

- small training sets
Classification datasets

Traditional approach to collect classification datasets:

- query search engine → fast
- manual post-processing result → slow

Small training sets

Caltech 101
101 classes, 9K images
Classification datasets

Traditional approach to collect classification datasets:

- query search engine → fast
- manual post-processing result → slow

Scene 15
15 classes, 5K images

- bedroom (FP)
- coast (OT)
- forest (OT)
- highway (OT)
- industrial (L)
- inside city (OT)
- kitchen (FP)
- living room (FP)
- mountain (OT)
- office (FP)
- open country (OT)
- store (L)
- street (OT)
- suburb (FP)
- tall building (OT)
Classification datasets

Traditional approach to collect classification datasets:
- query search engine $\rightarrow$ fast
- manual post-processing result $\rightarrow$ slow

- small training sets

PASCAL VOC’07
20 classes, 10K images

- PASCAL VOC’07
Classification datasets

How to scale the collection process?

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Images</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CALTECH101 (2003)</td>
<td>10K</td>
<td>1K</td>
</tr>
<tr>
<td>SCENE15 (2006)</td>
<td>100K</td>
<td>100</td>
</tr>
<tr>
<td>PASCAL VOC'07</td>
<td>1M</td>
<td>10K</td>
</tr>
<tr>
<td></td>
<td>10M</td>
<td>1M</td>
</tr>
</tbody>
</table>
Classification datasets

How to scale the collection process?

- no manual post-processing → Tiny images

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Images</th>
<th>Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCENE15, 2006</td>
<td>1K</td>
<td>100</td>
</tr>
<tr>
<td>CALTECH101, 2003</td>
<td>10K</td>
<td>1K</td>
</tr>
<tr>
<td>PASCAL VOC’07</td>
<td>100K</td>
<td>1000</td>
</tr>
<tr>
<td>TINY, 2008</td>
<td>1M</td>
<td>75K</td>
</tr>
</tbody>
</table>

Tiny (32x32) images
75K classes, 80M images
Classification datasets

How to scale the collection process:

- no manual post-processing \(\rightarrow\) Tiny images
- crowd-source manual post processing \(\rightarrow\) ImageNet

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Classes</th>
<th>Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tiny Images</td>
<td>10K</td>
<td>100</td>
</tr>
<tr>
<td>ImageNet</td>
<td>10K</td>
<td>1M</td>
</tr>
<tr>
<td>SCENE15</td>
<td>100</td>
<td>10K</td>
</tr>
<tr>
<td>PASCAL VOC'07</td>
<td>1K</td>
<td>100K</td>
</tr>
<tr>
<td>CALTECH101</td>
<td>10</td>
<td>1K</td>
</tr>
<tr>
<td>IMAGENET</td>
<td>1K</td>
<td>1M</td>
</tr>
<tr>
<td>TINY</td>
<td>1</td>
<td>10K</td>
</tr>
<tr>
<td>(2003)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2007)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2009)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classification datasets

How to scale the collection process:

- no manual post-processing → Tiny images
- crow-source manual post processing → ImageNet

ImageNet: cur. release
22K classes, 14M images
Classification datasets

How to scale the collection process:

- no manual post-processing → Tiny images
- crowd-source manual post processing → ImageNet

![Graph showing classification datasets with various numbers of images and classes.](image)
Advantages of ImageNet

Hierarchy of classes:

Deng, Dong, Socher, Li, Li and Fei-Fei, “Imagenet: a large-scale hierarchical image database”, CVPR’09.

Fine-grained subsets: generally more practical problems

→ Fungus: 134 classes, 90K images
Advantages of ImageNet

Hierarchy of classes:

Deng, Dong, Socher, Li, Li and Fei-Fei, “Imagenet: a large-scale hierarchical image database”, CVPR’09.

Fine-grained subsets: generally more practical problems

→ Ungulate: 183 classes, 173K images
Advantages of ImageNet

Hierarchy of classes:

Deng, Dong, Socher, Li, Li and Fei-Fei, “Imagenet: a large-scale hierarchical image database”, CVPR’09.

Fine-grained subsets: generally more practical problems

Vehicle: 262 classes, 226K images
Historical differences

**Query-by-example retrieval**

- Early focus on large-scale
- Focus on instances → little if no learning
- Take image features for granted, e.g. GIST, BOV
- CPU / memory optimized: e.g. compression, approx. search

**Classification / annotation**

- Recent focus on large-scale
- Focus on categories → heavy on machine learning
- Lots of work on features, e.g. how to encode and pool local descriptors
- Get best possible accuracy, whatever the cost
Convergence
Retrieval borrows from classification

Classifier learning from a single positive for cross-domain retrieval
Shrivastava, Malisiewicz, Gupta and Efros, “Data-driven visual similarity for cross-domain image matching”, SIGGRAPH Asia’11.

Classifier learning for query-expansion
Arandjelović, Zisserman, “Three things everyone should know to improve object retrieval”, CVPR’12.

Metric / subspace learning for semantically guided retrieval
Gong, Lazebnik, Iterative Quantization: a procrustean approach to learning binary codes, CVPR’11.

Using image representations designed for classification for retrieval
Convergence
Classification borrows from retrieval

Data-driven classification
Torralba, Fergus, Freeman, “80 million tiny images: a large dataset for non-parametric object and scene recognition”, TPAMI’08.

Classification as a patch matching problem
Boiman, Shechtman, Irani, “In defense of nearest-neighbor based image classification”, CVPR’08.

Compressed features for large-scale classification
Goals of this tutorial

Provide tools to handle large-scale datasets

Show convergence of large-scale retrieval and classification

Show that LSVR does not necessarily require gigantic resources
Outline

Part I: Image representations for LSVR
  • Local description
  • Bag-of-words representation
  • New patch aggregation techniques

Part II: Scalable matching and learning
  • Efficient matching
  • Large-scale learning