Photo Annotation

- Annotation task
  - For 99 concepts, rank 10,000 photos
  - Concept detection system
- Modalities: Visual & tags

- Concept-based retrieval task
  - For 40 topics, rank 200,000 photos
  - Can use existing concept detectors … or train new ones
- Modality: Visual only

Image Feature Extraction: Point Sampling Strategies

- Orientation and scale of object changes
- Salient point methods robustly detect regions
- Preferred for visual categorisation accuracy are
  - Harris-Laplace salient points
  - Dense sampling

Invariant Visual Descriptors

- Color SIFT:
  - Intensity-based SIFT
  - Opponent SIFT
  - C-SIFT
  - rSIFT
  - RGB-SIFT

- Add color, but also keep intensity information

Example: light color change

RGB-SIFT descriptor is invariant
Visual Feature Overview

Experimental results

ColorDescriptor Software
- Today at v3.0
- GPU-accelerated version (CUDA) for 64-bit Windows: end of next month

Train concept model
- SVM with nonlinear kernel
  - $\chi^2$ kernel
  - Histogram intersection kernel
- Fast to apply with intersection kernel
- Other approaches work with linear kernels (Fisher vector, supervector), but need much longer feature vectors

Annotation task
in 1 slide

Concept-based retrieval

Selected 1 likelihood threshold for all concepts
Concept-based retrieval
- 40 topics
- 5 examples per topic
- 200,000 images in test set

What to do?
- More discriminative features? Keep features fixed
- More powerful classifiers? Keep classifier fixed
- Better training examples? Experiment with training & retrieval method

Approach 1: Fully automatic
- Take 5 positive example images
- Take 10, 33 or 100 random negatives
- Train new model for topic

Visual-only

Approach 2: Human topic mapping
- Topics are often Boolean combinations of existing concepts
- Use 99 concept detectors from the annotation task
- Parsing the relations between the concepts from the topic text is challenging
- Let a human pick 1 or 2 concepts per topic, fusion of classification scores

Approach 3: Human topic inspection
- Start with 5 given positive images
- Apply automatic model on the annotation task train set (8,000 images)
- Human adds more examples & retrain
- At most 7.5 minutes per topic
- Increases #pos from 5 to ~42
- We never look at the test set

Results 1: Fully automatic
<table>
<thead>
<tr>
<th>Topic</th>
<th>Fully automatic</th>
<th>Sub10</th>
<th>Sub33</th>
<th>Sub100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Graffiti on buildings/walls</td>
<td>0.017</td>
<td>0.014</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>2 Toy vehicle</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>31 Person doing sports at sea</td>
<td>0.123</td>
<td>0.057</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td>4 Airplane in the sky</td>
<td>0.045</td>
<td>0.000</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>28 Fireworks</td>
<td>0.000</td>
<td>0.384</td>
<td>0.415</td>
<td></td>
</tr>
<tr>
<td>30 Close-up of flowers with raindrops</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>32 Underexposed photos of animals</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>33 Cars and motion blur</td>
<td>0.000</td>
<td>0.128</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>0.018</td>
<td>0.037</td>
<td>0.043</td>
<td></td>
</tr>
</tbody>
</table>

- Should have used >100 random negatives?
- Our overall best for 3/40 topics

Results 2: Human topic mapping
<table>
<thead>
<tr>
<th>Topic</th>
<th>Human topic mapping</th>
<th>Concept #1</th>
<th>Concept #2</th>
<th>Concept #3</th>
<th>Concept #4</th>
<th>Concept #5</th>
<th>Concept #6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Graffiti on buildings/walls</td>
<td>Graffiti</td>
<td>Building Sign</td>
<td>0.054</td>
<td>0.022</td>
<td>0.010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Toy vehicle</td>
<td>Toy</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31 Person doing sports at sea</td>
<td>Single Person</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 Airplane in the sky</td>
<td>Airplane</td>
<td>Sky</td>
<td>0.125</td>
<td>0.165</td>
<td>0.166</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 Fireworks</td>
<td>Fire</td>
<td>Outdoor</td>
<td>0.006</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 Close-up of flowers with raindrops</td>
<td>Flowers</td>
<td>Rain</td>
<td>0.002</td>
<td>0.015</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 Underexposed photos of animals</td>
<td>Underexposed</td>
<td>Animal</td>
<td>0.025</td>
<td>0.025</td>
<td>0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33 Cars and motion blur</td>
<td>Car</td>
<td>Motion Blur</td>
<td>0.019</td>
<td>0.108</td>
<td>0.110</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAP</td>
<td>0.055</td>
<td>0.089</td>
<td>0.088</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Didn’t use negation, only ‘AND’
- Works only if there are related concepts
- Our overall best for 21/40 topics
Results 3: Human topic inspection

<table>
<thead>
<tr>
<th>Topic Description</th>
<th>#pos</th>
<th>#neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graffiti on buildings/walls</td>
<td>0,184</td>
<td>0,253</td>
</tr>
<tr>
<td>Toy vehicle</td>
<td>0,000</td>
<td>0,000</td>
</tr>
<tr>
<td>Person doing sports at sea</td>
<td>0,007</td>
<td>0,009</td>
</tr>
<tr>
<td>Airplane in the sky</td>
<td>0,051</td>
<td>0,091</td>
</tr>
<tr>
<td>Fireworks</td>
<td>0,404</td>
<td>0,423</td>
</tr>
<tr>
<td>Close-up of flowers with raindrops</td>
<td>0,005</td>
<td>0,006</td>
</tr>
<tr>
<td>Cute toys arranged as still-life</td>
<td>0,044</td>
<td>0,060</td>
</tr>
<tr>
<td>Ship/boat on a river</td>
<td>0,012</td>
<td>0,024</td>
</tr>
<tr>
<td>Underexposed photos of animals</td>
<td>0,068</td>
<td>0,065</td>
</tr>
<tr>
<td>Cars and motion blur</td>
<td>0,340</td>
<td>0,345</td>
</tr>
<tr>
<td>Close-up of bodypart</td>
<td>0,230</td>
<td>0,261</td>
</tr>
</tbody>
</table>

MAP = 0,108

#pos (on average): 42
#neg (on average): 328

- Topic example images give starting point
- Build a new ‘concept’ on the fly
- Our overall best for 17/40 topics

Demo

#28. Fireworks. We like to find photos of fireworks at night in outdoor images
#15. Sea sunset or sunrise. We like to find pictures of a sunrise or sunset over the sea. So sunrises over cities or over lakes are not relevant. If the picture shows persons or vehicles it is also not relevant

Conclusions

- Annotation task
  - Focus on feature representations, machine learning algorithms
- Concept-based retrieval is a new and different task
  - Focus on training approaches, using proven visual features
- Lessons learned
  - Should have used the text tags
  - For ~50% of topics, selecting 1-2 ‘relevant’ concepts works
  - Should support complex relations: ‘A but not B’
  - If approach 1 & 2 don’t work, difficult to bootstrap 3… unless you would exploit tags

ColorDescriptor Software

Visit http://www.colordescriptors.com