CAP 6412 Advanced Computer Vision


Boqing Gong
March 17, 2016
Today

- Administrivia
- Datasets in Computer Vision
- Recurrent Neural Networks (RNNs) (?)
- Visual Genome, by Shreyas
Assignment 9: Due on Tuesday 03/22, 12pm


Next week: DAG-CNN & Transferability

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Niladri Basu Bal</td>
<td></td>
</tr>
<tr>
<td>Mert Ozerdem</td>
<td></td>
</tr>
</tbody>
</table>
What’s next

• Sign up for volunteer presentations at
  • https://docs.google.com/spreadsheets/d/1DxMQ_RVMx8BLmc5gij51dtXI2ZJzktR8EzJJZkVpGW4/edit#gid=0

• Suggest papers you would like to read, share, challenge

• We need another two or three volunteers
Today

• Administrivia
• Datasets in Computer Vision
• Recurrent Neural Networks (RNNs) (?)
• Visual Genome, by Shreyas
Datasets in Computer Vision

• Drive and enable research
• Benchmark research output: algorithms, systems, heuristics, etc.
• Represent real-world problems (?)
  • Simplify challenging problems

MIT + CMU
Frontal Faces
Datasets in Computer Vision

- Drive and enable research
- Benchmark research output: algorithms, systems, heuristics, etc.
- Represent real-world problems (?)
  - Simplify challenging problems
  - Introduce controlled variations

CMU Multi-PIE
Datasets in Computer Vision

• Drive and enable research
• Benchmark research output: algorithms, systems, heuristics, etc.
• Represent real-world problems (?)
  • Simplify challenging problems
  • Introduce controlled variations
  • Simulate the “wild”

Umass Labeled Faces in the Wild
How many object categories are there?

~10,000 to 30,000

Biederman 1987

slide credit Fei-Fei et. al.
### Size of existing datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th># of categories</th>
<th># of images per category</th>
<th># of total images</th>
<th>Collected by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caltech101</td>
<td>101</td>
<td>~100</td>
<td>~10K</td>
<td>Human</td>
</tr>
<tr>
<td>Lotus Hill</td>
<td>~300</td>
<td>~500</td>
<td>~150K</td>
<td>Human</td>
</tr>
<tr>
<td>LabelMe</td>
<td>183</td>
<td>~200</td>
<td>~30K</td>
<td>Human</td>
</tr>
<tr>
<td>Ideal</td>
<td>~30K</td>
<td>&gt;&gt;10^2</td>
<td>A LOT</td>
<td>Machine</td>
</tr>
</tbody>
</table>

*slide credit Fei-Fei et al.*
IMAGENET

- An ontology of images based on WordNet
- Collected using Amazon Mechanical Turk
IMAGENET

14,847 categories, 9,349,136 images

- Animals
  - Fish
  - Bird
  - Mammal
  - Invertebrate

- Scenes
  - Indoors
  - Geological formations

- Sport Activities
- Fabric Materials
- Instrumentation
  - Tool
  - Appliances
  - ...

- Plants
  - ...

slide credit Fei-Fei et al.

Deng, Wei, Socher, Li, Li, Fei-Fei, CVPR 2009
Computer Vision tasks beyond categorization

- 3D reconstruction from images for face, scenes
- Segmentation of medical images
- Registration of medical images
- Scene understanding
  - Relationships: spatial, activity, etc.
- Reflection and reconstruction (low-level vision)
- Retrieval tasks
- OCR in the wild
- Signals → Spectrum → Computer vision techniques
- Attribute categorization
- Pose estimation → Human-centric computer vision
- Tracking → Problem on videos
Middlebury Stereo Datasets

**2001 datasets** - 6 datasets of piecewise planar scenes [1]
(Sawtooth, Venus, Bull, Poster, Barn1, Barn2)

**2003 datasets** - 2 datasets with ground truth obtained using structured light [2]
(Cornes, Teddy)

**2005 datasets** - 9 datasets obtained using the technique of [2], published in [3, 4]
(Art, Books, Dolls, Laundry, Moebius, Reindeer, Computer, Drumsticks, Dwarves)

**2006 datasets** - 21 datasets obtained using the technique of [2], published in [3, 4]
(Aloe, Baby1-3, Bowling1-2, Cloth1-4, Flowerpots, Lampshade1-2, Midd1-2, Monopoly, Wood1-2)
Berkeley Segmentation Data Set 500
Large scale instance retrieval

Oxford Buildings Dataset

INRIA Holidays Dataset
The Indoor Scene Dataset

- 67 indoor categories
- 15620 images
- At least 100 images per category
- Training 67 x 80 images
- Testing 67 x 20 images
Caltech Pedestrian Dataset

- 350,000 labeled pedestrian bounding boxes
- 250,000 frames
Person layout

PASCAL VOC "Person Layout"

Oxford Buffy Stickmen
276 frames x 6 = 1656 body parts (sticks)

ETHZ Pascal stickmen set
549 images x 6 = 3294 body parts (sticks)
Human action recognition

Hollywood2 dataset

- 12 different action classes from 69 Hollywood movies
- 1707 video sequences in total
- Separate movies for training / testing
- Performance measure: mean average precision (mAP) over all classes
Fine grained visual categorization

The Oxford Flowers 102

Caltech-UCSD birds 200
Material recognition

Exploring Features in a Bayesian Framework for Material Recognition
Ce Liu, Lavanya Sharan, Edward H. Adelson, and Ruth Rosenholtz
CVPR 2010
Is it difficult to create datasets?

• The following slides are borrowed from Larry Zitnick
Space of all images

Interesting

Boring

Common   Unusual
A man riding a wave on a surfboard in the water.

“surfboard”
Unusual

Interesting

Common

Boring

Unusual

flickr™
A crazy zebra climbing a giraffe to get a better view.

The limits of vision and language models...
A man is rescued from his truck that is hanging perilously from a bridge.

The limits of knowledge...
Today

• Administrivia
• Datasets in Computer Vision
• Recurrent Neural Networks (RNNs) (?)
• Visual Genome, by Shreyas
Upload slides before or after class

• See “Paper Presentation” on UCF webcourse

• Sharing your slides
  • Refer to the original sources of images, figures, etc. in your slides
  • Convert them to a PDF file
  • Upload the PDF file to “Paper Presentation” after your presentation