Searching for Images in the era of Deep Learning

STM Innovations
December 5th, 2018

DR HARALAMBOS MARMANIS
EVP & CTO, CCC
“Drops splashing” on Google
What if?

• We had hand-drawn sketches for a book and wanted to search our image inventory to identify professional drawings or pictures?
• What if we wanted to organize the images for an astronomy journal so that all look-alike galaxies are grouped together?
• What if we wanted to autogenerate captions for our vast inventory of images?
• What if … ?
Before Deep Learning

- Metadata-based image indexing
- Content-based image indexing
  - Color histograms (HSV, RGB)
  - JPEG Coefficient histogram
  - Visual Bag of Words with SHIFT or SURF
  - Approximate fast search based on hashing and metric indexing

Great tool to use for the above: LIRE
ILSVRC

- Task: Image Classification (1000 Classes)
  - 1.2m images in the training set
  - 100k images in the test set
- Goal: Predict the best (top 5) class
- Evaluation: One of the top 5 is correct
ImageNet search for “dog”
Let the party begin!

# ILSVRC Progress

<table>
<thead>
<tr>
<th>Year</th>
<th>Winner</th>
<th>Error</th>
<th>Depth</th>
<th>Filter</th>
<th>C</th>
<th>T (Days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>AlexNet</td>
<td>15.4%</td>
<td>7</td>
<td>11x11</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>2013</td>
<td>ZFNet</td>
<td>11.2%</td>
<td>7</td>
<td>7x7</td>
<td>M</td>
<td>12</td>
</tr>
<tr>
<td>2014</td>
<td>GoogLeNet</td>
<td>6.7%</td>
<td>100</td>
<td>1x1, 3x3, 5x5</td>
<td>H</td>
<td>7*</td>
</tr>
<tr>
<td>2015</td>
<td>ResNet</td>
<td>3.6%</td>
<td>152</td>
<td>3x3</td>
<td>L</td>
<td>14-21</td>
</tr>
</tbody>
</table>
Deep Learning is not just CNNs

Deep learning methods

- CNN-based Methods
  - AlexNet [14]
  - Clarifai [22]
  - SPP [23]
  - VGG [24]
  - GoogLeNet [25]
  - Deep Belief Networks [8]
  - Deep Boltzmann Machines [26]
  - Deep Energy Models [27]

- RBM-based Methods

- Autoencoder-based Methods
  - Sparse Autoencoder [28]
  - Denoising Autoencoder [29]
  - Contractive Autoencoder [30]

- Sparse Coding-based Methods
  - Sparse Coding SPM [31]
  - Laplacian Sparse Coding [32]
  - Local Coordinate Coding [33]
  - Super-Vector Coding [34]
Typical CNN for image processing

- Input layer
- Convolution layer
- Pooling layer
- Convolution layer
- Pooling layer
- ………
- Dense layer
- Output layer

Image Representation
Image Classification
Enhanced Search with CNN models

• Build a CNN model
• Train the model
• Store the model
  • Get the CNN feature vectors and index them in Lucene as Points
• Perform a feedforward pass with our image
• Use the FloatPointNearestNeighbor.nearest(…)
• Select the top-k results
The art of the trade

• Network architecture
  • How many layers? What type of layers? How many nodes in each layer for a certain task?

• Training strategy
  • Dropout & DropConnect
  • Data Augmentation
    − Generate image translations and horizontal reflections
    − Alter the values of the RGB channels in training images

• Pre-Training and Fine-Tuning

• Training with limited data

• Time complexity