Collaborative Feature Learning from Social Media
Hailin Jin | Principal Scientist | Adobe Research
Collaborators

- Chen Fang (Dartmouth)
- Jianchao Yang and Zhe Lin (Adobe)
Deep Learning

- Traditional model of learning
  - Fixed/engineered features (or kernels) + trainable classifiers

- Deep Learning: End-to-end learning from data
  - Trainable features (or kernels) + trainable classifiers
ImageNet large-scale image classification challenge

- **Data**
  - 1000 categories
  - 1.2M images for training
  - 150K images for validation and testing

- **Task**
  - Classification
ImageNet large-scale image classification challenge

- Best top-5 error rate
  - ~6%
- Human performance
  - ~5%
Learning beyond labeled data

- Supervised learning
  - Powerful model
  - Lots of data
- Unsupervised and semi-supervised learning
- Transfer learning

- Can we use other kinds of “labels?”
Visual information in social data
<table>
<thead>
<tr>
<th></th>
<th>u1</th>
<th>u2</th>
<th>u3</th>
<th>u4</th>
<th>u5</th>
<th>u6</th>
<th>u7</th>
<th>u8</th>
<th>u9</th>
<th>u10</th>
<th>u11</th>
<th>u12</th>
<th>u13</th>
<th>u14</th>
<th>u15</th>
<th>u16</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p4</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p5</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p6</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>p7</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>p8</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>p9</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>p10</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

User-project matrix (views, appreciations, comments, ownership)
Recommender systems

- Matrix factorization – filling in missing information
- Discover user and project latent factors

\[
\min_{x^*, y^*} \sum_{V_{ij} \neq -1} (V_{ij} - y_i^T x_j)^2 + \lambda(||x_j||^2 + ||y_i||^2)
\]
Regression using deep convolutional neural network
Algorithm overview
Collaborative feature learning
Behance 2M dataset

- 1.9M users and 1.9M projects
- 45M appreciations and 273M views
- Matrix density: 0.0013% and 0.0091%
Social data pre-processing

- Remove least and most popular projects and users
- Processed data:
  - 309K users and 423K projects
  - 31M appreciations (70%) and 178M views (65%)
  - Matrix density: 0.03% and 0.16%
Image similarity in latent space

- beauty
- portrait
- woman
- hair

- wedding
- photography

- elegant
- graceful
- neat
- refined

- automotive
- classic

- automotive
- design
- industrial
- transportation

- Casa La Encantada
- house
Image retrieval

![Graphs showing viewer overlap and viewed NN as a function of nearest neighbor for Behance feature fc6, ImageNet feature fc6, and random guess.](image-url)
## Feature comparison

<table>
<thead>
<tr>
<th></th>
<th>Collaborative</th>
<th>ImageNet</th>
<th>Meta-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr style</td>
<td>37.2</td>
<td>37.1</td>
<td>32.8</td>
</tr>
<tr>
<td>Wiki Paintings</td>
<td>41.4</td>
<td>40.7</td>
<td>38.6</td>
</tr>
<tr>
<td>AVA style</td>
<td>56.0</td>
<td>51.3</td>
<td>53.9</td>
</tr>
<tr>
<td>Caltech 256</td>
<td>57.6</td>
<td>68.9</td>
<td>48.9</td>
</tr>
</tbody>
</table>
Summary

- Learn image feature from social data and images
- No labels are needed
- Scale to billions of users/images/views
- To be presented at CVPR 2015