Fine-Grained Visual Comparisons with Local Learning

Aron Yu
Kristen Grauman
University of Texas at Austin

Visual Comparisons
Which shoe is more sporty?

Problem: Fine-grained visual comparisons require accounting for subtle visual differences specific to each comparison pair.

Status Quo: Learning a Global Ranking Function

Our Approach
We propose a local learning approach for fine-grained comparisons.

Analogous Neighboring Pairs
Detect analogous pairs based on individual similarity & paired contrast.
- select neighboring pairs that accentuate fine-grained differences
- take product of pairwise distances of individual members
- i.e., highly analogous if both query-training pairs are similar

Learned Attribute Distance
Learn a Mahalanobis metric per attribute (similarity computation).
- attribute similarity doesn’t rely equally on each dim of feature space
- constraints similar images be close, dissimilar images be far

Results: UT-Zap50K
- FG-LocalPair: our proposed fine-grained approach
- LocalPair: our approach w/o the learned metric
- RandPair: local approach with random neighbors
- GlobalRank: a baseline w/o fine-grained ranking
- ReTrec [ICCV 13]: non-linear relative attribute approach

Accuracy Comparison (10 iterations @ K=100)

<table>
<thead>
<tr>
<th>Approach</th>
<th>Open</th>
<th>Pointy</th>
<th>Sporty</th>
<th>Comfort</th>
</tr>
</thead>
<tbody>
<tr>
<td>FG-LocalPair</td>
<td>87.77</td>
<td>60.27</td>
<td>91.20</td>
<td>86.93</td>
</tr>
<tr>
<td>LocalPair</td>
<td>82.53</td>
<td>63.70</td>
<td>86.30</td>
<td>84.77</td>
</tr>
<tr>
<td>RandPair</td>
<td>88.53</td>
<td>68.07</td>
<td>92.20</td>
<td>90.90</td>
</tr>
<tr>
<td>ReTrec</td>
<td>90.47</td>
<td>92.67</td>
<td>93.37</td>
<td></td>
</tr>
</tbody>
</table>

Learned vs Global Models
Our approach w/o the learned metric

Results: PubFig & Scenes
We form supervision pairs using the category-wise comparisons ➔ avg. 20,000 ordered labels / attribute.
- Public Figures Face (PubFig): 772 images w/ 11 attributes
- Outdoor Scene Recognition (OSR): 2,688 images w/ 6 attributes

Observation: We outperform all baselines, demonstrating strong advantage for detecting subtle differences on the harder comparisons (~20% more).

UT Zappos50K Dataset
We introduce a new large shoe dataset UT-Zapp50K, consisting of 50,025 catalog images from Zappos.com.

- 4 relative attributes (open, pointy, sporty, comfort)
- High confidence pairwise labels from SKU workers
- 6,751 ordered labels = 4,612 “equal” labels
- 4,334 twice-labeled fine-grained labels (no “equal” option)

Observation: Nearest analogous pairs most suited for local learning need not be those closest in raw feature space.

Key Idea: having the right data ➔ having more data

Project webpage here