Discovering Visual Patterns in Art Collections with Spatially-consistent Feature Learning

Xi Shen ¹, Alexei A. Efros ², Mathieu Aubry ¹

¹École des Ponts ParisTech
²UC Berkeley

2019 CVPR
Motivation

Figure: Duplicated details in Rossetti’s paintings
Challenges

- Artworks in different medias, color with geometric deformation.
- No training data available.

Figure: (a) *Nymphs Sleeping After the Hunt, Spied on by Satyr* (oil); (b) *Diana’s Nymphs After the Hunt* (oil); (c) *Seventeen Studies of Different Dogs* (drawing). Images are from Brueghel dataset.
Main Contribution

A self-supervised feature fine-tuning for matching:
- **instance**, not categories
- **across domains**, e.g. engraving, oil painting...
Feature Learning: Candidates from Matching in the Database
Figure: Hard positive training sample, Green Regions.
Feature Learning : Metric Learning
Discovery Score:

\[ S(\mathcal{I}) = \frac{1}{N} \sum_{i \in \mathcal{I}} e\left(-\frac{e_i^2}{2 \sigma^2}\right) s_i \]

- \( \mathcal{I} \): inlier set;
- \( e_i \): error to fit the geometric model;
- \( s_i \): similarity of the descriptors;
- \( N \): number of features in the source image.
Datasets

- Large Time Lags Location (LTLL);
- Oxford 5K;
- Brueghel.

Figure: Our detection results in Brueghel with learned feature for 10 annotated categories.
Qualitative Results

Figure: One shot detection results obtained with cosine similarity with ImageNet feature (top) and our trained features (middle) as well as the ones obtained with our features and the discovery score (bottom).
# Quantitative Results

<table>
<thead>
<tr>
<th>Feature \ Method</th>
<th>Cosine similarity</th>
<th>Discovery score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ImageNet pre-taining</td>
<td>58.0</td>
<td>54.8</td>
</tr>
<tr>
<td>C. Doersch et al. 2015</td>
<td>58.8</td>
<td>64.29</td>
</tr>
<tr>
<td>Ours (trained on Brueghel)</td>
<td><strong>75.3</strong></td>
<td><strong>76.4</strong></td>
</tr>
<tr>
<td>Ours (trained on LTLL)</td>
<td>65.2</td>
<td>69.95</td>
</tr>
</tbody>
</table>

Table: Experimental results on Brueghel, IoU $> 0.3$.

<table>
<thead>
<tr>
<th>Method</th>
<th>LTLL (%)</th>
<th>Oxford (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. Fernando et al. 2015</td>
<td>56.1</td>
<td>-</td>
</tr>
<tr>
<td>F. Radenović et al. 2017</td>
<td>-</td>
<td>87.8</td>
</tr>
<tr>
<td>ResNet18 max-pool, image level</td>
<td>59.8</td>
<td>14.0</td>
</tr>
<tr>
<td>ResNet18 + discovery</td>
<td>80.9</td>
<td>85.0</td>
</tr>
<tr>
<td>Ours (trained LTLL + discovery)</td>
<td><strong>88.5</strong></td>
<td>83.6</td>
</tr>
<tr>
<td>Ours (trained Oxford + discovery)</td>
<td>85.6</td>
<td><strong>85.7</strong></td>
</tr>
</tbody>
</table>

Table: Classification accuracy on LTLL and retrieval mAP on Oxford5K.
Discovery between Images during Training (LTLL)

Iter 1800
Source Image

Iter 3600
Target Image

Iter 5400
Discovery between Images during Training (Brueghel)

Source Image

Iter 1800

Iter 3600

Iter 5400

Target Image
Visual results Brueghel (1)
Visual results Brueghel (2)
Visual results Brueghel (3)
Visual results Brueghel (4)
Visual results Dante Gabriel Rossetti (1)
Visual results Dante Gabriel Rossetti (2)
Visual results Peter Paul Rubens (1)
Visual results Canaletto (1)
Visual results Canaletto (2)
Visual results Canaletto (3)