Artistic Image Analysis using Graph-based Learning Approaches

Gustavo Carneiro
Joint work with Joao P. Costeira, Nuno P. Silva, Alessio del Bue

Australian Center for Visual Technologies and
The University of Adelaide
ICMR’11, ECCV’12, TIP’13

Background...

- National Tile Museum (Museu Nacional do Azulejo), Lisbon

PROBLEM 1

Artist?
Influences?
When?
Where?
PRINTART

PROBLEM 2
PrintArt

- System for organizing art image databases
- Image annotation
- Image retrieval
  - Text query
  - Image query

Image Annotation

Query image:

1. Initial annotation: Black and white print, Italy, XV century, Christian Art, Angel, Donkey, Robes, Saint Mary, Blessed Virgin, Saint Joseph
2. User correction: infant Jesus Christ, XVII century

Image Annotation & Print Retrieval


4. Image returned.
Image Retrieval

• Query:

“Jesus Christ”, “Saint Mary Blessed Virgin”, and “Saint Joseph”.

The system then returns the following images:

User selects the last two images above as relevant.

User selects the last image above as the target image, and search is done.

Art Prints

• Source of inspiration for generations of artists

• Art Historian
  • Discover influences between art works
  • In artistic image analysis, we should study prints because
    • Larger availability, widespread = influenced more artists
Art Prints

- Source of inspiration for generations of artists

Art Historian

Discover influences between art works

In artistic image analysis, we should study prints because

- Larger availability, widespread = influenced more artists
Art Prints

• Source of inspiration for generations of artists

• In artistic image analysis, we should study prints because
  • Larger availability, widespread = influenced more artists

Art Prints

• Source of inspiration for generations of artists

• Art Historian
  • Discover influences between art works
  • In artistic image analysis, we should study prints because
    • Larger availability, widespread = influenced more artists
Art Prints

- Printmaking
  - Create art works by printing
  - 1400s (engraving, etching, etc.)
  - Even a “copy” of a painting is an original work of art (impression)

Different from Photo and Painting

- Cons: no colour, texture does not represent visual classes
- Pros: artistic influence network, composition
Interesting example...

- How can you characterize the visual class “sea” in these examples

Fig. 1. Different paintings showing the visual class “sea” with remarkably different patterns of color and texture. In (a), we show Pieter Brueghel II's Christ on the Storm on the Sea of Galilee; in (b) we have Claude Monet’s Shadows on the Sea; and (c) displays John Marin’s Sea Piece.

Example of Network of Influences & Composition (Annunciation)

Anonymous, 1580  Allaert Claes, ?  Virgil Solis, 1550
Background - Art Analysis

- Fake vs Original

![Original vs Forged Images]

[A Digital Technique for Art Authentication. Lyu et al. Nat. Acad. Sciences'04]

- Multiclass classification of brushwork

![Brushwork Categories]

[Yelizaveta et al. Semi-supervised annotation of brushwork in paintings domain using serial combinations of multiple experts. ACM Multimedia'06]

Our Goal

- Produce three levels of artistic image annotation for a previously unseen print
  - Global (theme, things present in the scene)
  - Local (localize in the image the things identified in the global annotation)
  - Pose (localize head, torso and limbs of human/animal subjects from local annotation)

![Image Annotations]

-Holy Family, Christ child, Mary, St. Joseph
-Christ child (blue), Mary (red), St. Joseph (green)
-Christ child (blue), Mary (red), St. Joseph (green)
Why is this Interesting

- Although a common task for photographic images, never done before for artistic images
  - Shed some light in computer vision problems?

- Can be used in tools to annotate artistic images and find influential prints

- Therefore, interdisciplinary (art history + computer vision)
  - New projects
  - Education

---

Sculpture by Karel Nepras, entitled “Great Dialogue,” Museum of Modern and Contemporary Art in Prague
Problem Set-up

• Database
  - 988 annotated images
  - Global annotation: 49 classification problems (1 multi-class with 27 classes and 48 binary problems)
    label cardinality = 4.22, label density = 0.05
  - Local annotation: 48 detection problems
  - Pose: 40 pose identification problems (37 human, 3 animal)

• \( \mathcal{D} = \{ (\mathbf{x}_i, y_i, L_i, P_i) \}_{i=1}^{\mathcal{D}} \)
  - feature vector representing an image \( I_i \)
  - \( Y = [y_i(1), ..., y_i(M)] \in \{0, 1\}^M \)
  - \( L = \{ l_{i,j} \}_{j=1}^{C} \) with \( l_{i,j} = [x, y] \)
  - \( P_i = \{ p_{i,j} \}_{j=1}^{P} \) with \( p_{i,j} = [y, h, w, \theta_v, \theta_h] \)

Problem Set-up

• Training set with 90% (889 images) and test set with 10% (99 images)

• Global Annotation
  
  \[ \text{maximize } p(\mathbf{y}|\mathbf{x}) \]
  
  subject to \( \mathbf{y} = [y(1), ..., y(M)] \in \{0, 1\}^M \),
  
  \[ \|\mathbf{y}(k)\|_1 = 1 \text{ for } k \in \{1, ..., M\} | y(k) > 1 \].

• Local Annotation
  
  \[ \text{maximize } p(L|\mathbf{y}, \mathbf{x}) \]
  
  each \( k \) that \( |y(k)| = 1 \) and \( y(k) = 1 \) has a respective bounding box \( l_k \in L \).

• Pose Annotation
  
  \[ \text{maximize } p(P|L, \mathbf{y}, \mathbf{x}), \]
  
  head and torso are within the bounds of the local annotation

• Retrieval
  
  \[ \hat{\mathbf{x}} = \arg\max_{\mathbf{x} \in \mathcal{F}} p(\mathbf{x}|q). \]
### Image Representation

- Scale Invariant Feature Transform (SIFT) [Lowe’04]
  ![SIFT Diagram](image)

- Visual vocabulary – hierarchical K-means [Zisserman’03, Nister’06]
  - 3 levels, 10 descendants per node = 1111 histogram bins
  ![Visual Vocabulary Diagram](image)

- Spatial pyramid representation [Lazebnik’06]

### Methodologies

- **Random (RND)**
  
  - **Global:**
    - Multiclass: \( \{ k : |y(k)| > 1 \} \)
    - Binary: \( \{ k : |y(k)| = 1 \} \)
    
    \[
    y^*(k) = \begin{cases} 
    y(k) & \sum_{j=1}^{|y(k)|-1} p(y(k) = \pi_j) \leq r < 1 \\
    \pi_1 & r < p(y(k) = \pi_1)
    \end{cases}
    \]
    
    \( r \sim \mathcal{U}(0, 1) \): uniform distribution

  - **Local and pose** (search training image and propagate label)
    
    \[
    i^* = \arg\min_{j \in \{1, \ldots, |\mathcal{D}|\}} \Delta(y^*, y_j)
    \]

  - **Retrieval**
    - Annotate all test images, and given a query annotation, search for closest ones in terms of Hamming distance
Methodologies

• Bag of features (BoF) [Zisserman’03, Csurka’04]
  - Train $Y$ support vector machine classifiers (SVM) with one-versus-all
  - $p(y(k) = \pi_j | x, \theta_{SVM}(k, j))$, $k \in \{1, ..., M\}, j \in \{1, ..., |y(k)|\}, \pi_j \in (0, 1)^{|y(k)|}$
  - Annotate by maximizing the global, local and pose annotation objective functions
  - Retrieval by first annotating test images, and then retrieving using query vector

• Label Propagation (LP) [Scholkopf’04]
  - Find annotation matrix $F^*$ with
    
    $$\text{minimize } 0.5 \text{ tr}(F^T (D - W) F)$$
    
    subject to $f_i = y_i$, for $i = 1, ..., |D|$
    
    where $W, F, D \in \mathbb{R}^{(|D|+|T|) \times (|D|+|T|)}$, with $W_{ij} = \exp(-0.5 \|x_i - x_j\|^2 / \sigma^2)$
    
    $D$ is a diagonal matrix with its $(i,i)$-element is sum of the $i^{th}$ row of $W$

  - Closed form solution: $F^* = \beta (I - \alpha (D - W))^{-1} Y$

• Label Propagation with Label Correlation (LP-CC) [Wang’09, Zha’08]
  - $\text{minimize } 0.5 \text{ tr}(F^T (D - W) F) + (1 - \mu) \text{ tr}((F - Y) A (F - Y)) + \rho \text{ tr}(FCF^T)$
    
    $A$ is a matrix containing ones in the diagonal from indices 1 to $|D|$
    
    $C \in [-1, 1]^{n \times n}$ containing the correlation between classes

  - Closed form solution: $F^* = (D - W)^{-1} Y (I - \mu C)$
Methodologies

- Inverted Label Propagation (ILP)
- Graph \( G = (V, E) \)
- Test image \( \tilde{x} \), random walk \( t = [(x^{(1)}, y^{(1)}), ..., (x^{(R)}, y^{(R)})] \)
- Label given random walk: \( p(y | \tilde{x}) = Z \sum_{r=1}^{R} p(y | t_r) p(t_r | \tilde{x}) \)
- Assume Markov process:

\[
p(t | \tilde{x}) = \prod_{u=2}^{R} p((x^{(u)}, y^{(u)}) | (x^{(u-1)}, y^{(u-1)}), \tilde{x}) p((x^{(1)}, y^{(1)}) | \tilde{x})
\]

Inverse Label Propagation
Methodologies

- **Inverted Label Propagation (ILP)**
  - Instead of finding $P$, estimate a vector containing the probability of landing in one of the training images after a random walk process.
  - Combinatorial Harmonics (CH) approach [Carneiro'11]
  - Adjacency matrix takes into consideration visual and label similarity:
    \[
    U(j, i) = I_y(y, y_j) \times I_x(x, x_j) \times I_y(x, \bar{x})
    \]
    \[
    I_y(y, y_j) = \sum_{k=1}^{M} \lambda_k \times y(k) \times y(k_j)
    \]
    \[
    I_x(x, x_j) = \sum_{d=1}^{N} \min(x(d), x_j(d))
    \]
  - Extend adjacency matrix with test image:
    \[
    \tilde{U} = \begin{bmatrix} U & \tilde{u} \\ \tilde{u}^T & 0 \end{bmatrix}
    \]
    with \( \tilde{u} = [I_y(x, \bar{x}), ..., I_y(x_{|D|}, \bar{x})]^T \)
  - Minimize the energy function:
    \[
    E(G, g) = \frac{1}{2} \left\| G g \left[ \begin{array}{c} \tilde{g} \\ \tilde{g}^T \end{array} \right] \right\|_2
    \]
    $L$ is the Laplacian from the adjacency matrix $\tilde{U}$.
  - Closed form solution:
    \[
    g^* = (-L^{-1} B^T 1)^T \quad \text{with} \quad \tilde{L} = \begin{bmatrix} L_1 & B \\ B^T & L_2 \end{bmatrix}
    \]

- **Matrix Completion (MC) [Nowak'10]**
  - Minimize \( \text{rank}(Z) \)
    subject to \( Z_y = [y_1,...,y_{|D|}], Z_x = [x_1,...,x_{|D|}], Z_{\bar{x}} = [\bar{x}_1,...,\bar{x}_{|D|}] \)
    where \( z = [x, y] \)
  - Replace \( \text{rank}(\cdot) \) by convex nuclear norm \( \|Z\|_* = \sum_{k=1}^{\min(|x|,|y|)} \sigma_k(Z) \)
  - Equality constraints replaced by squared losses
  - Find annotation for each test image and estimate local/pref annotations and retrieval
Methodologies

- Structural Learning (SL) [Joachims’05]
  - Margin maximization quadratic problem

\[
\begin{aligned}
\min_w & \|w\|^2 + C \sum_{i=1}^{[D]} \xi_i \\
\text{s.t.} & \quad w^\top \Psi(y_i, x_i) - w^\top \Psi(y, x_i) + \xi_i \geq \Delta(y_i, y), \quad y = 1...[D], \quad \forall y \in \{0, 1\}^Y \\
\xi_i & \geq 0, \quad i = 1...[D]
\end{aligned}
\]

where \( \Delta(y_i, y) = |y_i - y| \), \( \Psi(y, x) = x \otimes y \in \mathbb{R}^X \times Y \)

- Again, find global annotations for test images and then estimate local/pose annotations and retrieval

Results

- Average precision, recall and F1 score for annotations
  - Label based (global annotation)

\[
\begin{aligned}
p_{ga}(y) &= \frac{\sum_{\sigma=1}^{|\Sigma|} \mathbb{1}[y_{\sigma} = y]}{\sum_{\sigma=1}^{|\Sigma|} \mathbb{1}[y_{\sigma} = y]}, \quad r_{ga}(y) = \frac{\sum_{\sigma=1}^{|\Sigma|} \mathbb{1}[y_{\sigma} = y]}{|\Sigma|}, \quad f_{ga}(y) = \frac{2p_{ga}(y)r_{ga}(y)}{p_{ga}(y) + r_{ga}(y)}
\end{aligned}
\]

- Example based (global annotation)

\[
\begin{aligned}
p_{ge} &= \frac{1}{|\gamma|} \sum_{\sigma=1}^{|\gamma|} \mathbb{1}[y_{\sigma} = y], \quad r_{ge} = \frac{1}{|\gamma|} \sum_{\sigma=1}^{|\gamma|} \mathbb{1}[y_{\sigma} = y], \quad f_{ge} = \frac{1}{|\gamma|} \sum_{\sigma=1}^{|\gamma|} \frac{2(y_{\sigma} - y_{\sigma})}{|y_{\sigma} - y_{\sigma}|}
\end{aligned}
\]

- Mean average precision (MAP) for retrieval

\[
\begin{aligned}
p_{r}(q, Q) &= \frac{\sum_{i=1}^Q \delta(y_q - q) - 1^{-q}}{Q}, \quad \text{and} \quad r_{r}(q, Q) = \frac{\sum_{i=1}^Q \delta(y_q - q - 1^{-q})}{\sum_{i=1}^Q \delta(y_q - q - 1^{-q})},
\end{aligned}
\]
### Results

- **Global Annotation/Retrieval Results:**

<table>
<thead>
<tr>
<th>Models</th>
<th>Label-based global annotation</th>
<th>Example-based global annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Retrieval</td>
<td>MAP</td>
</tr>
<tr>
<td>RND</td>
<td>0.08 ± 0.06</td>
<td>0.06 ± 0.01</td>
</tr>
<tr>
<td>BoF</td>
<td>0.12 ± 0.05</td>
<td>0.14 ± 0.11</td>
</tr>
<tr>
<td>LP</td>
<td>0.11 ± 0.01</td>
<td>0.12 ± 0.02</td>
</tr>
<tr>
<td>LP-CC</td>
<td>0.11 ± 0.01</td>
<td>0.13 ± 0.02</td>
</tr>
<tr>
<td>LP-O</td>
<td>0.14 ± 0.02</td>
<td>0.19 ± 0.03</td>
</tr>
<tr>
<td>MC</td>
<td>0.17 ± 0.01</td>
<td>0.24 ± 0.03</td>
</tr>
<tr>
<td>SL</td>
<td>0.14 ± 0.01</td>
<td>0.18 ± 0.04</td>
</tr>
</tbody>
</table>

- **Local Annotation Results:**

<table>
<thead>
<tr>
<th>Models</th>
<th>Label-based local annotation</th>
<th>Example-based local annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Precision</td>
<td>Average Recall</td>
</tr>
<tr>
<td>RND</td>
<td>0.04 ± 0.01</td>
<td>0.04 ± 0.01</td>
</tr>
<tr>
<td>BoF</td>
<td>0.25 ± 0.08</td>
<td>0.05 ± 0.03</td>
</tr>
<tr>
<td>LP</td>
<td>0.12 ± 0.05</td>
<td>0.06 ± 0.02</td>
</tr>
<tr>
<td>LP-CC</td>
<td>0.08 ± 0.02</td>
<td>0.06 ± 0.01</td>
</tr>
<tr>
<td>ILP-O</td>
<td>0.06 ± 0.03</td>
<td>0.10 ± 0.03</td>
</tr>
<tr>
<td>MC</td>
<td>0.07 ± 0.01</td>
<td>0.03 ± 0.01</td>
</tr>
<tr>
<td>SL</td>
<td>0.09 ± 0.00</td>
<td>0.06 ± 0.01</td>
</tr>
</tbody>
</table>
### Results

- **Pose Annotation Results**

<table>
<thead>
<tr>
<th>Models</th>
<th>Label-based Pose annotation</th>
<th>Example-based Pose annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Precision</td>
<td>Average Recall</td>
</tr>
<tr>
<td>RND</td>
<td>0.00 ± .01</td>
<td>0.00 ± .01</td>
</tr>
<tr>
<td>BoP</td>
<td>0.01 ± .01</td>
<td>0.01 ± .01</td>
</tr>
<tr>
<td>LP</td>
<td>0.00 ± .00</td>
<td>0.00 ± .00</td>
</tr>
<tr>
<td>LP-CC</td>
<td>0.00 ± .00</td>
<td>0.00 ± .00</td>
</tr>
<tr>
<td>ILP</td>
<td>0.01 ± .01</td>
<td>0.01 ± .01</td>
</tr>
<tr>
<td>ILP-O</td>
<td>0.05 ± .04</td>
<td>0.08 ± .06</td>
</tr>
<tr>
<td>MC</td>
<td>0.00 ± .00</td>
<td>0.00 ± .00</td>
</tr>
<tr>
<td>SL</td>
<td>0.00 ± .00</td>
<td>0.00 ± .00</td>
</tr>
</tbody>
</table>

### Results

- **Global**
  - Judith,
  - Holofores head,
  - Maid servant

- **Local**
  - Holofores head (brown),
  - Judith (orange),
  - Maid servant (red)

- **Pose**
  - Holofores head (brown),
  - Judith (orange),
  - Maid servant (red)

- **Global**
  - Calvary,
  - Christ,
  - Cross

- **Local**
  - Christ (orange),
  - Cross (gray),

- **Pose**
  - Christ (orange)
Results

Fig. 5.

Nativity, Angels, Angels, Angels, Christ child, Mary, St. Joseph, St. Joseph

Global

Local

Pose

Fig. 6.

Annunciation, Dove, Gabriel, Mary, Wing, Wing

Global

Local

Pose

Resurrection, Christ, Soldier, Soldier, Soldier, Soldier

Global

Local

Pose

Holy Family, Christ child, Mary, St. Joseph, St. Joseph

Global

Local

Pose

Christ child (purple), Mary (green), St. Joseph (blue)

Christ child (purple), Mary (green), St. Joseph (blue)

Christ (orange), Soldier (purple), Soldier (yellow)

Christ (orange), Soldier (purple), Soldier (yellow)

Holy Family, Christ child, Mary, St. Joseph

Global

Local

Pose

Nativity, Angels, Angels, Angels, Christ child, Mary, Wing

Global

Local

Pose

Angels (purple), Christ child (blue), Mary (blue), Wing (gray)

Angels (purple), Christ child (blue), Mary (blue), Wing (gray)

Angels (purple), Christ child (blue), Mary (blue)


Fig. 5. Retrieval results of the **ILP-O**. Each row shows the top five matches to the following queries (from top to bottom): ‘**Holy Family**’, and ‘**Christ child**’. Below each image, it is indicated whether the image is annotated with the class.
Conclusions

• Inverted label propagation produces best results

• Small training sets for most of classes are a problem for inductive methods (BoF, SL)

• Artistic influence network – facilitates good results from random walk processes

• Face/person detector to improve results

• Sketch-based Interface and Modeling