Artistic Image Classification: An Analysis on the PRINTART Database

Gustavo Carneiro[1], Nuno Pinho da Silva[2],
Alessio Del Bue[3], and João Paulo Costeira[2,⋆]

[1] Australian Centre for Visual Technologies, The University of Adelaide, Australia
[2] Instituto de Sistemas e Robótica, Instituto Superior Técnico, Portugal
[3] PAVIS, Istituto Italiano di Tecnologia (IIT), Italy

Abstract. Artistic image understanding is an interdisciplinary research field of increasing importance for the computer vision and the art history communities. For computer vision scientists, this problem offers challenges where new techniques can be developed; and for the art history community new automatic art analysis tools can be developed. On the positive side, artistic images are generally constrained by compositional rules and artistic themes. However, the low-level texture and color features exploited for photographic image analysis are not as effective because of inconsistent color and texture patterns describing the visual classes in artistic images. In this work, we present a new database of monochromatic artistic images containing 988 images with a global semantic annotation, a local compositional annotation, and a pose annotation of human subjects and animal types. In total, 75 visual classes are annotated, from which 27 are related to the theme of the art image, and 48 are visual classes that can be localized in the image with bounding boxes. Out of these 48 classes, 40 have pose annotation, with 37 denoting human subjects and 3 representing animal types. We also provide a complete evaluation of several algorithms recently proposed for image annotation and retrieval. We then present an algorithm achieving remarkable performance over the most successful algorithm hitherto proposed for this problem. Our main goal with this paper is to make this database, the evaluation process, and the benchmark results available for the computer vision community.

1 Introduction

Artistic image understanding is a field of research that stimulates the development of interdisciplinary work. In this paper, we consider artistic image to be an artistic expression represented on a flat surface (e.g., canvas or sheet of paper) in the form of a painting, printing, or drawing. Even though we have observed a increasing interest in this area, there is still a lack of common evaluation

⋆ This work was supported by the FCT (ISR/IST plurianual funding) through the PIDDAC Program funds and FCT Project PRINTART (PTDC/EEA-CRO/098822/2008).

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part IV, LNCS 7575, pp. 143–157, 2012.

Fig. 1. Examples of a scene depicting the same artistic theme "The Annunciation". Figure (a) shows a real photo of the scene, while in figure (b) a painting is displayed, and (c) shows an art print.

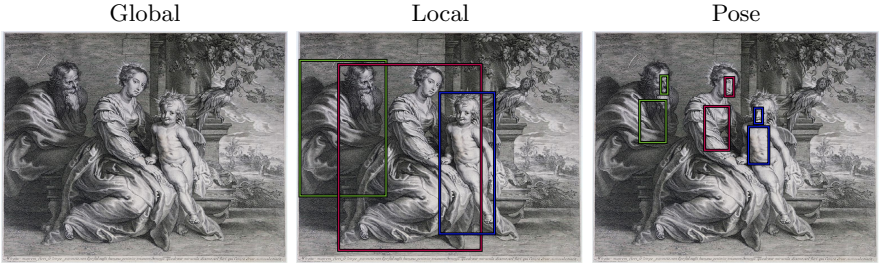

Fig. 2. Examples of the global, local and pose annotations made by the art historians. More training samples are provided in the supplementary material.

databases and procedures, similar to the ones found in photographic image retrieval and annotation, such as: Pascal VOC, Imagenet, TinyImages, Lotus Hill, SUN database to cite a few. Different from photographic images, art images can be better constrained based on compositional rules and themes. However, the texture and color patterns of visual classes (e.g., sky, sea, sand) are not consistently expressed in the artistic images, which makes the exploitation of low-level image features more challenging. In fact, current art image processing has shown that texture and color patterns in artistic images are more effectively used to classify painting styles [1] or artists [2] than to identify visual classes. For instance, Fig. 1 shows examples of a photo, a painting, and a print of a scene depicting the artistic theme "The Annunciation". Notice how the low-level features in the photographic image are more likely to successfully represent visual classes in the photo than in the artistic images.

We define artistic image understanding as a process that receives an artistic image and outputs a set of global, local and pose annotations. The global annotations consist of a set of artistic keywords describing the contents of the image. Local annotations comprise a set of bounding boxes that localize certain visual classes, and pose annotations consist of a set of body parts that indicate the

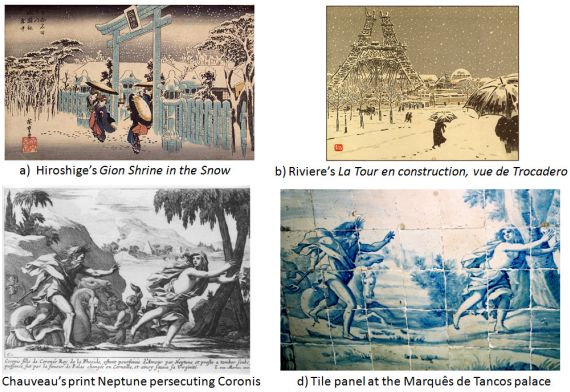

Fig. 3. Influence of Japanese art prints (a) on impressionist paintings (b), and of monochromatic art prints (c) on tile panel paintings (d)

pose of humans and animals in the image (see Fig. 2). Another process involved in the artistic image understanding is the retrieval of images given a query containing an artistic keyword. Systems developed for such end are of paramount importance to art historians for the task of analyzing artistic production, or can be part of an augmented reality method that provides information of an object of art given a digital picture of it.

A visual art form that is particularly important for the analysis of art images is printmaking. Printmaking is the process of creating prints from the impression that the print creator has of a painting (i.e., the print produced is similar to the original painting, but not identical). Cheap paper production and advancements in graphical arts resulted in an intensive use of printmaking methods over the last five centuries, which generated prints that have reached a significantly large number of artists. The main consequence of this wide availability of prints is their influence over several generations of artists, who have used them as a source of inspiration for their own production. For instance, Fig. 3 displays the influence of Japanese art prints on impressionist artists of the XIX century [3], and the influence of monochromatic art prints on artistic tile painters in Portugal. Therefore, a system that can automatically annotate and retrieve art prints has the potential to become a key tool for the understanding of the visual arts produced in the last five centuries.

In this paper, we present a new annotated database composed of artistic images that will be available for the computer vision community in order to start a comprehensive and principled investigation on artistic image understanding. Given the expert knowledge required for annotating this kind of images, it is not possible to use crowdsourcing tools (e.g., Amazon mechanical turk). Hence, art historians annotated 988 monochromatic artistic images, representing prints of religious themes made between the XV and XVII centuries in Europe. In this multi-label multi-class problem, 75 visual classes are annotated, from which 27 are related to the theme of the art image, and 48 are visual classes that can be

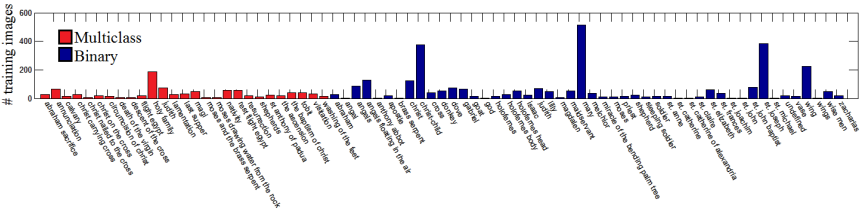

Fig. 4. Number of training images per class

localized in the image with bounding boxes. Out of these 48 classes, 40 visual classes have pose annotation, where 37 denote human subjects and 3 represent animal types. Figure 2 shows an example of the global, local and pose annotations produced by an art historian. We suggest error measures for the problems of global image annotation, image retrieval, local visual object detection, and pose estimation. Moreover, we test several methodologies and report their error measures that will be used as benchmarks for the problem. Specifically, we consider the following methodologies: random, bag of features [4], label propagation [5], inverted label propagation [6], matrix completion [7], and structural learning [8]. In particular, we introduce an improved inverted label propagation method that produces the best results, both in the automatic (global, local and pose) annotation and retrieval problems. This database will be freely available on the web [9], together with a table containing up-to-date results, a list of suggested error measures (with the respective code), and links to the evaluated techniques.

Literature Review. The current focus of art image analysis is on the forgery detection problem [10,2] and on the classification of painting styles [1]. The methodologies being developed can be regarded as adaptations of systems that work for photographic images, where the main changes are centered on the type of feature used and on spatial dependencies of local image descriptors. A particularly similar database to the one presented in this paper is the ancient Chinese painting data-set used for the multi-class classification of painting styles [11], which consists of monochromatic art images. Another important reference for our paper is the work by Yelizaveta et al. [12], which handles the multi-class classification of brush strokes, but they do not consider the multi-label problem being handled in our paper. Recently, Carneiro [6] shows a methodology for art image retrieval and *global* annotation, but he did not propose a database of artistic images, nor did he investigate local and pose annotation problems.

2 Database Collection and Evaluation Protocols

The artistic image database comprises 988 images with global, local and pose annotations (Fig. 2). All images have been collected from the Artstor digital image library [13], and annotated by art historians. The first stage consists of

a global annotation containing one multi-class problem (theme with 27 classes) and 48 binary problems (Fig. 4 shows the class names and the respective number of training images). All these 48 binary problems comprise visual classes that can be localized in the image with bounding boxes forming the local annotation, as depicted in the central frame of Fig. 2. Finally, out of these 48 visual classes, 37 are annotated with the pose of the human subject and 3 are annotated with animal pose. The pose annotation is composed of torso and head (both represented by a bounding box), as shown in the right frame of Fig. 2.

Notation. The training set is represented by $\mathcal{D} = \{(\mathbf{x}, \mathbf{y}, \mathcal{L}, \mathcal{P})_i\}_{i=1}^{|\mathcal{D}|}$, where $\mathbf{x}_i$ is a feature vector representing an image I_i, $\mathbf{y}_i$ is the global annotation of that image representing the M multi-class and binary problems, so $\mathbf{y}_i = [\mathbf{y}_i(1), ..., \mathbf{y}_i(M)] \in \{0,1\}^Y$, where each problem is denoted by $\mathbf{y}_i(k) \in \{0,1\}^{|\mathbf{y}_i(k)|}$ with $|\mathbf{y}_i(k)|$ denoting the dimensionality of $\mathbf{y}_i(k)$ (i.e., $|\mathbf{y}_i(k)| = 1$ for binary problems and $|\mathbf{y}_i(k)| > 1$ with $\|\mathbf{y}_l\|_1 = 1$ for multi-class problems). This means that binary problems involve an annotation that indicates the presence or absence of a visual class, while multi-class annotation regards problems that one and only one of the possible classes is present. The set $\mathcal{L}_i$ represents the local annotation of image I_i denoted by a set of bounding boxes, each related to one of the binary classes of $\mathbf{y}_i$. Specifically, we have $\mathcal{L}_i = \{\mathbf{l}_{i,j}\}_{j=1}^{|\mathcal{L}_i|}$ with $\mathbf{l}_{i,j} = [y, \mathbf{b}]$, where $y \in \{1, ..., Y\}$ represents the visual class of the bounding box, $\mathbf{b} = [\mathbf{z}, w, h]$ with $\mathbf{z} \in \Re^2$ being the top-left corner and w and h, the width and height of the box, respectively. Finally, the set $\mathcal{P}_i = \{\mathbf{p}_{i,j}\}_{j=1}^{|\mathcal{P}_i|}$ denotes the pose annotation of image I_i, where $\mathbf{p}_{i,j} = [y, \mathbf{b}^{head}, \mathbf{b}^{torso}]$, where $\mathbf{b}^{head}$ denotes the bounding box of the head, and $\mathbf{b}^{torso}$ is the bounding box of the torso annotation. An annotated test set is represented by $\mathcal{T} = \{(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}, \widetilde{\mathcal{L}}, \widetilde{\mathcal{P}})_i\}_{i=1}^{|\mathcal{T}|}$, but the annotations in the test set are used only for the purpose of methodology evaluation.

The label cardinality of the database, computed as $LC = \frac{1}{|\mathcal{D}|+|\mathcal{T}|} \sum_{i=1}^{|\mathcal{D}|+|\mathcal{T}|} \|\mathbf{y}_i\|_1$, is 4.22, while the label density $LD = \frac{1}{(|\mathcal{D}|+|\mathcal{T}|)Y} \sum_{i=1}^{|\mathcal{D}|+|\mathcal{T}|} \|\mathbf{y}_i\|_1$, is 0.05, where $Y = 75$ and $|\mathcal{D}| + |\mathcal{T}| = 988$.

2.1 Annotation and Retrieval Problems

For computing the error measures, 10 different training and test sets are available, with training sets comprising $|\mathcal{D}| = 889$ images (90% of the annotated images) and test sets with $|\mathcal{T}| = 99$ images (10% of the annotated images). The results are reported based on the performance computed over the test set $\mathcal{T}$ after training the methodology with the training set $\mathcal{D}$. Below, we define the error measures for the global annotation, retrieval, local and pose annotation.

Global Annotation. The global annotation process of a test image $\widetilde{\mathbf{x}}$ is achieved by finding $\mathbf{y}^*$ that solves the following optimization problem:

$$\begin{aligned} &\text{maximize } p(\mathbf{y}|\widetilde{\mathbf{x}}) \\ &\text{subject to } \mathbf{y} = [\mathbf{y}(1), ..., \mathbf{y}(M)] \in \{0,1\}^Y, \\ &\qquad \|\mathbf{y}(k)\|_1 = 1 \text{ for } \{k \in \{1, ..., M\} | |\mathbf{y}(k)| > 1\}, \end{aligned} \tag{1}$$

where $p(\mathbf{y}|\widetilde{\mathbf{x}})$ is a probability function that computes the confidence of annotating the test image $\widetilde{\mathbf{x}}$ with vector $\mathbf{y}$. We assess the *label-based global annotation* of each visual class y using the following precision, recall and F1 measures:

$$pga(y) = \frac{\sum_{i=1}^{|\mathcal{T}|}(\pi_y \odot \mathbf{y}_i^*)^\top \widetilde{\mathbf{y}}_i}{\sum_{i=1}^{|\mathcal{T}|} \pi_y^\top \mathbf{y}_i^*}, \; rga(y) = \frac{\sum_{i=1}^{|\mathcal{T}|}(\pi_y \odot \mathbf{y}_i^*)^\top \widetilde{\mathbf{y}}_i}{\sum_{i=1}^{|\mathcal{T}|} \pi_y^\top \widetilde{\mathbf{y}}_i}, \; fga(y) = \frac{2pga(y)rga(y)}{pga(y)+rga(y)}, \tag{2}$$

where $\pi_y \in \{0,1\}^Y$ is one at the y^{th} position and zero elsewhere, and $\odot$ denotes the element-wise multiplication operator. The values of $pga(y)$, $rga(y)$ and $fga(y)$ are averaged over the visual classes. Notice in (2) that we only assess the result class by class independently. We also need to measure the performance considering all the annotated classes jointly. The following *example-based global annotation* measures (precision, recall and F1) are used in order to assess the performance in multi-label problems [14]:

$$pge = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \frac{(\mathbf{y}_i^*)^\top \widetilde{\mathbf{y}}_i}{\|\mathbf{y}_i^*\|_1}, \; rge = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \frac{(\mathbf{y}_i^*)^\top \widetilde{\mathbf{y}}_i}{\|\widetilde{\mathbf{y}}_i\|_1}, \; fge = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|} \frac{2(\mathbf{y}_i^*)^\top \widetilde{\mathbf{y}}_i}{\|\mathbf{y}_i^*\|_1+\|\widetilde{\mathbf{y}}_i\|_1}. \tag{3}$$

Image Retrieval. The retrieval problem is defined as the most relevant test image returned from $\mathcal{T}$ given a query represented by a vector $\mathbf{q}$, as in:

$$\widetilde{\mathbf{x}}^* = \arg\max_{\widetilde{\mathbf{x}} \in \mathcal{T}} p(\widetilde{\mathbf{x}}|\mathbf{q}), \tag{4}$$

where $p(\widetilde{\mathbf{x}}|\mathbf{q})$ computes the probability of returning the image $\widetilde{\mathbf{x}} \in \mathcal{T}$ given the query vector $\mathbf{q} \in \{0,1\}^Y$. Although $\mathbf{q}$ can represent any combinations of classes, in this paper, we restrict $\mathbf{q}$ to have only one class (i.e., $\|\mathbf{q}\|_1 = 1$). The *label-based retrieval* is evaluated from the following precision and recall measures computed using the first $Q \leq |\mathcal{T}|$ images retrieved (sorted by $p(\widetilde{\mathbf{x}}|\mathbf{q})$ in (4) in descending order):

$$pr(\mathbf{q}, Q) = \frac{\sum_{i=1}^{Q} \delta(\widetilde{\mathbf{y}}^\top \mathbf{q} - \mathbf{1}^\top \mathbf{q})}{Q}, \text{ and } rr(\mathbf{q}, Q) = \frac{\sum_{i=1}^{Q} \delta(\widetilde{\mathbf{y}}^\top \mathbf{q} - \mathbf{1}^\top \mathbf{q})}{\sum_{i=1}^{|\mathcal{T}|} \delta(\widetilde{\mathbf{y}}^\top \mathbf{q} - \mathbf{1}^\top \mathbf{q})}, \tag{5}$$

where $\delta(.)$ is the Kronecker delta function. These precision and recall measures are used to compute the mean average precision (MAP), which is defined as the average precision over all queries, at the ranks that the recall changes.

Local Annotation. The local annotation aims at finding the bounding boxes of the visual classes present in the image. The following optimization problem finds the local annotation $\mathcal{L}^*$ given the test image and its global annotation:

$$\text{maximize } p(\mathcal{L}|\mathbf{y}, \widetilde{\mathbf{x}}), \tag{6}$$

where each k that $|\mathbf{y}(k)| = 1$ and $\mathbf{y}(k) = 1$ has a respective bounding box $\mathbf{l}_j^* \in \mathcal{L}^*$. The *label-based local annotation* of each visual class y is assessed with the following precision, recall and F1 measures [15]:

$$pla(y) = \frac{\sum_{i=1}^{|\mathcal{T}|} a(\mathbf{l}_i^*(y) \bigcap \widetilde{\mathbf{l}}_i(y))}{\sum_{i=1}^{|\mathcal{T}|} a(\widetilde{\mathbf{l}}_i(y))}, \; rla(y) = \frac{\sum_{i=1}^{|\mathcal{T}|} a(\mathbf{l}_i^*(y) \bigcap \widetilde{\mathbf{l}}_i(y))}{\sum_{i=1}^{|\mathcal{T}|} a(\widetilde{\mathbf{l}}_i(y))}, \; fla(y) = \frac{2pla(y)rla(y)}{pla(y)+rla(y)}, \tag{7}$$

where the function $a(\mathbf{l})$ returns the area (in pixels) of the bounding box defined by $\mathbf{l}$ (see above in Sec. 2), and operator $\bigcap$ returns the intersection between the bounding boxes from estimation $\mathbf{l}_i^*(y)$ and from ground truth $\widetilde{\mathbf{l}}_i(y)$ in the test image (note that both boxes are related to class y). The values of $pla(y)$, $rla(y)$ and $fla(y)$ are then averaged over the visual classes. Notice that in (7) we only assess the result class by class independently. We also need to measure the performance considering all the annotated classes jointly. The following *example-based local annotation* measures (precision, recall and F1) are used in order to assess the performance in multi-label problems:

$$\begin{gathered} ple = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\frac{a(\mathbf{l}_i^*(y)\bigcap\widetilde{\mathbf{l}}_i(y))}{a(\mathbf{l}_i^*(y))},\ rle = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\frac{a(\mathbf{l}_i^*(y)\bigcap\widetilde{\mathbf{l}}_i(y))}{a(\widetilde{\mathbf{l}}_i(y))}, \\ fle = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\frac{2(a(\mathbf{l}_i^*(y)\bigcap\widetilde{\mathbf{l}}_i(y)))}{a(\mathbf{l}_i^*(y))+a(\widetilde{\mathbf{l}}_i(y))}. \end{gathered} \tag{8}$$

Pose Annotation. Finally, for the pose annotation, we assume the knowledge of global and local annotations in order to arrive at the pose annotation $\mathcal{P}^*$, as follows:

$$\text{maximize } p(\mathcal{P}|\mathcal{L},\mathbf{y},\widetilde{\mathbf{x}}), \tag{9}$$

where each k that $|\mathbf{y}(k)| = 1$ and $\mathbf{y}(k) = 1$ has a respective bounding box $\mathbf{l}_j \in \mathcal{L}$, and the head and torso bounding boxes are within the local annotation bounding box. The *label-based pose annotation* of the *head* visual class is assessed with the following precision, recall and F1 measures [15]:

$$\begin{gathered} ppa(y,head) = \frac{\sum_{i=1}^{|\mathcal{T}|}a(\mathbf{p}_i^*(y,head)\bigcap\widetilde{\mathbf{p}}_i(y,head))}{\sum_{i=1}^{|\mathcal{T}|}a(\mathbf{p}_i^*(y,head))}, \\ rpa(y,head) = \frac{\sum_{i=1}^{|\mathcal{T}|}a(\mathbf{l}_i^*(y,head)\bigcap\widetilde{\mathbf{l}}_i(y,head))}{\sum_{i=1}^{|\mathcal{T}|}a(\widetilde{\mathbf{p}}_i(y,head))}, \\ fpa(y,head) = \frac{2ppa(y,head)rpa(y)}{ppa(y,head)+rpa(y,head)}, \end{gathered} \tag{10}$$

and similarly for *torso*, where the function $a(\mathbf{p}(y,head))$ returns the area (in pixels) of the bounding box defined by $\mathbf{p}$ (see above in Sec. 2), and operator $\bigcap$ returns the intersection between the bounding boxes from estimation $\mathbf{p}_i^*(y,head)$ and from ground truth $\widetilde{\mathbf{p}}_i(y,head)$ in test image i (note that both boxes are related to class y). The values of $ppa(y)$, $rpa(y)$ and $fpa(y)$ are then averaged over the visual classes. Notice in (10) that we only assess the result class by class independently. We also need to measure the performance considering all the annotated classes jointly. The following *example-based pose annotation* measures (precision, recall and F1) are used in order to assess the performance in multi-label problems:

$$\begin{gathered} ppe = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\sum_{m\in\{head,torso\}}\frac{a(\mathbf{p}_i^*(y,m)\bigcap\widetilde{\mathbf{p}}_i(y,m))}{a(\mathbf{p}_i^*(y,m))}, \\ rpe = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\sum_{m\in\{head,torso\}}\frac{a(\mathbf{p}_i^*(y,m)\bigcap\widetilde{\mathbf{p}}_i(y,m))}{a(\widetilde{\mathbf{p}}_i(y,m))}, \\ fpe = \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\sum_{y=1}^{Y}\sum_{m\in\{head,torso\}}\frac{2(a(\mathbf{l}_i^*(y,m)\bigcap\widetilde{\mathbf{l}}_i(y)))}{a(\mathbf{p}_i^*(y,m))+a(\widetilde{\mathbf{p}}_i(y,m))}. \end{gathered} \tag{11}$$

3 Image Annotation and Retrieval Procedures

In this section, we describe the image representation and the different methodologies used to solve for the annotation and retrieval problems.

3.1 Image Representation

The images are represented with the spatial pyramid [16] (with three levels), which is an extension of the bag of visual words [4], where each visual word is formed with a collection of local descriptors. The local descriptors are extracted with the scale invariant feature transform (SIFT) [17] using a uniform grid over the image and scale space in order to have 10000 descriptors per image. The vocabulary is built by gathering the descriptors from all images and running a hierarchical clustering algorithm with three levels, where each node in the hierarchy has 10 descendants [18]. This results in a directed tree with $1+10+100+1000 = 1111$ vertexes, and the image feature is formed by using each descriptor of the image to traverse the tree and record the path (note that each descriptor generates a path with 4 vertexes). The histogram of visited vertexes is weighted by the node entropy (i.e., vertexes that are visited more often receive smaller weights). The spatial pyramid representation is achieved by tiling the image in three levels, as follows: the first level comprises the whole image, the second level divides the image into 2×2 regions, and the third level breaks the image into 3×1 regions. This tiling has shown the best results in the latest Pascal VOC image classification competitions [19]. This means that there are 8 histograms describing an image, represented by $\mathbf{x} \in \Re^X$, where $X = 8 \times 1111$.

3.2 Methodologies

We explored different annotation methodologies that have recently shown state-of-the-art results in several photographic image annotation processes. Specifically, we evaluate the performance of inductive and transductive methodologies, and use a random annotation approach for comparison. For the inductive learning, we study the performance of bag of feature and structural learning approaches. The transductive methodology is tested with different types of label propagation methods.

Random. The random global annotation takes into consideration the priors of the visual classes as follows:

$$\mathbf{y}^*(k) = \begin{cases} \pi_1, & r < p(\mathbf{y}(k) = \pi_1) \\ \vdots & \\ \pi_{|\mathbf{y}(k)|}, & \sum_{j=1}^{|\mathbf{y}(k)|-1} p(\mathbf{y}(k) = \pi_j) \leq r < 1 \end{cases} \begin{array}{c} \text{Multiclass: } \{k : |\mathbf{y}(k)| > 1\} \\ \\ \\ \\ \end{array} \quad \text{Binary: } \{k : |\mathbf{y}(k)| = 1\}$$
$$, \mathbf{y}^*(k) = \begin{cases} 1, & r < p(\mathbf{y}(k) = 1) \\ 0, & \text{otherwise} \end{cases}, \tag{12}$$

where $r \sim \mathcal{U}(0,1)$ (with $\mathcal{U}(0,1)$ denoting the uniform distribution between 0 and 1), $p(\mathbf{y}(k) = \pi_j) = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbf{y}(k)_i^\top \pi_j$ (with $\pi_j = 1$ for binary problems and

$\pi_j \in \{0,1\}^{|\mathbf{y}(k)|}$ with zeros everywhere except at the j^{th} position). The retrieval is done by first computing the global annotations for the test images in the set $\mathcal{T}$, and then the images are ranked based on the Hamming distance between query and test image annotations, as in:

$$\Delta(\mathbf{q},\mathbf{y}) = \|\mathbf{q} - \mathbf{y}^*\|_1. \tag{13}$$

The local and pose annotations are achieved for each visual class by first selecting the training image with the smallest value for

$$i^* = \arg \min_{j \in \{1,...,|\mathcal{D}|\}} \Delta(\mathbf{y}^*, \mathbf{y}_j), \tag{14}$$

and assign $\mathcal{L}^* = \mathcal{L}_{i^*}$ and $\mathcal{P}^* = \mathcal{P}_{i^*}$. The acronym for this approach is **RND**.

Bag of Features. The bag of features model is based on Y support vector machine (SVM) classifiers using the one-versus-all training method. Specifically, we train the Y classifiers (each classifier for each label) $p(\mathbf{y}(k) = \pi_j | \widetilde{\mathbf{x}}, \theta_{SVM}(k,j))$, for $k \in \{1,...,M\}$, $j \in \{1,...,|\mathbf{y}(k)|\}$, $\pi_j \in \{0,1\}^{|\mathbf{y}(k)|}$ (with the j^{th} element equal to one and rest are zero), and the annotation and retrieval use the same methods in (1) and (4), respectively, replacing $p(\mathbf{y}|\widetilde{\mathbf{x}})$ by $p(\mathbf{y}(k) = \pi_j | \widetilde{\mathbf{x}}, \theta_{SVM}(j))$. The penalty factor of the SVM for the slack variables is determined via cross-validation, where the training set $\mathcal{D}$ is divided into a training and validation sets of 90% and 10% of $\mathcal{D}$, respectively. This model roughly represents the state-of-the-art approach for image annotation and retrieval problems [20]. The extension to the retrieval problem is based on (13), and the local and pose annotations follow the method in (14). The acronym for this approach is **BoF**.

Label Propagation. The label propagation encodes the similarity between pairs of images using the graph Laplacian, and estimate the annotations of test image using transductive inference. This method has been intensively investigated, and we only present the main developments, which are the following. Find the annotation matrix $\mathbf{F}^*$ using the following optimization problem [5]:

$$\begin{array}{l} \text{minimize} \;\; 0.5 \; \text{tr}(\mathbf{F}^\top(\mathbf{D} - \mathbf{W})\mathbf{F}) \\ \text{subject to } \mathbf{f}_i = \mathbf{y}_i, \;\; \text{for } i = 1,...,|\mathcal{D}| \end{array}, \tag{15}$$

where $\mathbf{W}, \mathbf{F}, \mathbf{D} \in \Re^{(|\mathcal{D}|+|\mathcal{T}|)\times(|\mathcal{D}|+|\mathcal{T}|)}$ with $\mathbf{W}_{ij} = \exp\{-0.5\,\|\mathbf{x}_i - \mathbf{x}_j\|_2^2/\sigma^2\}$ such that the index for the training set is from 1 to $|\mathcal{D}|$ and for the test set from $|\mathcal{D}|+1$ to $|\mathcal{D}|+|\mathcal{T}|$, $\mathbf{D}$ is a diagonal matrix with its (i,i)-element equal to the sum of the i^{th} row of $\mathbf{W}$, and tr(.) computes the trace of a matrix. This problem has the closed form solution $\mathbf{F}^* = \beta(\mathbf{I} - \alpha(\mathbf{D} - \mathbf{W}))^{-1}\mathbf{Y}$, where $\mathbf{I}$ denotes the identity matrix,and α and β are regularization parameters such that $\alpha + \beta = 1$. In the experiments,this approach is named **LP**. The problem in (15) has been extended in order to include label correlation [21,22], as follows

$$\text{minimize } 0.5\,\text{tr}(\mathbf{F}^\top(\mathbf{D}-\mathbf{W})\mathbf{F}) + (1-\mu)\text{tr}((\mathbf{F}-\mathbf{Y})\Lambda(\mathbf{F}-\mathbf{Y})) + \mu\text{tr}(\mathbf{F}\mathbf{C}\mathbf{F}^\top), \tag{16}$$

where Λ is a matrix containing ones in the diagonal from indices 1 to $|\mathcal{D}|$, and zero otherwise, and $\mathbf{C} \in [-1,1]^{Y \times Y}$ containing the correlation between classes.The problem in (16) has closed form solution $\mathbf{F}^* = (\mathbf{D} - \mathbf{W})^{-1}\mathbf{Y}(\mathbf{I} - \mu\mathbf{C})$, where μ is a regularization parameter. We represent this approach by **LP-CC** in the experiments. After finding $\mathbf{F}^*$, we need to define the values for $\mathbf{y}_i^*$ for each test image. We tried some alternatives present in the literature, but obtained the best performance with class mass normalization [23], which adjusts the class distributions to match the priors. The extension to the retrieval problem is based on (13), and the local and pose annotations follow the approach described in (14).

Inverted Label Propagation. By inverting the problem described in (15), it is possible to produce the global, local, and pose annotations simultaneously. Specifically, instead of inferring the labels of the test images (using matrix $\mathbf{F}$ in Eq. 15), the inverted label propagation returns a vector representing the probability of landing in one of the training images after starting the random walk process from a test image. Furthermore, the similarity between annotations (which in LP requires a reformulation of the problem) is incorporated in the adjacency matrix. Then, the annotation can be finalized using the training images annotations weighted by the probability of random walk process. Recently, Carneiro [6] has formulated the global annotation problem with the combinatorial harmonic (CH) approach [24], which computes the probability that a random walk starting at the test image $\widetilde{\mathbf{x}}$ first reaches each of the database samples $(\mathbf{x}, \mathbf{y}, \mathcal{L}, \mathcal{P})_i \in \mathcal{D}$. Assuming that the test image is represented by $\widetilde{\mathbf{x}}$, the adjacency matrix in this inverted problem is defined by taking into consideration both the image and label similarities, as in:

$$\mathbf{U}(j,i) = I_y(\mathbf{y}_i, \mathbf{y}_j) \times I_x(\mathbf{x}_i, \mathbf{x}_j) \times I_x(\mathbf{x}_j, \widetilde{\mathbf{x}}), \tag{17}$$

where $I_y(\mathbf{y}_i, \mathbf{y}_j) = \sum_{k=1}^{M} \lambda_k \times \mathbf{y}(k)_i^\top \mathbf{y}(k)_j$ (λ_k is the weight associated with the label k), and $I_x(\mathbf{x}_i, \mathbf{x}_j) = \sum_{d=1}^{X} \min(\mathbf{x}_i(d), \mathbf{x}_j(d))$ (i.e., this is the histogram intersection kernel over the spatial pyramid representation described in Sec. 3.1). Note that the matrix $\mathbf{U}$ in (17) is row normalized. The computation of the CH solution extends the adjacency matrix in (17), as in: $\widetilde{\mathbf{U}} = \begin{bmatrix} \mathbf{U} & \widetilde{\mathbf{u}} \\ \widetilde{\mathbf{u}}^T & 0 \end{bmatrix}$, where $\widetilde{\mathbf{u}}$ is the un-normalized initial distribution vector defined as $\mathbf{u} = [I_x(\mathbf{x}_1, \widetilde{\mathbf{x}}), ..., I_x(\mathbf{x}_{|\mathcal{D}|}, \widetilde{\mathbf{x}})]^\top$. Our goal is to find the distribution $\mathbf{g}^* \in \Re^{|\mathcal{D}|}$ ($\|\mathbf{g}^*\|_1 = 1$), representing the probability of first reaching each of the training images in a random walk procedure, where the labeling matrix $\mathbf{G} = \mathbf{I}$ (i.e., an $|\mathcal{D}| \times |\mathcal{D}|$ identity matrix) denotes a problem with $|\mathcal{D}|$ classes, with each training image representing a separate class. The estimation of $\mathbf{g}^*$ is based on the minimization of the following energy function:

$$E([\mathbf{G}, \mathbf{g}]) = \frac{1}{2} \left\| [\mathbf{G}, \mathbf{g}] \widetilde{\mathbf{L}} \begin{bmatrix} \mathbf{G}^T \\ \mathbf{g}^T \end{bmatrix} \right\|_2^2, \tag{18}$$

where $\widetilde{\mathbf{L}}$ is the Laplacian matrix computed from the the adjacency matrix $\widetilde{\mathbf{U}}$. This Laplacian matrix can be divided into blocks of the same sizes as in $\widetilde{\mathbf{U}}$, that is $\widetilde{\mathbf{L}} = \begin{bmatrix} \mathbf{L}_1 & \mathbf{B} \\ \mathbf{B}^T & \mathbf{L}_2 \end{bmatrix}$. Solving the following optimization problem produces $\mathbf{g}^*$ [24]:

$$\begin{array}{l} \text{minimize} \quad E([\mathbf{G}, \mathbf{g}]) \\ \text{subject to } \mathbf{G} = \mathbf{I}, \end{array} \tag{19}$$

which has the closed form solution [24]: $\mathbf{g}^* = (-\mathbf{L}_2^{-1}\mathbf{B}^T\mathbf{I})^\top$. Note that $\mathbf{g}^* \in [0,1]^{|\mathcal{D}|}$ and $\|\mathbf{g}^*\|_1 = 1$. In order to annotate the test image, one can use class mass normalization [6], but we propose an alternative way, which is to simply take the annotation of the training sample $\mathbf{y}_{i^*}, \mathcal{L}_{i^*}, \mathcal{P}_{i^*}$ with $i^* = \arg\max \mathbf{g}^*$. This allows to produce global, local and pose annotations, and the extension to the retrieval problem is based on (13). In the experiments, this approach is named **ILP-O**. Note that the original ILP [6] (with class mass normalization) is denoted by **ILP**.

Matrix Completion. The matrix completion formulation consists of forming a joint matrix with annotation and features $\mathbf{Z} = \begin{bmatrix} \mathbf{Z_y} & \mathbf{Z_{y^*}} \\ \mathbf{Z_x} & \mathbf{Z_{\widetilde{x}}} \end{bmatrix}$, where the goal is to find the values for $\mathbf{Z_{y^*}} = [\mathbf{y}_1^* ... \mathbf{y}_{|\mathcal{T}|}^*]$ giving [7]:

$$\begin{array}{l} \text{minimize} \quad \text{rank}(\mathbf{Z}) \\ \text{subject to } \mathbf{Z_y} = [\mathbf{y}_1 ... \mathbf{y}_{|\mathcal{D}|}],\ \mathbf{Z_x} = [\mathbf{x}_1 ... \mathbf{x}_{|\mathcal{D}|}],\ \mathbf{Z_{\widetilde{x}}} = [\widetilde{\mathbf{x}}_1 ... \widetilde{\mathbf{x}}_{|\mathcal{T}|}]. \end{array} \tag{20}$$

In (20), the non-convex minimization objective function rank is replaced by the convex nuclear norm $\|\mathbf{Z}\|_* = \sum_{k=1}^{min\{|\mathcal{D}|, Y+X\}} \sigma_k(\mathbf{Z})$, where the $\sigma_k(\mathbf{Z})$ are the singular values of $\mathbf{Z}$. Moreover, the equality constraints for $\mathbf{Z_x}$ and $\mathbf{Z_{\widetilde{x}}}$ are replaced by squared losses, and the one for $\mathbf{Z_y}$ is relaxed to a logistic loss. After finding $\mathbf{Z_{y^*}}$, we need to define the values for $\mathbf{y}_i^*$ for each test image, and we obtained the best results with class mass normalization [23]. This approach is extended for the retrieval problem using (13), and the local and pose annotations follow the approach described above in (14). This approach is represented by the acronym **MC** in the experiments.

Structural Learning. The structural learning formulation follows the structured SVM implementation [8], which is based on the margin maximization quadratic problem, defined by:

$$\begin{array}{ll} \min\limits_{\mathbf{w},\xi} & \|\mathbf{w}\|^2 + C\sum\limits_{i=1}^{|\mathcal{D}|}\xi_i \\ s.t. & \mathbf{w}^\top\Psi(\mathbf{y}_i,\mathbf{x}_i) - \mathbf{w}^\top\Psi(\mathbf{y},\mathbf{x}_i) + \xi_i \geq \Delta(\mathbf{y}_i,\mathbf{y}),\ \ i = 1...|\mathcal{D}|,\ \ \forall \mathbf{y} \in \{0,1\}^Y, \\ & \xi_i \geq 0,\ \ i = 1...|\mathcal{D}| \end{array} \tag{21}$$

where $\Delta(\mathbf{y}_i, \mathbf{y}) = \|\mathbf{y}_i - \mathbf{y}\|_1$ (13), $\Psi(\mathbf{y}, \mathbf{x}) = \mathbf{x} \otimes \mathbf{y} \in \Re^{X \times Y}$ (i.e., this is a tensor product combining the vectors $\mathbf{x}$ and $\mathbf{y}$ by replication the values of $\mathbf{x}$ in

every dimension $y \in \{1, ..., Y\}$ where $\mathbf{y}^\top \pi_y = 1$), C is penalty for non-separable points, and ξ_d denotes the slack variables to deal with non-separable problems. The retrieval problem is based on (13), and the local and pose annotations follow (14). We represent this approach with the acronym **SL** in the experiments.

4 Experiments

In the experiments, we first compare the results of the global annotation and retrieval using all methods listed in Sec. 3.2 with the 10-fold cross validation experimental setup described in Sec. 2. For the **BoF**, we used the code implemented by Vedaldi and Fulkerson [25]. We implemented the code for **LP** following the algorithm by Zhou et al. [5]. For **LP-CC** we used the method by Wand et al. [21]. For **ILP** we follow the methodology by Carneiro [6], which was extended in this paper to produce the **ILP-O**. The **MC** was implemented based on the code MC-1 by Goldberg et al. [7], and for the **SL**, we used the code SVM^{struct} available from the page *svmligh.joachims.org/svm_ struct.html*. All regularization parameters in the algorithms above are learned via cross validation.

5 Discussion and Conclusions

According to the experiments, our extension of the inverted label propagation produces the best results. However, we note that the small training sets do not allow the inductive methodologies to build robust models for the majority of visual classes, and we believe that this is the main reason why **BoF** and **SL** do not produce the best results. We believe, that the superior performance of the inverted linear propagation is explained by the similar images from the same theme, containing the similar composition, visual classes and setting. Such similarities in art images arise from the artists' influence network. Therefore, given that the random walk process is highly likely to select the most similar images, the global annotation is often correct for the query image. The results for the local and pose annotation present an interesting challenge for the community. For

Table 1. Retrieval and global annotation performances in terms of the average ± standard deviation of measures (2)-(5) computed in a 10-fold cross validation experiment (the best performance for each measure is highlighted).

	Retrieval	Label-based global annotation			Example-based global annotation		
Models	Label MAP	Average Precision	Average Recall	Average F1	Average Precision	Average Recall	Average F1
RND	0.08 ± .06	0.06 ± .01	0.07 ± .01	0.06 ± .01	0.26 ± .02	0.21 ± .01	0.22 ± .01
BoF	0.12 ± .05	0.14 ± .11	0.10 ± .06	0.11 ± .08	0.35 ± .03	0.26 ± .08	0.30 ± .05
LP	0.11 ± .01	0.12 ± .02	0.12 ± .02	0.12 ± .02	0.32 ± .03	0.28 ± .02	0.26 ± .02
LP-CC	0.11 ± .01	0.13 ± .02	0.14 ± .02	0.13 ± .02	0.27 ± .03	0.26 ± .03	0.25 ± .03
ILP	0.14 ± .02	0.19 ± .03	**0.35 ± .03**	0.25 ± .04	0.24 ± .02	**0.48 ± .05**	0.30 ± .02
ILP-O	**0.18 ± .04**	**0.26 ± .05**	0.26 ± .05	**0.26 ± .05**	**0.39 ± .03**	0.39 ± .04	**0.38 ± .03**
MC	0.17 ± .01	0.24 ± .03	0.11 ± .02	0.15 ± .02	0.37 ± .02	0.28 ± .02	0.32 ± .02
SL	0.14 ± .01	0.18 ± .04	0.14 ± .03	0.16 ± .03	0.34 ± .04	0.31 ± .04	0.32 ± .04

Table 2. Local Annotation performance in terms of the average ± standard deviation of measures (7)-(8) computed in a 10-fold cross validation experiment (the best performance for each measure is highlighted).

	Label-based local annotation			Example-based local annotation		
Models	Average Precision	Average Recall	Average F1	Average Precision	Average Recall	Average F1
RND	0.04 ± .01	0.04 ± .01	0.04 ± .01	0.13 ± .03	0.18 ± .04	0.15 ± .02
BoF	**0.25 ± .08**	0.05 ± .03	0.07 ± .03	**0.28 ± .05**	0.17 ± .06	0.20 ± .04
LP	0.12 ± .05	0.06 ± .02	0.08 ± .02	0.21 ± .02	0.19 ± .04	0.20 ± .02
LP-CC	0.08 ± .02	0.06 ± .01	0.07 ± .01	0.12 ± .02	0.17 ± .04	0.14 ± .02
ILP	0.06 ± .03	0.10 ± .03	0.07 ± .03	0.13 ± .02	0.19 ± .03	0.16 ± .02
ILP-O	0.15 ± .05	**0.16 ± .05**	**0.15 ± .05**	0.21 ± .03	**0.24 ± .03**	**0.23 ± .03**
MC	0.07 ± .01	0.03 ± .01	0.04 ± .01	0.12 ± .03	0.14 ± .06	0.13 ± .03
SL	0.09 ± .00	0.06 ± .01	0.07 ± .01	0.18 ± .03	0.20 ± .04	0.19 ± .01

Table 3. Pose Annotation performance in terms of the average ± standard deviation of measures (10)-(11) computed in a 10-fold cross validation experiment (the best performance for each measure is highlighted).

	Label-based Pose annotation			Example-based Pose annotation		
Models	Average Precision	Average Recall	Average F1	Average Precision	Average Recall	Average F1
RND	0.00 ± .01	0.00 ± .01	0.00 ± .01	0.00 ± .02	0.00 ± .01	0.00 ± .01
BoF	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01
LP	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00
LP-CC	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00
ILP	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01	0.01 ± .01
ILP-O	**0.05 ± .04**	**0.08 ± .06**	**0.06 ± .05**	**0.06 ± .02**	**0.07 ± .02**	**0.06 ± .02**
MC	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00
SL	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00	0.00 ± .00

instance, exploring context cues may improve these results. Another point that can be explored is the use of people and face detectors in art images (we applied several state-of-the-art people and face detectors, but only obtained uninspiring results). In order to stimulate even more the research in this sub-field, we plan to add the delineation of arms and legs for the pose annotation. One final point, which is not evaluated in this work, concerns the image representation. Recently, wavelets produced excellent results on the forgery detection problem [2], but a more systematic comparison to other features is still necessary.

In conclusion, we believe that this database has the potential to spur a new sub-field of art image analysis within the computer vision community. The error measures and results provided can be used by the community to assess the progress made in this area. We believe that proper art image understanding has the potential to influence a more complete general image understanding.

The Table 1 shows the results (2)-(5) described for the global annotations process. The local annotation results explained in (7)-(8) are shown in Tab. 2, and the experimental results for the pose annotation are displayed in Tab. 3 using the measures (10)-(11). Figures 5 and 6 shows examples of retrieval and annotation results produced by the proposed **ILP-O**.

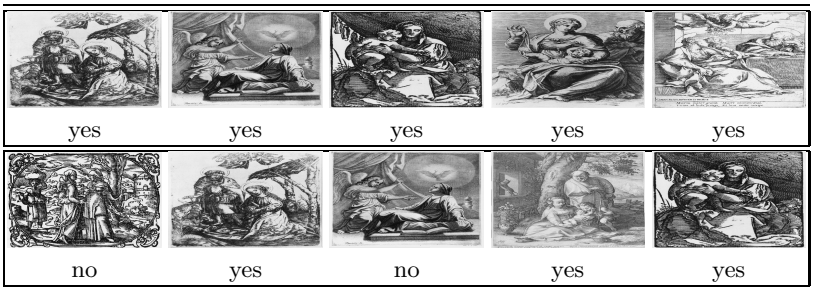

Fig. 5. Retrieval results of the **ILP-O**. Each row shows the top five matches to the following queries (from top to bottom): *'Holy Family'*, and *'Christ child'*. Below each image, it is indicated whether the image is annotated with the class.

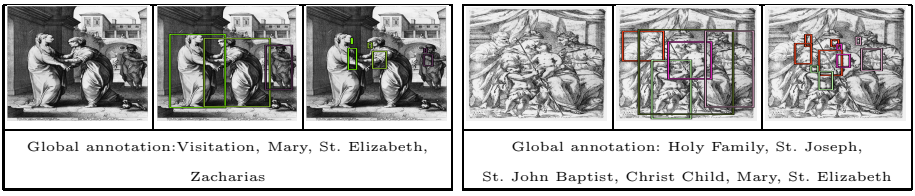

Fig. 6. Annotation result of **ILP-O**. Note that the global annotation shown produced a perfect match with respect to the art historian's annotation.

Acknowledgments. The authors thank D. Lázaro and R. Carvalho for their help with the art history issues. We acknowledge the matrix completion code by R. Cabral. Finally, we thank D. Lowe for valuable suggestions on the development of this work.

References

1. Graham, D., Friedenberg, J., Rockmore, D., Field, D.: Mapping the similarity space of paintings: image statistics and visual perception. Visual Cognition 18, 559–573 (2010)
2. Li, J., Yao, L., Hendriks, E., Wang, J.: Rhythmic brushstrokes distinguish van gogh from his contemporaries: Findings via automated brushstroke extraction. IEEE TPAMI (accepted for publication in 2012)
3. Baumann, F., et al.: Degas Portraits. Merrell Holberton, London (1994)
4. Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1–22 (2004)
5. Zhou, D., Bousquet, O., Lal, T., Weston, J., Scholkopf, B.: Learning with local and global consistency. In: NIPS, pp. 321–328 (2004)
6. Carneiro, G.: Graph-based methods for the automatic annotation and retrieval of art prints. In: Proceedings of the ACM ICMR (2011)
7. Goldberg, A.B., Zhu, X., Recht, B., Xu, J., Nowak, R.D.: Transduction with matrix completion: Three birds with one stone. In: NIPS, pp. 757–765 (2010)

8. Tsochantaridis, I., Joachims, T., Hofmann, T., Altun, Y.: Large margin methods for structured and interdependent output variables. JMLR 6, 1453–1484 (2005)
9. http://printart.isr.ist.utl.pt
10. Johnson, C., Hendriks, E., Berezhnoy, I., Brevdo, E., Hughes, S., Daubechies, I., Li, J., Postma, E., Wang, J.: Image processing for artistic identification: Computerized analysis of Vincent Van Goghs brushstrokes. IEEE Sig. Proc. Ma., 37–48 (2008)
11. Li, J., Wang, J.: Studying digital imagery of ancient paintings by mixtures of stochastic models. IEEE Trans. Image Processing 13, 340–353 (2004)
12. Yelizaveta, M., Tat-Seng, C., Jain, R.: Semi-supervised annotation of brushwork in paintings domain using serial combinations of multiple experts. In: ACM Multimedia, pp. 529–538 (2006)
13. http://www.artstor.org
14. Nowak, S., Lukashevich, H., Dunker, P., Rüger, S.: Performance measures for multilabel evaluation: a case study in the area of image classification. In: Multimedia Information Retrieval, pp. 35–44 (2010)
15. Everingham, M., Van Gool, L.J., Williams, C., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. IJCV 88, 303–338 (2010)
16. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features, spatial pyramid matching for recognizing natural scene categories. In: CVPR (2006)
17. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV 60, 91–110 (2004)
18. Nistér, D., Stewénius, H.: Scalable recognition with a vocabulary tree. In: CVPR, pp. 2161–2168 (2006)
19. http://pascallin.ecs.soton.ac.uk/challenges/VOC/
20. de Sande, K.V., Gevers, T., Smeulders, A.: The university of amsterdams concept detection system at imageclef 2009. In: CLEF Working Notes 2009 (2009)
21. Wang, H., Huang, H., Ding, C.: Image annotation using multi-label correlated green's function. In: ICCV, pp. 2029–2034 (2009)
22. Zha, Z., Mei, T., Wang, J., Wang, Z., Hua, X.: Graph-based semi-supervised learning with multi-label. In: ICME, pp. 1321–1324 (2008)
23. Zhu, X., Ghahramani, Z., Lafferty, J.: Semi-supervised learning using gaussian fields and harmonic functions. In: ICML, pp. 912–919 (2003)
24. Grady, L.: Random walks for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 28, 1768–1783 (2006)
25. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008)