Of Gods and Goats: Weakly Supervised Learning of Figurative Art

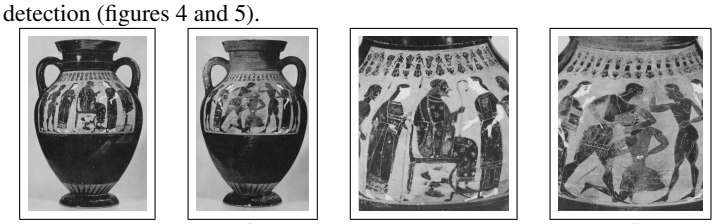
Elliot J. Crowley
elliot@robots.ox.ac.uk
Andrew Zisserman
az@robots.ox.ac.uk

Department of Engineering Science
University of Oxford

The objective of this paper is to automatically annotate the decorations on Greek vases with the gods and animals depicted there, given a dataset [1] of tens of thousands of images with associated text descriptions; such images often require expert labelling knowledge. Several papers have considered this 'words and pictures' problem [2] of automatically annotating image regions given only images and associated text. However, here there are additional challenges of: noisy supervision, non-naturalistic renderings, high intra-class variability, and low inter-class variability. Figure 1 shows a typical vase entry in the dataset.

Motivation. Automatic annotations of gods and animals is a very useful resource for archaeologists studying classical art, as assembling this type of material (e.g. all depictions of the god Zeus, aligned and size normalized from thousands of vases) manually would take quite some time. Apart from being a useful computer vision application, the method developed is directly applicable to other such art/archaeological collections with similar annotations, and the algorithm of progressive reduction of visual search space is useful in general for 'words and pictures' datasets.

Summary of method. To solve the annotation problem we propose a weakly supervised learning approach that proceeds in a number of stages. The key idea is that each stage strengthens the supervisory information available to allow for successful learning – this is akin to increasing the signal to noise ratio in signal processing. We proceed in three steps: (i) we use text mining methods to select sets of images that are visually consistent for a god depicted in a particular style (figure 2); (ii) we employ a form of multiple instance learning [5] to identify the image regions depicting the god in those images where he/she appears; (iii) the image regions are used to train a DPM sliding window detector [3] and all images in the dataset associated with the god can then be annotated by object category detection (figures 4 and 5).



(a) (b) (c) (d)

A. Theseus and minotaur, with youths and women; B. Zeus seated, between women and onlookers;

Figure 1: **An example vase entry.** It consists of four images (a, b, c, d) and two text descriptions (A, B). Image (c) is a detail of (a) and both images correspond to description B. Similarly, image (d) is a detail of (b) and both correspond to description A. None of this information (correspondences, details) is provided. Affine transformations are automatically estimated between the images associated with each vase in order to determine and represent each vase by only the most zoomed images (c and d in this case).

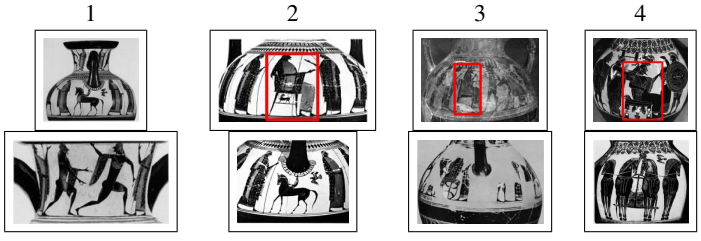


Figure 2: **Learning visually consistent clusters.** There is a large amount of variation in the depiction of each god. To overcome this difficulty we separate vase sets into groups corresponding to a similar depiction. We mine for verbs and non-person nouns in the text descriptions of each vase and greedily assign each vase to a cluster associated with such a word. A subset of the vases for the 'Zeus Seated' cluster can be seen above (4 vases shown from a cluster of 170 vases). Zeus, where present, is indicated by a red rectangle. Note that the data is noisy: if perfect, Zeus would be expected to be in one of the images for Vase no. 1 but he is not.

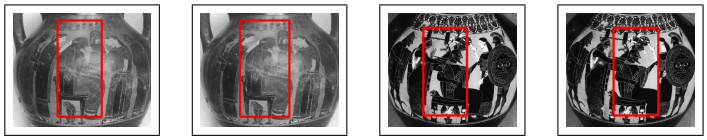
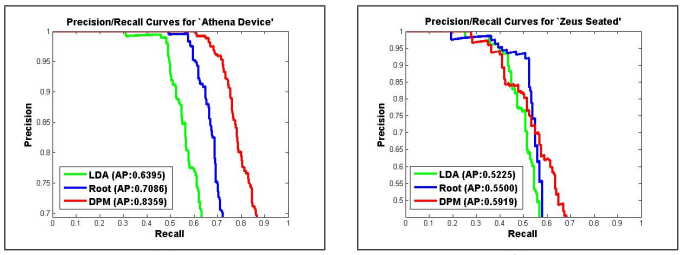


Figure 3: **Obtaining Candidate Regions**. To find the image regions within each visually consistent cluster containing the god we propose candidate regions and then determine which occur in other images within the cluster, but not outside of it by assessing the discriminative performance of an Exemplar-LDA based detector [4] trained on that region. The highest scoring regions are then re-ranked based on their visual consistency with each other and aligned. The windows corresponding to these top ranked detectors are suitable positive examples of the god. The regions obtained in this way for 'Zeus Seated' can be seen above.



(a) (b)

Figure 4: **Precision/Recall curves for (a) 'Athena Device' and (b) 'Zeus Seated'**. The green curves are for LDA models trained with the same positive examples as the DPM, the blue curves are for the DPM root-filters and the red curves are for the full DPM.

Figure 5: **Top Detections**. Above: Athena Device, Below: Zeus Seated.

[1] The Beazley Archive of Classical Art Pottery Database, July 2013. URL `http://www.beazley.ox.ac.uk/pottery`.

[2] K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan. Matching words and pictures. 3:1107–1135, February 2003.

[3] P. Felzenszwalb, R. Grishick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. 2010.

[4] B. Hariharan, J. Malik, and D. Ramanan. Discriminative decorrelation for clustering and classification. In *Proc. ECCV*, 2012.

[5] O. Maron and T. Lozano-Pérez. A framework for multiple-instance learning. In *Proc. NIPS*, pages 570–576. MIT Press, 1998.