

The State of the Art: Object Retrieval in Paintings using Discriminative Regions

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

Visual Geometry Group
 Department of Engineering Science
 University of Oxford

The objective of this work is to recognize object categories (such as animals and vehicles) in paintings, whilst learning these categories from natural images. This is a challenging problem given the substantial differences between paintings and natural images, and variations in depiction of objects in paintings [5] – see figure 1.

Contributions. (i) We show that object category classifiers learnt using Fisher Vectors [4] extracted from natural images can retrieve paintings containing that category with some success; (ii) we then introduce a method of re-ranking these retrieved paintings based on spatial consistency of Mid-Level Discriminative Patch (MLDP) correspondences with the original training images and show that the precision of the top ranked paintings (i.e. the ones that would appear on the first webpage in an image search) can be significantly improved using this method.

Motivation. Obtaining paintings with a particular object is of interest to Art Historians who currently find paintings manually or from memory. They can then study the change in the depiction style over time or determine when an object first appeared in paintings.

Summary of method. Object category classifiers are learnt from training sets of natural images (e.g. PASCAL VOC) and applied to paintings. The top ranked paintings for each category are re-ranked based on their spatial consistency with the natural images as follows: (i) discriminative regions are extracted from the natural images using the method of Aubry *et al.* [2] (figure 2); (ii) these regions are used to learn LDA [3] classifiers which are applied as sliding window detectors to the top ranked paintings to find matching regions and a RANSAC style algorithm is used to remove outlying matches (figure 3); (iii) each painting is scored by the maximum number of inlying matches shared with a natural image and are re-ranked accordingly (figure 4).

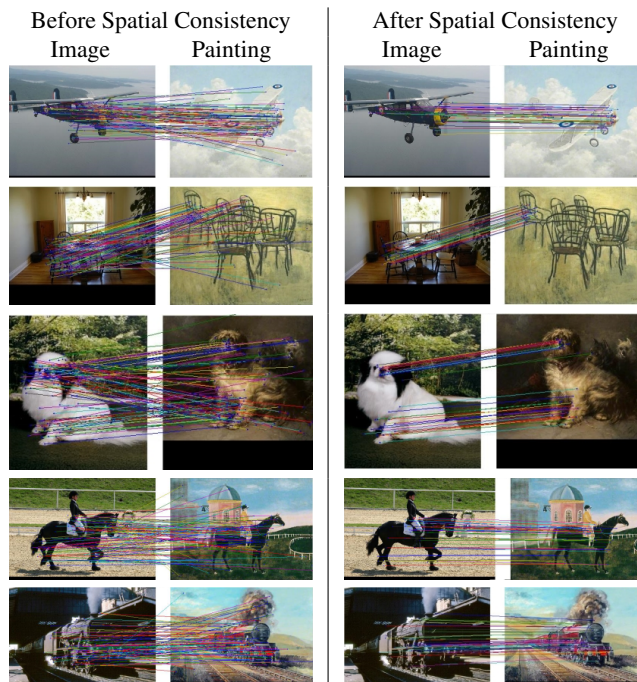


Figure 3: **Obtaining correspondences.** Each sliding window detector obtained from a mid-level discriminative patch (MLDP) on the natural image, defines a possible correspondence at the highest scoring detection window on each painting. This gives a set of provisional correspondences between each image-painting pair for an object. For each pair a RANSAC style algorithm is used to select a subset of these correspondences that are spatially consistent, and the image-painting pair is scored based on the size of this subset. Note, that the MLDP correspondences are able to generalize slightly over viewpoint, intra-class differences, and between natural images and paintings.

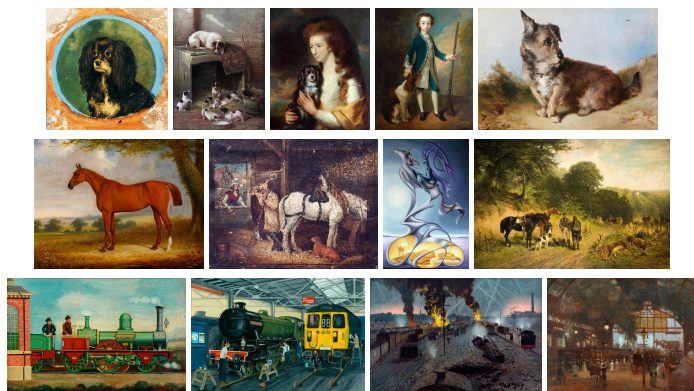


Figure 1: **Example Paintings** from top to bottom row, those containing: dog, horse, train. Objects have a variety of sizes, poses and depictive styles, and can be partially occluded or truncated. The paintings have been obtained from [1].

Rank	1	2	3	4	5
Dog Classifier					
Dog MLDP					
Sheep Classifier					
Sheep MLDP					

Figure 4: Top 5 ranked paintings before and after re-ranking using MLDPs for the dog and sheep category. A green border indicates a correct classification and a red border an incorrect one. Classification results are improved using our method.

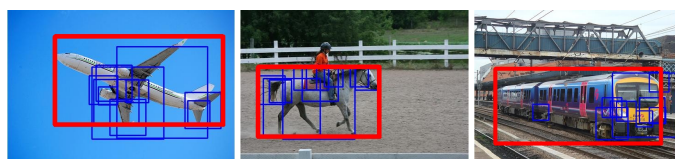


Figure 2: **Discriminative regions are extracted from natural images.** For a given object category square regions are sampled from each object ROI at a variety of scales; each region is represented by a HOG descriptor and assigned a score based on how much that HOG descriptor differs from the mean HOG descriptor of many natural images. Regions that score the highest this way are retained and are considered to be MLDPs for the object. The figure above shows a subset of discriminative regions (blue) overlapping with PASCAL VOC ROIs (red) for several images. Notice that informative areas of the objects are picked out such as a horse’s head, and even within the ROI no indiscriminate background patches are selected.

[1] BBC – Your Paintings. <http://www.bbc.co.uk/arts/yourpaintings/>.
 [2] M. Aubry, B. Russell, and J. Sivic. Painting-to-3D model alignment via discriminative visual elements. In *ACM Transactions of Graphics*, 2013.
 [3] B. Hariharan, J. Malik, and D. Ramanan. Discriminative decorrelation for clustering and classification. In *Proc. ECCV*, 2012.
 [4] F. Perronnin, J. Sánchez, and T. Mensink. Improving the Fisher kernel for large-scale image classification. In *Proc. ECCV*, 2010.
 [5] Q. Wu and P. Hall. Modelling visual objects invariant to depictive style. In *Proc. BMVC.*, 2013.