Enhancing Exemplar SVMs using Part Level Transfer Regularization

Yusuf Aytar
yusuf@robots.ox.ac.uk
Andrew Zisserman
az@robots.ox.ac.uk

Department of Engineering Science
University of Oxford
Parks Road
Oxford, OX1 3PJ, UK

Content based image retrieval (CBIR), the problem of searching digital images in large databases according to their visual content, is a well established research area in computer vision. In this work we are particularly interested in retrieving subwindows of images which are similar to the given query image, i.e. the goal is detection rather than image level classification. The notion of *similarity* is defined as being the same object class but also having similar viewpoint (e.g. frontal, left-facing, rear etc.). A query image can be a part of an object (e.g. head of a side facing horse), a complete object (e.g. frontal car image), or a composition of objects (visual phrases, e.g. person riding a horse). For instance, given a query of a horse facing left, the aim is to retrieve any left facing horse (intra-*class* variation) which might be walking or running with different feet formations (exemplar deformation).

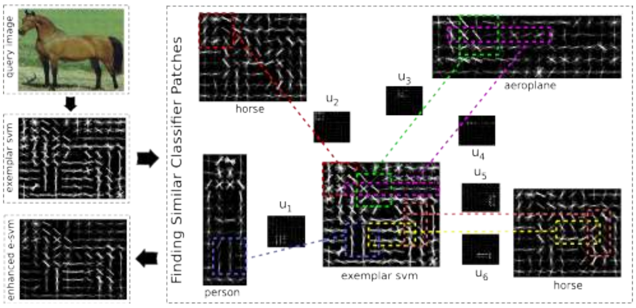

Figure 1: **Overview of the EE-SVM learning procedure.** The box on the right shows mining classifier patches from existing classifiers by matching subparts of E-SVM trained from the given query image. Comparing E-SVM and EE-SVM, better suppression of the background can be seen from the visualized classifiers. Note, here and in the rest of the paper we only visualize the positive components of the HOG classifier.

Recently exemplar SVMs (E-SVM) [1], where an SVM is trained with only a single positive sample, have found applications in the areas of CBIR [2] and object detection [1]. Since the E-SVM is trained from a single positive sample (together with many negatives), it is specialized to that given sample. This means that it can be strict (on viewpoint for example), and the negatives give some background suppression. However, the single positive is also a limitation: only so much can be learnt about the foreground of the query, and more significantly it can lead to lack of generalization. In our context, *generalization* refers to intra-class variation and deformation whilst maintaining the viewpoint. Learning such generalization from a single positive is challenging given the lack of examples of allowable deformations and intra-class variation.

In this work we propose a transfer learning approach for boosting the performance of E-SVMs using part-like patches of previously learned classifiers. The formulation softly constrains the learned template to be constructed from classifiers that have been fully trained (i.e. using many positives). For instance, the neck of a horse can be transferred from the tail of an aeroplane (see figure 1), or a jumping bike can borrow part of wheel patches from regular side facing bike or motorbike classifiers (see figure 2). The intuitive reason behind borrowing patches from other well trained classifiers is that these classifier patches bring with them a better sense of discriminative features and background suppression. The classifier patches also bring some generalization properties which an E-SVM may lack because it is only trained on a single positive sample. The result of the transfer learning is an enhancement of background suppression and tolerance to intra-class variation. However, these enhancements incurs no (significant) additional cost in learning and testing. We term the boosted E-SVM, Enhanced Exemplar SVM (EE-SVM).

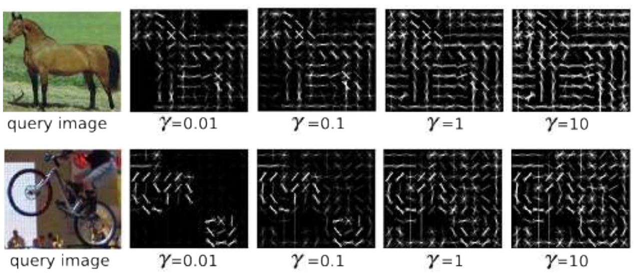

Figure 2: **Two limits of EE-SVM from reconstruction($\gamma = 0.01$) to E-SVM($\gamma = 10$).** Learned EE-SVM templates with varying γ values are displayed. λ is fixed to 1.

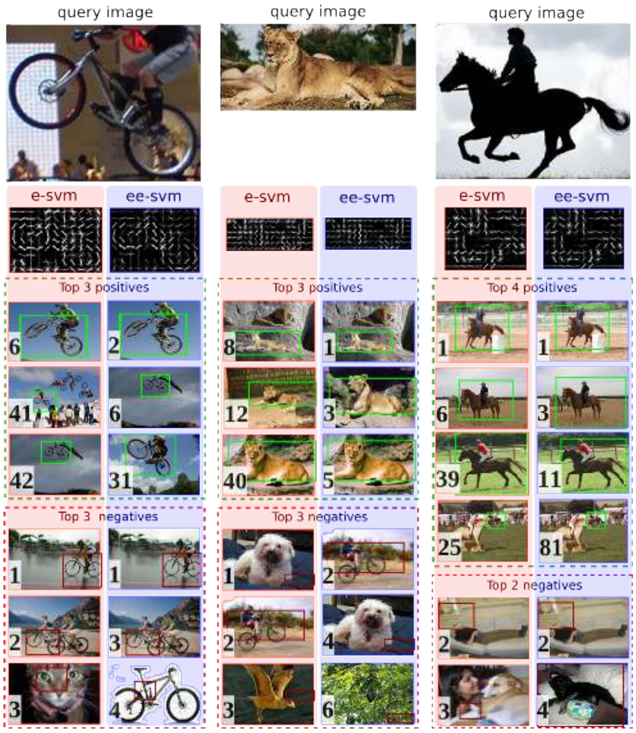

Figure 3: Retrieval of unusual poses on ImageNet. A visual phrase retrieval is also shown on the rightmost column.

We make the following contributions: (a) introduce the EE-SVM objective function; (b) demonstrate the improvement in performance of EE-SVM over E-SVM for CBIR; and, (c) show that there is an equivalence between transfer regularization and feature augmentation for this problem and others, with the consequence that the new objective function can be optimized using standard libraries.

Enhanced E-SVM incorporates the part based transfer regularization using the objective:

$$min_{w,b,\alpha} \quad \lambda||w - \sum_{i}^{M} \alpha_i u_i||^2 + \gamma \sum_{i}^{M} \alpha_i^2 + \sum_{i}^{N} \max\left(0, 1 - y_i(w^{\mathsf{T}} x_i + b)\right) \quad (1)$$

where λ and γ controls the balance between the two regularization terms as well as the tradeoff between error term and regularization terms. u_i's are the classifier patches cropped from source classifiers and relocated on a w sized template padded with zeros other than the classifier patch (see Figure 1), and α_i's are transfer weights. Note that given a fixed set of u_i's the formulation is convex.

EE-SVM is evaluated both quantitatively and qualitatively on the PASCAL VOC 2007 and ImageNet datasets for pose specific object retrieval. It achieves a significant performance improvement over E-SVMs, with greater suppression of negative detections and increased recall, whilst maintaining the same ease of training and testing.

[1] T. Malisiewicz, A. Gupta, and A. A. Efros. Ensemble of exemplar-SVMs for object detection and beyond. In *Proc. ICCV*, 2011.

[2] A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros. Data-driven visual similarity for cross-domain image matching. *ACM Trans. Graph.*, 30(6), 2011.