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

Abstract

Exemplar SVMs (E-SVMs, Malisiewicz et al, ICCV 2011), where a SVM is trained with only a single positive sample, have found applications in the areas of object detection and Content-Based Image Retrieval (CBIR), amongst others.

In this paper we introduce a method of part based transfer regularization that boosts the performance of E-SVMs, with a negligible additional cost. This Enhanced E-SVM (EE-SVM) improves the generalization ability of E-SVMs by softly forcing it to be constructed from existing classifier parts cropped from previously learned classifiers. In CBIR applications, where the aim is to retrieve instances of the same object class in a similar pose, the EE-SVM is able to tolerate increased levels of intra-class variation and deformation over E-SVM, and thereby increases recall.

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.

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 Introduction

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 as in [20], 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).

Recently exemplar SVMs (E-SVM) [15], where an SVM is trained with only a single positive sample, have found applications in the areas of CBIR [21] and object detection [15]. 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),

http://dx.doi.org/10.5244/C.26.79

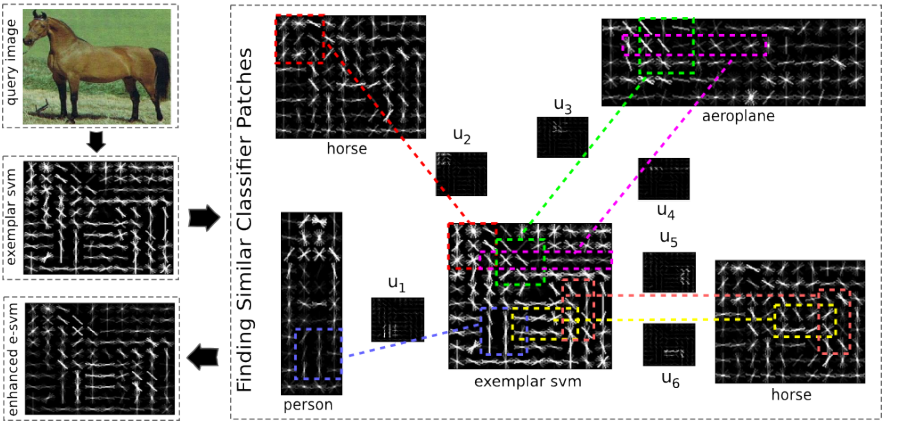

Figure 1: **Overview of the EE-SVM learning procedure.** The box on the right shows mining classifier patches from existing classifiers by matching subparts of E-SVM trained from the given query image. Comparing E-SVM and EE-SVM, better suppression of the background can be seen from the visualized classifiers.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

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 this can lead to false detections), 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).

We describe the enhanced E-SVM in section 2 and give a quantitative and qualitative evaluation in section 4. Although it might be feared that judging the quality of retrieval results will be very subjective, we show that available annotation and measures from PASCAL VOC [7] can be used for this task. In addition to introducing the EE-SVM we show that transfer learning can also be equivalently formulated as feature augmentation. This equivalence has not been explicitly noted before and is another of the contributions of this paper.

1.1 Relation to prior work

Transfer learning [13, 28], has been applied to computer vision primarily for image classification [17, 23, 24, 28], rather than detection, and we discuss the relation of the EE-SVM

formulation to the standard objectives of transfer learning in section 2. Certainly, a possible solution for improving the E-SVM generalization would be transfer learn on the complete classifier (i.e. use the entire template). However this requires a visually similar classifier trained with the same object class, pose, scale and aspect ratio to transfer from. With the help of part level transfer, these constraints become less problematic due to the facts that (i) parts can be relocated, (ii) the possibility of finding a good match for transfer increases when we look at smaller classifier patches. Deformable transfer [1] of the complete classifier level would be another alternative solution, however we [1] observed little significant boost over simpler rigid transfer. Our EE-SVM approach can also be viewed as a deformable transfer considering that parts are being relocated.

Another line of work, that facilitates shared parts across different classes, builds upon the observation that the classes share some common visually coherent substructures, such as wheels, feet, heads, etc. Torralba *et al.* [25] introduced a method for sharing small patch oriented templates in a boosting framework, and Opelt *et al.* [16] extended this approach to shared boundary fragments. Fidler *et al.* [10] explored the shareability of features among object classes in a generative hierarchical framework. Stark *et al.* [22] proposed a method for transferring part-like shape features through explicit migration of model parameters for each part, however this transfer is manual at the moment. Ott and Everingham [18] introduced part sharing across classes for object detection in the framework of discriminatively trained part-based models [9]. In a slightly different way, our work uses the notion of parts as patches of classifier templates. These patches are mined from a set of previously learned classifiers depending on the quality of match with the subparts of a E-SVM template.

The proposed approach is mainly described from the transfer learning perspective, however it has very strong relations to the line of work that focuses on enriching the image descriptors with the responses of high-level classifiers. One popular branch is representing the image by responses of a set of attribute classifiers which are learned in a supervised fashion. This attribute-based representation is successfully employed for object classification tasks [8, 11, 12, 29]. In a similar but an unsupervised fashion, Torresani *et al.* introduced the "classeme" descriptor [19, 26] which is composed of boolean outputs of a set of nonlinear object classifiers that are learned from images returned by text-based image search engines. Building upon the attribute-based representation, Douze *et al.* [6] incorporated Fisher vectors to the representation and proposed an efficient coding technique for compressing the descriptor. All these approaches either replace or augment the original low-level descriptor with the outputs of higher level classifiers. The proposed method also employs a similar augmentation scheme, however we augment the feature vector with the responses of previously learned classifier patches which are selected and relocated based on the quality of match with a E-SVM template learned from the query image.

Combining these two views of the proposed method constructs an equivalence between transfer regularization and feature augmentation. We explicitly prove this equivalence and discuss its implications in section 2.1.

2 Enhanced Exemplar SVM

This section discusses the E-SVM formulation and introduces the enhanced E-SVM objective. The formulation of the E-SVM [15] is:

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

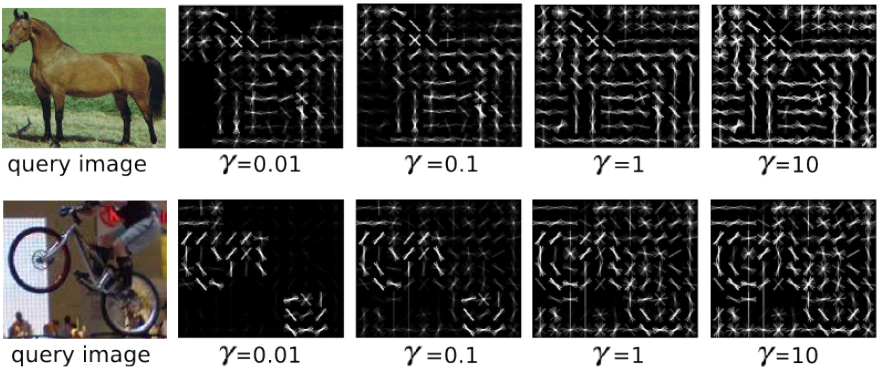

Figure 2: **Two limits of EE-SVM from reconstruction($\gamma = 0.01$) to E-SVM($\gamma = 10$).** Learned EE-SVM templates with varying γ values are displayed. λ is fixed to 1.

where λ controls the weight of regularization term, w is the classifier vector, b is the bias term, x_i and y_i are the training samples and their labels, respectively. Note that there is only one positive sample in the training set and its error is weighted more (50 times in [15]) than the negative samples. In order to simplify the formulation, different weightings of positive and negative samples are not explicitly shown.

In enhanced E-SVM, part based transfer regularization is incorporated to the E-SVM formulation. The objective is:

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

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.

The two limits of this formulation are E-SVM and reconstruction from the classifier patches. As $\gamma \to \infty$, since α_i's will be forced to be zero due to infinite penalization, $\sum_i^M \alpha_i u_i$ will be a zero vector and (2) converges to the E-SVM formulation (1). As $\lambda \to \infty$, w will be forced to be equal to $\sum_i^M \alpha_i u_i$ and thus it will be forced to be constructed as a weighted combination of u_i's. Therefore by tweaking λ and γ we can obtain a midway solution between E-SVM and reconstruction from the other classifiers. Figure 2 shows the smooth transition from reconstruction to E-SVM by changing γ with a fixed λ.

Discussion. Transfer regularization is introduced with Adaptive SVM (A-SVM) [13, 28] which transfers information from a single auxiliary classifier. Subsequently A-SVMs are extended to transfer from multiple classes [27] and similar formulations are employed for a variety of classification [6, 23, 24] and detection tasks [1]. The proposed formulation is also a transfer regularization objective which transfers from the parts of previously learned classifiers. The main difference is that we control the weight of transfer with an additional regularization term $(\gamma \sum_i^M \alpha_i^2)$ where $\gamma \to \infty$ indicates no transfer and $\gamma \to 0$ indicates maximum transfer. The advantages of this representation will be elaborated in the next section. Note that this formulation is not specific to E-SVM and this transfer regularization can also be applied to "classical" SVM formulations.

2.1 Feature Augmentation vs. Transfer Regularization

In the previous section the EE-SVM (2) is mainly described as a transfer learning approach. In this section it will be transformed from the transfer learning perspective to the feature augmentation perspective. The derivation below steps through the rearrangements for mapping (2) to an equivalent "classical" SVM formulation where the feature vector is augmented with the responses of u_i's.

$$\lambda||w-\sum_i^M \alpha_i u_i||^2+\gamma\sum_i^M \alpha_i^2+\sum_i^N \max\left(0,1-y_i(w^\mathsf{T}x_i+b)\right) \quad (w=\Delta w+\sum_i^M \alpha_i u_i) \tag{3}$$

$$= \lambda||\Delta w||^2+\gamma\sum_i^M \alpha_i^2+\sum_i^N \max\left(0,1-y_i\left(\Delta w^\mathsf{T}x_i+\left(\sum_i^M \alpha_i u_i\right)^\mathsf{T}x_i+b\right)\right) \tag{4}$$

$$= ||\bar{w}||^2+\sum_i^N \max\left(0,1-y_i(\bar{w}^\mathsf{T}a_i+b)\right) \qquad \textit{where} \tag{5}$$

$$\bar{w}=[\sqrt{\lambda}\Delta w;\sqrt{\gamma}\alpha_1;\sqrt{\gamma}\alpha_2;...;\sqrt{\gamma}\alpha_M] \quad a_i=[\tfrac{1}{\sqrt{\lambda}}x_i;\tfrac{1}{\sqrt{\gamma}}u_1^\mathsf{T}x_i;\tfrac{1}{\sqrt{\gamma}}u_2^\mathsf{T}x_i;...;\tfrac{1}{\sqrt{\gamma}}u_M^\mathsf{T}x_i] \tag{6}$$

$\bar{w}$ is the transformed classifier and a_i is the augmented feature vector with the responses of u's on x_i. The classifier w, the solution to the original problem (2), can easily be computed from $\bar{w}$ since $w=\Delta w+\sum_i^M \alpha_i u_i$. As is clear from (5), the transformed problem is a "classical" SVM formulation with feature augmentation, and it can be optimized efficiently using existing powerful SVM solvers. Note that this derivation is not limited to the E-SVMs and it can be applied to any transfer regularization objective.

The major implication of this derivation is that transfer regularization can also be stated as a classical SVM minimization problem where the feature vector is augmented with the responses of source classifiers. This equivalence constructs a bridge between papers implementing feature augmentation or populating the features with the responses of high-level classifiers [5, 8, 11, 12, 14, 26, 29] and papers performing transfer regularization [23, 24, 27, 28]. Another direct implication is that transfer regularization approaches [23, 24, 27, 28], which requires specialized optimization, can be reformulated to be efficiently optimized with state-of-the-art SVM solvers.

3 Implementation

In this section the details of the implementation will be discussed. Initially training source classifiers and E-SVM will be described. Afterwards, the EE-SVM training procedure will be explained in two phases: (a) mining regularization parts from source classifiers and (b) optimizing the EE-SVM objective.

The classifiers are linear SVM classifiers (templates) over HOG [3, 9] features. Each HOG cell is composed of a 32 dimensional vector which stores the weight of oriented gradients and the total gradient energy normalized with four neighboring block energies [9]. The source classifiers are trained using PASCAL VOC 2007 [7] training and validation sets using two components for each class without parts, similar to the procedure in [9] . In total we have 20×2 templates. The mirror and upside down (vertical mirror) versions of these templates are also used which adds up to $20\times2\times4=160$ source classifiers. Each E-SVM is composed of 100 or slightly less HOG cells where the aspect ratio is chosen according to the query image. The E-SVM is trained with the given query image as the positive sample and randomly selected 2000 negative images from the PASCAL'07 training set. The training is performed iteratively in a similar fashion to [15] where mined hard negatives are incorporated to the learning after each iteration.

The training procedure of EE-SVM, which is briefly visualized in figure 1, starts with training an E-SVM classifier from the given query image. After obtaining the E-SVM, for each 3×3 cell classifier patch a *good match* is searched for within the source classifiers. This search can be efficiently done using fast nearest neighbor methods. However we performed it as a sliding window search since we have a limited number of source classifiers. Even though we use 3×3 cell classifier patches for experimental validation, any other varying size and aspect ratio can also be applied. A *good match* is defined by thresholding the cosine similarity (normalized dot product) between E-SVM patch (a $3 \times 3 \times 32$ dimensional vector) and classifier patches. This threshold value is fixed to 0.35, but the level can be increased when a larger set of source classifiers exist which would increase the possibility to find much better matchings. After determining where to transfer from, each patch is relocated on a w sized HOG template padded with zeros other than the transferred classifier patch. Finally learning of EE-SVM is performed using the same set of training samples used for training E-SVM and no new hard negatives are collected. The optimization of the EE-SVM objective is performed through the feature augmentation version of the formulation (5) using the LIBSVM [2] package. The only additional cost of EE-SVM over E-SVM is the transformation of training samples, and training another SVM, which constitutes less than 1% of the training time (i.e. mining hard negatives is costly). The test time complexity of EE-SVM is exactly the same as that of E-SVM.

4 Experiments

In this section the experimental results will be described. Initially we'll give the experimental settings, evaluation metrics and the defaults for the hyperparameters. In the next two sections, we'll discuss two set of experiments performed on PASCAL VOC 2007 [7] dataset and ImageNet [4]. Average precision (AP), and precision at top K (PR@5, PR@10, PR@50, PR@100) retrievals are used for evaluating the quality of retrieval results. A correct retrieval is defined as the same object class with the same pose as the query image and the retrieved subwindow should have at least 50% overlap with the true bounding box around the object class. The definition of the pose is inherited from the PASCAL VOC metrics [7] where four main poses exist namely *left, right, frontal, rear* (the pose "unspecified" is omitted) and each pose accepts ± 20 degrees separation from its canonical view. In all the experiments the proposed approach is compared with E-SVM method. λ parameter is fixed to 1 and γ fixed to 5 in all the experiments unless otherwise stated. The matching similarity threshold, which determines the *good* classifier patch matches based on the normalized dot product of two vectors, is fixed to 0.35.

4.1 PASCAL VOC Experiments

The retrievals of PASCAL'07 classes with four main poses are evaluated. The query images are selected as all non-truncated images of the 17 classes (bottle, dining table and potted plant are omitted since they don't have poses) with 4 main poses from the training set. For each query image, an E-SVM and an EE-SVM are trained and run on the test set. Ground truth is identified as the same object class with same pose label. The detections of the same object class other than the target pose is omitted and not counted towards AP computation. For instance if we are searching for a bicycle facing left, we ignore (i.e neither counts as positive nor negative) the detections of front, rear, left or unspecified poses of bicycle. In total 1598 queries from 17 classes are evaluated, and the pose distribution is: 453 left, 440 right, 490 frontal, and 215 rear.

For some query images, due to being unusual examples of the pose (e.g. left facing bicycle with front wheel up as in figure 2), the AP results can be very low. Conversely for

Figure 3: **Retrieval results of PASCAL'07 queries.** Top 3 positives and negatives are being displayed. Orders in the ranked list is shown left bottom corner of each image.

some others, which are canonical examples of the pose, the AP results are much higher. In order to have a better insight on the results and see the boost in different quality of samples, we grouped the queries as being above some AP threshold. The query belongs to the quality group $AP > threshold$, if either AP of E-SVM or EE-SVM is above the defined *threshold* (for instance group $AP \geq 0$ means all the queries). In all the tables improvement in AP is shown as the relative improvement.

Table 1 shows the overall results and the AP improvement of EE-SVM over E-SVM. In all the quality groups EE-SVM significantly improves over E-SVM. Moreover, as the quality of samples increase the boost of EE-SVM increases. In table 2 the AP results and improvements are shown for individual classes for the quality level ($AP \geq 0.05$). For statistical significance only the classes which have more then 10 queries are shown. Except for the *tvmonitor* class, for all the other classes EE-SVM significantly outperforms E-SVM. The reason for the decrease in MAP for *tvmonitor* class is due to the frontal poses where EE-SVM focuses more on what is being displayed rather than the frame of the monitor.

A few qualitative results can be seen in figure 3 where the top three positives and neg-

AP $\geq$	0.00	0.01	0.05	0.10	0.15	0.20	0.30	0.50
# of queries	1598	746	516	314	209	145	68	13
Relative Imp. in MAP	0.115	0.116	0.119	0.135	0.148	0.160	0.171	0.195

Table 1: **Relative MAP (Mean Average Precision) improvements of EE-SVM over E-SVM with changing quality groups.** For instance $AP \geq 0.01$ means the queries which achieved $AP = 0.01$ or above. Queries are the images of 17 classes with four major poses (i.e. left, right, frontal, rear) from PASCAL'07 training set. As the quality of samples increase the boost of EE-SVM increases.

class	plane	bicycle	bus	car	chair	cow	horse	m.bike	person	sheep	tvm.
# of queries	14	66	18	158	13	18	49	32	57	16	41
E-SVM	0.086	0.237	0.109	0.184	0.070	0.081	0.125	0.142	0.082	0.113	0.135
EE-SVM	0.112	0.308	0.121	0.196	0.083	0.088	0.141	0.171	0.085	0.125	0.125
MAP imp.	0.303	0.298	0.116	0.066	0.185	0.083	0.126	0.201	0.045	0.104	-0.077

Table 2: **MAP results and relative improvements of EE-SVM over E-SVM for individual classes for the quality group ($AP \geq 0.05$).** Only classes which have more than 10 queries are shown.

atives are shown with their ranks in the ordered list of retrieved subwindows. In EE-SVM retrievals the ranks of the top three negatives are much later, this shows that EE-SVM better suppresses the negatives and thus increases the recall.

The effect of parameter selections. There are few parameters of the system that can be adjusted for different purposes. For instance, in order to handle the occlusions we need a more aggressive transfer (i.e. small γ), and larger patches to transfer from (i.e. 4×4 or 5×5). If the query sample is not a common pose of a common class we can prefer small patches (i.e. 3×3) to increase the chance of a possible match and perhaps decrease the patch similarity threshold.

Handling occlusion and truncation via EE-SVM. It is quite common to came across truncated and occluded query images. With the help of partial classifier patch matchings we can complete the truncated parts or remove the effect of partial occlusion. A partial match is defined as, given a partial match ratio β as a threshold, two classifier patches are matching if any $\beta\%$ subselection of the two patches match with the confidence above the similarity threshold. Figure 5 shows two examples of handling occlusion and completing the truncated parts. In these examples we used 5×5 patches and the partial match ratio β is 70%. When we increase the size of the patches even though the chance of finding a good match decreases, if there is one it improves the result drastically (see precision recall curves in figure 5). γ, which defines the strength of transfer, is set to 1 for these experiments.

4.2 ImageNet Experiments

These experiments are conducted on ImageNet (though transfer is still from detectors learnt on PASCAL VOC). For quantitative experiments five ImageNet classes (synsets) are selected: *lion, deer, tandem, bulldozer,* and *ambulance*. For each of these classes three queries (from one of the main canonical poses) are selected from web images and evaluated on the

	lion		deer		tandem		bulldozer		ambulance		MEAN	
	e-svm	ee-svm	e-svm	ee-svm	e-svm	ee-svm	e-svm	ee-svm	e-svm	ee-svm	e-svm	ee-svm
PR@5	0.47	**0.60**	0.60	**0.73**	1.00	1.00	0.93	0.93	1.00	1.00	0.80	**0.85**
PR@10	0.40	0.40	0.50	0.50	1.00	1.00	0.90	0.90	1.00	1.00	0.76	0.76
PR@50	0.19	**0.21**	0.30	**0.32**	0.98	**0.99**	0.62	**0.67**	0.85	**0.87**	0.59	**0.61**
PR@100	0.13	**0.14**	0.22	**0.28**	0.93	**0.97**	0.42	**0.49**	0.76	**0.80**	0.49	**0.54**

Table 3: **Precision at top K comparison of ImageNet Queries.** Three queries with varying poses are evaluated for each class from ImageNet and the mean precisions are presented. Retrieval is performed on the collection composed of PASCAL'07 test images and corresponding ImageNet category.

Figure 4: Retrieval of unusual poses on ImageNet. A visual phrase retrieval is also shown on the rightmost column.

corresponding ImageNet class images. PASCAL'07 test set is also added to the evaluated samples in order introduce noise in the image database. The evaluations are compared using precision at the the top K retrievals. From the results, displayed in table 3, we can conclude that the recall of EE-SVM is much better than the recall of E-SVM, particularly for top 50 and top 100 retrievals.

In addition to canonical poses, the method is also qualitatively demonstrated for unusual poses. With the help of part based transfer, since parts can be relocated and migrated across classes, even for quite unusual poses we can obtain significant improvements. The left facing bicycle with the front wheel up (see figure 2 and figure 4) is a nice example where the wheel patches are transferred from motorbike and bicycle classifiers with regular poses. Another example, displayed in figure 4, is a sitting lion where the ranks of positives clearly show EE-SVM's ability for better recall.

For visual phrases our method successfully reconstructs the classifier template from the patches of existing source classifiers. A visual phrase example (i.e. person riding horse) is also displayed in figure 4.

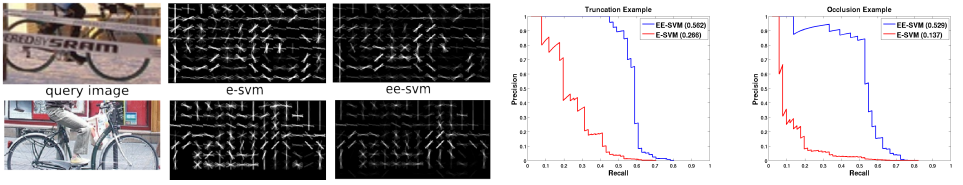

Figure 5: Occlusion and truncation handling via EE-SVM. It is better visualized with zooming into the document.

5 Conclusion

As has been shown, part level transfer regularization can be used not only for enhancing classifiers, but also for going beyond the spatial extent of the query by completing occlusions and truncations via partial part matchings. These matchings can be further improved by exploring the co-occurrence relations between part classifiers.

The equivalence between feature augmentation and transfer regularization introduces a new perspective to re-explore the papers from both subjects, and also a more convenient method for implementing transfer regularization by using standard SVM packages.

Acknowledgements. We are grateful for financial support from ERC grant VisRec no. 228180.

References

[1] Y. Aytar and A. Zisserman. Tabula rasa: Model transfer for object category detection. In *Proc. ICCV*, 2011.

[2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2011.

[3] N. Dalal and B Triggs. Histogram of Oriented Gradients for Human Detection. In *Proc. CVPR*, 2005.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *Proc. CVPR*, 2009.

[5] M. Douze, A. Ramisa, and C. Schmid. Combining attributes and fisher vectors for efficient image retrieval. In *Proc. CVPR*, 2011.

[6] L.X. Duan, D. Xu, I.W. Tsang, and J.B. Luo. Visual event recognition in videos by learning from web data. In *Proc. CVPR*, 2010.

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. *IJCV*, 2010.

[8] A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. Describing objects by their attributes. In *Proc. CVPR*, 2009.

[9] P. F. Felzenszwalb, R. B. Grishick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part based models. *IEEE PAMI*, 2010.

[10] S. Fidler, M. Boben, and A. Leonardis. Evaluating multi-class learning strategies in a generative hierarchical framework for object detection. In *NIPS*, 2009.

[11] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Efficient discriminative learning of parts-based models. In *Proc. ICCV*, 2009.

[12] C. H. Lampert and M. B. Blaschko. Structured prediction by joint kernel support estimation. *Machine Learning*, 2009.

[13] X. Li. *Regularized Adaptation: Theory, Algorithms and Applications*. PhD thesis, University of Washington, USA, 2007.

[14] J. Luo, T. Tommasi, and B. Caputo. Multiclass transfer learning from unconstrained priors. In *Proc. ICCV*, 2011.

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

[16] A. Opelt, A. Pinz, and A. Zisserman. A boundary-fragment-model for object detection. In *Proc. ECCV*, 2006.

[17] F. Orabona, C. Castellini, B. Caputo, A.E. Fiorilla, and G. Sandini. Model adaptation with least-squares svm for adaptive hand prosthetics. In *Proc. Intl. Conf. on Robotics and Automation*, 2009.

[18] P. Ott and M. Everingham. Shared parts for deformable part-based models. In *Proc. CVPR*, 2011.

[19] M. Rastegari, C. Fang, and L. Torresani. Scalable object-class retrieval with approximate and top-k ranking. In *Proc. ICCV*, 2011.

[20] M.A. Sadeghi and A. Farhadi. Recognition using visual phrases. In *Proc. CVPR*, 2011.

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

[22] M. Stark, M. Goesele, and B. Schiele. A shape-based object class model for knowledge transfer. In *Proc. ICCV*, 2009.

[23] T. Tommasi and B. Caputo. The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories. In *Proc. BMVC.*, 2009.

[24] T. Tommasi, F. Orabona, and B. Caputo. Safety in numbers: Learning categories from few examples with multi model knowledge transfer. In *Proc. CVPR*, 2010.

[25] A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Proc. CVPR*, 2004.

[26] L. Torresani, M. Szummer, and A. Fitzgibbon. Efficient object category recognition using classemes. In *Proc. ECCV*, 2010.

[27] J. Yang, R. Yan, and A.G. Hauptmann. Cross-domain video concept detection using adaptive svms. In *ACM Multimedia*, 2007.

[28] J. Yang, R. Yan, and A.G. Hauptmann. Adapting svm classifiers to data with shifted distributions. In *ICDM Workshops*, 2007.

[29] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. J. Guibas, and F. F. Li. Human action recognition by learning bases of action attributes and parts. In *Proc. ICCV*, 2011.