

Image Detection Using Exemplar-SVM with Augmented Features

Shuwan Sun^a, Mingyan Jiang^{b,*}

School of Information Science and Engineering, Shandong University, Jinan, 250100, China

^a sunshuwan112@163.com, ^{b,*} jiangmingyan@sdu.edu.cn

*corresponding author

Keywords: Exemplar-SVM, Feature Augmentation, calibration.

Abstract. Exemplar-SVM (E-SVM) was first introduced by Malisiewicz et al, and achieved good performance in the application of object detection and Content-Based Image Retrieval. E-SVM is about training a linear SVM with a single positive sample and many negative samples. All the E-SVMs thus form an ensemble that performs well. Training a good E-SVM requires the negative set to be very large. In this paper we propose to apply E-SVM with augmented features to the area of image detection by softly forcing it to be constructed from existing classifier parts cropped from previously trained classifiers. Our method helps to improve the generalization of the original E-SVM. The proposed method is evaluated on the subset of PASCAL VOC 2007 to detect different orientation of a specific object, and it achieves better performance over E-SVM whilst helps reducing the size of negative set.

1. Introduction

An E-SVM[1] is a large collection of linear SVM classifiers, each trained from one positive example and many negative ones (an E-SVM). At test time each window is scored by all E-SVMs, and the highest score is assigned to the window. We particularly interested in retrieving subwindows of images which are similar to the given query image[6].

There is some good properties of E-SVMs: (1) Little overfitting. Although each SVM has only a single positive example, the huge number of negatives appear to provide enough to constrain the problem. In a way, the exemplar's decision boundary is defined, in large part, by what it is not. At the same time, each classifier is solving a much simpler problem than in the full category case. (2) While the large imbalance between the positive and negative sets can often lead to a poor decision boundary, we have empirically found that the induced ordering of the detections with respect to that boundary is still good. Thus, the E-SVM can be interpreted as ordering the negatives by visual similarity to the exemplar. (3) While a standard category classifier treats positive and negative examples in the same way, the Ensemble of E-SVMs handles them differently. One way to think about it is that the positives are represented non-parametrically, while the negatives are represented parametrically[10].

In this paper we studied to incorporated part based representation with a feature mapping function into the E-SVM object formulation. This approach can also be interpreted as transfer learning[2], which requires a classifier trained by visually similar images of the same object with different scale, pose, orientation et al. Thus features of exemplar object are augmented and explicitly represented.

2. Algorithm Formulation

2.1. Exemplar-SVM Approach Formulation

Given a set of training exemplars x_e we create a descriptor from the ground-truth bounding box of each exemplar with a cell size of 8 pixels using a sizing heuristic which attempts to represent each exemplar with roughly 100 cells. Instead of warping each exemplar to a canonical frame. We create negative samples of the same dimensions as x_e by extracting negative windows from images not containing any objects from the exemplar’s category.

Weight vector w_E and bias term b of each E-SVM (w_E, b) can be learned by optimizing the following convex object using the hinge loss function:

$$\min_{w_E, b} \lambda \|w\|^2 + \sum_i \max(0, 1 - y_i(w^T x_i + b)) \quad (1)$$

Where λ controls the weight of regularization term. Since that there is only one positive sample in the training set, its error is weighted more than the negative ones. The difference is not explicitly shown to simplify the formulation.

Because of this max operation, it is necessary to calibrate the E-SVMs to make their scores comparable. A common procedure is to calibrate each SVM independently, by fitting a logistic sigmoid to its output on a validation set. Our calibration [3] step can be interpreted as a simple rescaling and shifting of the decision boundary. Given a detection x and the learned sigmoid parameters (α_E, β_E) , the calibrated detection score for exemplar E is as follows:

$$f(x | w_E, \alpha_E, \beta_E) = \frac{1}{1 + e^{-\alpha_E(w_E^T x - \beta_E)}} \quad (2)$$

While the logistic fitting is performed independently for each exemplar, we found that it gives us a considerable boost in detection performance over using raw SVM output scores. In case of E-SVM, every training data is trained against a large number of negative examples. Thus, for 1000 bounding boxes, we train 1000 SVMs.

2.2. Exemplar-SVM with Augmented Feature

Since too much generalization can be extracted from a positive exemplar image, we try to address this question by relocating parts of the exemplar, and finding a good match for transfer increases when we look at smaller classifier patches.

In Augmented E-SVM [5], part based transfer regularization is incorporated to the E-SVM formulation. The objective is:

$$\min_{w_E, b, \alpha} \lambda \left\| w - \sum_i \alpha_i u_i \right\|^2 + \gamma \sum_i \alpha_i^2 + \sum_i \max(0, 1 - y_i(w^T x_i + b)) \quad (3)$$

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 are the classifier patches cropped from source classifiers and relocated on a w sized template padded with zeros other than the classifier patch ,

and α_i are transfer weights. Note that given a fixed set of u_i the formulation is convex.

The two limits of this formulation are E-SVM and reconstruction from the classifier patches. As $\gamma \rightarrow \infty$, since α_i will be forced to be zero due to infinite penalization, $\sum_i^M \alpha_i u_i$ will be a zero vector and Eq(3) converges to the E-SVM formulation Eq(1). As $\lambda \rightarrow \infty$, 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 . Therefore by tuning λ and γ we can obtain a midway solution between E-SVM and reconstruction from the other classifiers.

3. Experiment

In this section we will describe the experimental results. The PASCAL VOC 2007 is a popular dataset used to benchmark object detection algorithms, which contains 20 classes annotated by bounding-box using *Libsvm*[11]. We chose four categories (bus, car, bike and horse) to do our experiment on, and learning classifier parameters for each of the exemplar. Parameter λ and γ is fixed to 1 and 5 respectively. The performance of proposed method is compared with the original E-SVM.

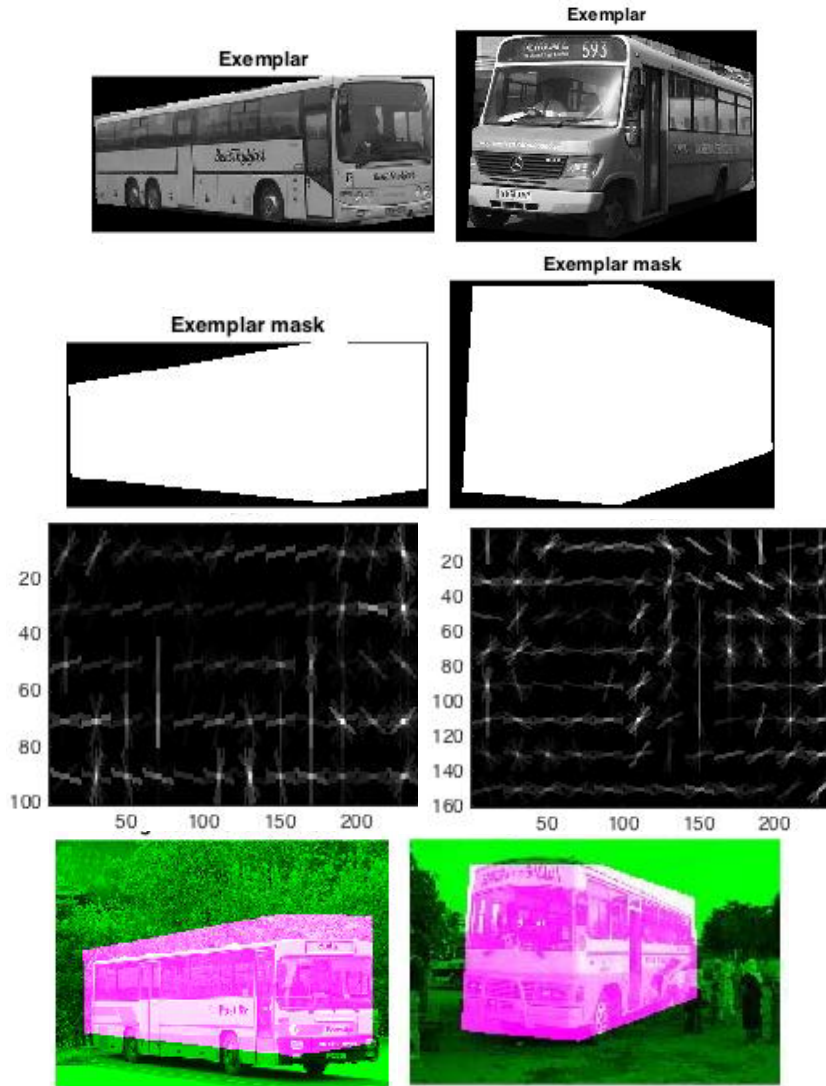




Figure 1 Detection Results. Row 1: positive sample with different orientation. Row 2: Exemplar mask of the given positive sample. Row 3: Hog weight vector. Row 4-6:Top 3 detected pictures from the testset.

Table 1 Map improvements of AE-SVM over E-SVM

Approach	bus	Car	bike	horse
Exemplar-SVM	0.109	0.185	0.143	0.126
AE-SVM	0.122	0.195	0.172	0.140
MAP Improvement	0.117	0.065	0.202	0.125

As is shown in figure 1, top 3 positive examples is retrieved from the VOC 2007dataset. In table 1 the MAP results and improvements are shown for individual classes. We can see AE-SVM significantly outperforms E-SVM.

4. Conclusions

We present a method to improve the performance of E-SVM. E-SVM is based on training a separate classifier for each exemplar, and can achieve fairly good generalization. As has been shown, Our method can further improve the generalization of E-SVM and help reduce the size of the negative training set.

Acknowledgements

This work is supported by the National Natural Science Foundation of Shandong Province of China under Grant No.ZR2014FM039.

References

- [1] Malisiewicz T, Gupta A, Efros A A. (2011) Ensemble of exemplar-svms for object detection and beyond. 2011 IEEE International Conference on Computer Vision (ICCV), 89-96.
- [2] Ayta Y, Zisserman A. (2012) Enhancing Exemplar SVMs using Part Level Transfer Regularization, BMVC, 1-11.

- [3] Modolo D, Vezhnevets A, Russakovsky O, et al. (2015) Joint calibration of ensemble of exemplar svms. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3955-3963.
- [4] Zepeda J, Perez P. (2015) Exemplar SVMs as visual feature encoders. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3052-3060.
- [5] Li W, Duan L, Xu D, et al. (2014) Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. IEEE transactions on pattern analysis and machine intelligence, 36(6): 1134-1148.
- [6] Mitro J. (2016) Content-based image retrieval tutorial. arXiv preprint arXiv:1608.03811.
- [7] Cauwenberghs G, Poggio T. (2000) Incremental and decremental support vector machine learning. NIPS, 13.
- [8] Malisiewicz T, Shrivastava A, Gupta A, et al. (2012) Exemplar-SVMs for Visual Object Detection, Label Transfer and Image Retrieval, ICML.
- [9] Felzenszwalb P F, Girshick R B, McAllester D, et al. (2010) Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9): 1627-1645.
- [10] Malisiewicz T. (2011) Exemplar-based representations for object detection, association and beyond. Carnegie Mellon University.
- [11] Chang C C, Lin C J. (2011) LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3): 27.