Regularizing Max-Margin Exemplars by Reconstruction and Generative Models

Jose C. Rubio, Björn Ommer
Heidelberg Collaboratory for Image Processing & IWR, Heidelberg University, Germany

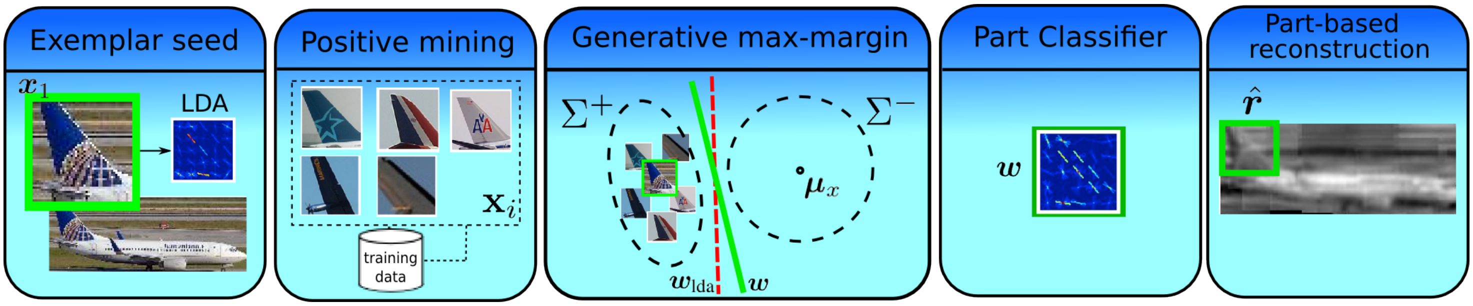

Figure 1. Scheme of the approach. First local patches are extracted from positive boxes. With each, an LDA classifier is trained. Then, positive mining is performed using that classifier on positive bounding boxes. Following, our generative max-margin classifier is trained using those positives as well as the distributions Σ^+ and Σ^-. Note that the red dashed line indicates the classifier obtained with regular LDA, while the green solid line represents our classifier, that *aligns* acoording to the shape of positive and negative covariances. Finally, the improved part filters are obtained, and part-based reconstruction of the original instance can be performed.

Part-based models are one of the leading paradigms in visual recognition. In the absence of costly part annotations, associating and aligning different training instances of a part classifier and finding characteristic negatives is challenging and computationally demanding. Our main goal is to solve those two problems consistently by estimating generative models for both positives and negatives and integrating them into a max-margin exemplar-based linear classifier. The outline of our approach is detailed in Figure 1. Our contributions can be summarized as follows:

1. Contrarily to LDA [3], we explicitly model both negative and positive distributions of the training data: the negative distribution allows for fast training by avoiding mining hard-negatives. The positive distribution serves to reduce the classifier degrees of freedom imposing a category-specific structured prior that assures the quality of the filters. Quadratic Discriminant Analysis also models both distributions explicitly, but becomes intractable for large-scale problems due to its quadratic decission boundary.
2. We propose a novel perceptual regularizer that prevents the classifier from changing the concept represented by its original exemplar, when training with multiple positive instances.
3. Such regularization term links our model to the image domain, thus enabling the generation of visual representations of our part-based models.
4. Our optimization strategy trains models efficiently (on the order of LDA) while achieving performance superior to exemplar-SVM with hard negative mining.

Our goal is to determine a hyper-plane $\boldsymbol{w}$ with maximal Mahalanobis distance to the distribution of negatives $(\boldsymbol{\mu}_x \Sigma^-)$ and to the distribution of positives $(\boldsymbol{x}_i, \Sigma^+)$. This family of classifiers is closely related to Support Vector Machines [4], and aims to establish a decision hyper-plane based on *local* information (positive exemplar features) as well as *global* evidence (positive and negative probability densities). We include a L1 normalization to encourage structured feature activations and drive the solution towards good quality optima. The objective function is formally defined as:

$$\begin{aligned}
\underset{\gamma,b,\boldsymbol{w},\xi}{\arg\max} \quad & \gamma - \left[\|\boldsymbol{w}\|_1 + C\sum_{i=1} \xi_i + \beta \|\hat{\boldsymbol{r}} - \boldsymbol{r}_1\|_2^2 \right] \qquad (1)\\
\text{s.t} \quad & \boldsymbol{w}^\top \boldsymbol{x}_i - b \geq \gamma \sqrt{\boldsymbol{w}\Sigma^+\boldsymbol{w}} - \xi_i \quad \forall i\\
& -\boldsymbol{w}^\top \boldsymbol{\mu}_x + b \geq \gamma \sqrt{\boldsymbol{w}\Sigma^-\boldsymbol{w}}\\
& \hat{\boldsymbol{r}} = \Sigma^{rx}\boldsymbol{w} + \boldsymbol{\mu}_r\\
& \xi \geq 0, \ \gamma \geq 0
\end{aligned}$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

where $\boldsymbol{r}_1$ is a wavelet representation of the exemplar image patch. To make our models robust to badly chosen positive samples, the regularization term $\|\hat{\boldsymbol{r}} - \boldsymbol{r}_1\|_2^2$ prevents the classifier from drifting away from the original object region that the exemplar represented in the first place. The term $\hat{\boldsymbol{r}} = \Sigma^{rx}\boldsymbol{w} + \boldsymbol{\mu}_r$ inverts the classifier to the wavelet domain applying a ridge-regressor, in a similar fashion as [6]. The right-most box of Figure 1. shows an example of a part-based reconstruction.

The following table shows the evaluation of object detection performance obtained with different algorithms for object detection in PASCAL07. We emphasize the training time for a given amount of part filters to show that our models with margin prediction achieve competitive performance with extremely fast training times.

	Ours/Fast	ESVM[5]	LDA[3]	DPM	LLDA[2]	RM^2C[1]
av.AP	34.6 / 33.6	33.0	31.6	33.7	24.4	32.9
time	14h / 14 min.	40h	1sec.	5h	7 min.	40h
#parts	1000	1000	48	48	1000	1000

To conclude, we have integrated a generative regularization into the discriminative part training and enforcing good visual reconstruction. The proposed part training algorithm is significantly faster than popular ESVM or DPM while retaining object recognition performance and improving part localization and the ability to reconstruct and explain images of objects.

[1] Angela Eigenstetter, Masato Takami, and Bjorn Ommer. Randomized max-margin compositions for visual recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[2] Ross Girshick and Jitendra Malik. Training deformable part models with decorrelated features. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2013.

[3] Bharath Hariharan, Jitendra Malik, and Deva Ramanan. Discriminative decorrelation for clustering and classification. In *European Conference on Computer Vision (ECCV)*, 2012.

[4] Gert R.G. Lanckriet, Laurent El Ghaoui, Chiranjib Bhattacharyya, and Michael I. Jordan. A robust minimax approach to classification. *Journal of Machine Learning Research*, 2003.

[5] Tomasz Malisiewicz, Abhinav Gupta, and Alexei A. Efros. Ensemble of exemplar-svms for object detection and beyond. In *ICCV*, 2011.

[6] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba. Hoggles: Visualizing object detection features. *ICCV*, 2013.