

Spotlight the Negatives: A Generalized Discriminative Latent Model

Hossein Azizpour¹

www.csc.kth.se/~azizpour/

Mostafa Arefiyan²

mostafa@brown.edu

Sobhan Naderi Parizi²

<http://vision.lems.brown.edu/students/Parizi>

Stefan Carlsson¹

stefanc@csc.kth.se

¹ CVAP

Royal Institute of Technology (KTH)
Stockholm, Sweden

² LEMS

Brown University
Providence, USA

Discriminative latent variable models (LVM) are frequently and successfully applied to various visual recognition tasks [1, 2, 3, 4, 5, 6, 7]. In these systems the latent (hidden) variables provide a formalism for modeling structured variation of visual features. These variations can be the result of different possible object locations, deformations, viewpoint, sub-categories, etc. Conventionally, these have all been defined only based on the foreground (positive) class. We call such latent variables “positive”. In this work we introduce generalized discriminative LVM (GLVM) which use both “positive” and “negative” latent variables. Negative latent variables are defined on the background (negative) class and provide *counter evidence* for presence of the foreground class. They can, for instance, learn mutual exclusion constraints, model scene subcategories where the positive object class is unlikely to be found, or capture specific parts which potentially indicate the presence of an object of a similar but not the same class.

The objective of the proposed framework is to learn a model which maximizes the evidence (characterized by positive latent variables) for the foreground class and at the same time minimizes the counter evidence (characterized by negative latent variables) lying in background class. Thereby GLVM empowers the latent variable modeling by highlighting the role of negative data in its own right. An interesting analogy is with game theory; when playing a game a simple strategy is to maximize your score. This resembles LVM’s objective which is to maximize evidence for the presence of the foreground class. A better strategy, however, is to maximize your scores at the same time as minimizing the opponents’ scores. This is analogous to GLVM’s objective which takes into account counter evidence emerging from negative latent variables while looking for evidence from the foreground class. The score of the opponents is counter evidence in our hypothetical example.

The concept of *negative parts* was noted in [3]. However, [3] focuses on automatic discovery and optimization of a part based model with negative parts. In this paper we extend the notion of negative parts to negative latent variables and propose a framework for defining models that use both positive and negative latent variables.

Many different modeling techniques benefit from latent variables. Felzenszwalb *et al.* [1] introduced latent SVM to train deformable part models, a state of the art object detector [2]. Yu and Joachims [7] transferred the idea to structural SVMs. Yang *et al.* [6] proposed a kernelized version of latent SVM and latent structural SVM and applied it to various visual recognition tasks. Razavi *et al.* [4] introduced latent variables to the Hough transform and achieved significant improvements over a Hough transform object detection baseline. And-Or trees use latent variable modeling for object recognition and occlusion reasoning [5].

In this paper we introduce a novel family of models which generalizes discriminative latent variable models. We review LVMs and formulate our generalization of LVMs. Then, we discuss training of GLVMs and propose an algorithm for learning the parameters of a GLVM. Then we discuss how various computer vision models can benefit from GLVM. Finally, we experiment on generalized DPMs.

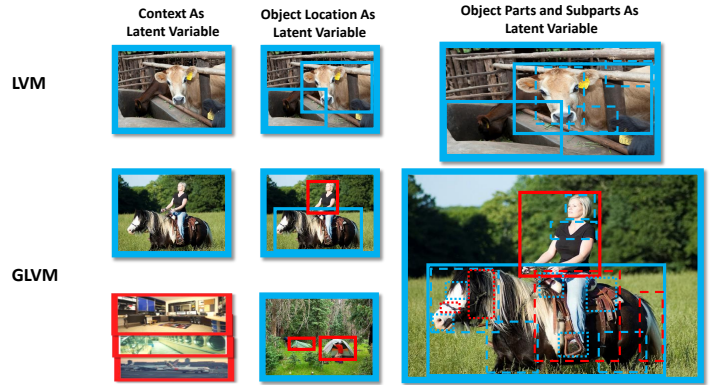


Figure 1: **Intuitive example of LVM and GLVM.** Assume the task of cow image classification. This figure visualizes examples of possible latent variables for LVM and GLVM models. A latent variable can be assigned at image level such as the background scene and then gradually go down a hierarchy of objects, parts and subparts. Positive latent variables (modeling evidences for cow) of an LVM may correspond to different scenes where cow is likely to be seen *e.g.* stable, meadow (blue boxes in the first column). Another positive latent variable can be assigned to the location of the cow in the image (second column, first row). In addition, location of different parts of cow *e.g.* head, back, legs can be modeled as other latent variables dependent to the latent variable of cow location (top right). All these examples of latent variables in LVMs look for evidences for the existence of cow. In GLVM we encourage the modeling of counter evidences through negative latent variables. For example at high level a negative latent variable may correspond to scenes where cow is unlikely to be found *e.g.* office, laundromat, or airport (bottom left). Meadow is an evidence for cow, but cows are less likely to be seen in a meadow beside camping tents (a dependent negative latent variable to positive latent variable meadow). Horses are visually similar to cow but people usually do not ride cows thus a human on top of cow can be negative evidence (a negative latent variable dependent to the positive latent variable of cow location). Furthermore, human, a negative latent variable for cow, can have its own positive latent variable as parts *e.g.* face and shoulders (bottom right).

- [1] Pedro F. Felzenszwalb, David A. McAllester, and Deva Ramanan. A discriminatively trained, multiscale, deformable part model. In *CVPR*, 2008.
- [2] Ross B. Girshick, Forrest N. Iandola, Trevor Darrell, and Jitendra Malik. Deformable part models are convolutional neural networks. In *CVPR*, 2015.
- [3] Sobhan Naderi Parizi, Andrea Vedaldi, Andrew Zisserman, and Pedro F. Felzenszwalb. Automatic discovery and optimization of parts for image classification. In *ICLR*, 2015.
- [4] Nima Razavi, Juergen Gall, Pushmeet Kohli, and Luc J. Van Gool. Latent Hough transform for object detection. In *ECCV*, pages 312–325, 2012.
- [5] Xi Song, Tianfu Wu, Yunde Jia, and Song-Chun Zhu. Discriminatively trained and-or tree models for object detection. In *CVPR*, pages 3278–3285, 2013.
- [6] Weilong Yang, Yang Wang, Arash Vahdat, and Greg Mori. Kernel latent SVM for visual recognition. In *NIPS*, pages 818–826, 2012.
- [7] Chun-Nam John Yu and Thorsten Joachims. Learning structural svms with latent variables. In *ICML*, pages 1169–1176, 2009.