Some objects can be represented in a deformable configuration. The model provides a fast semi-supervised training method without relying on feature detection. Location of object is used as a latent variable.

The image is represented by HOG representation. The main body of object is labeled by a bounding box and it is defined in a low resolution image pyramid. A coarse root filter covers the major body of object by the bounding box around the object. The part filters are defined in low levels of image pyramid with higher solution, in a specific size. Higher resolution helps to capture more details for part filters. The placements of part filters are computed at different levels of image pyramid and relative possible position corresponding to the root location, by given the model. The best location is a function in terms of model parameters and root location. Also a valid placement of part filter should overlap with the root filter.

For latent SVM and classical SVM, hard negative data is used to reduce the training time and dataset. Hard negative data are selected from a large set of possible negative data. The idea is that if the hard negative data can be classified correctly, then true negative data can be also classified correctly. So training results would be the same while the training data would be reduced.

Six part filters are selected based on the root filter and fill into higher resolution image. To update model, existing detector is moved around the bounding box with more than 50% overlap to create more positive examples. The method performs well on rigid objects and persons. Those objects have more or less spatial homogeneity.