Self-learning A Scene-specific Pedestrian Detector using a Progressive Latent Model

University of Chinese Academy of Sciences

Qixiang Ye[1], Tianliang Zhang[1], Wei Ke[1], Qiang Qiu[2], Jie Chen[3], Guillermo Sapiro[2]

[1] University of Chinese Academy of Sciences, Beijing, China [2]Duke University, [3] University of Oulu, Finland

IEEE 2017 Conference on Computer Vision and Pattern Recognition

CVPR July 21-26 HONOLULU 2017

Motivation

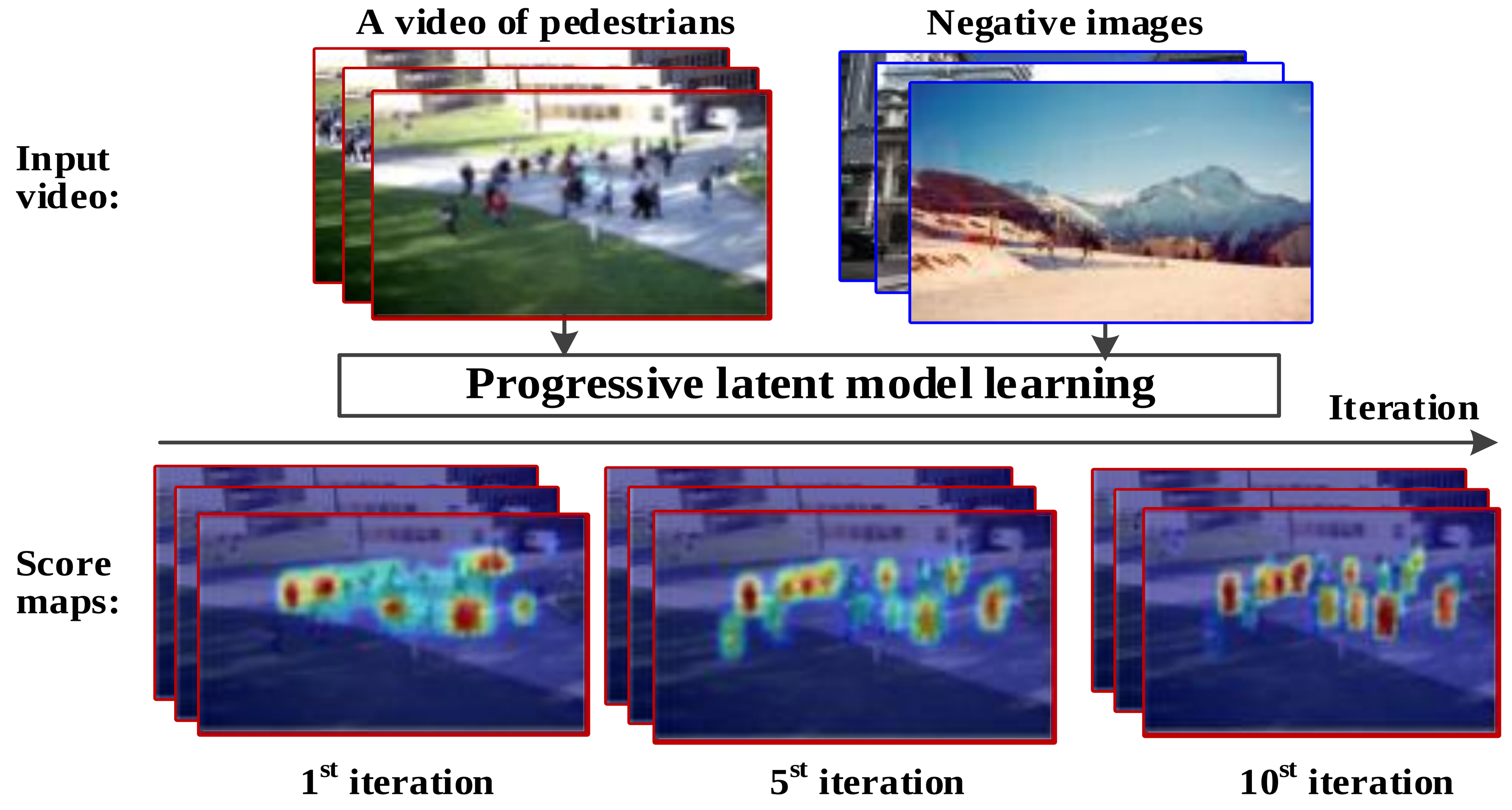
1) Supervised learning of detectors for different scenes requires repeated human effort;

2) Offline-trained detectors degrade with changes of the scene or views of camera;

3) Scene specific cues, e.g., object resolution, occlusions, and background structures are not incorporated into the detectors

Dense pedestrians Sparse pedestrians Flat view Bird view

Self-learning Framework



Given a video where pedestrians are dominant moving objects, self-learning progressively constructs a scene-specific detector using object discovery, object enforcement, and label propagation procedures

Progressive Latent Model (PLM)

$$\{h^*, \beta^*\} = \min_{\beta,h} \mathcal{F}_{(\mathcal{X},\mathcal{Y})}(\beta, h)$$

$$= \min_{\beta,h} \mathcal{F}_l(\beta, h) - \lambda \mathcal{F}_s(\beta) + \gamma \mathcal{F}_g(\beta, h)$$

Object discovery Object enforcement Label propagation

Object discovery: implemented with a latent SVM (LSVM) model to choose object proposals that best discriminate positive frames from negative images

$$\min_{\beta,h} \mathcal{F}_l(\beta, h) = \min_{\beta,h} \frac{1}{2}||\beta||^2 + \mathcal{C}\sum_{i=1}^{N} l(\beta, x_i, y_i, h)$$

$$l(\beta, x_i, y_i, h) = \max_{y,h} \left(\beta^T \cdot v(x_i, y, h) + \Delta(y_i, y)\right)$$

$$- \max_{h} \beta^T \cdot v(x_i, y_i, h)$$

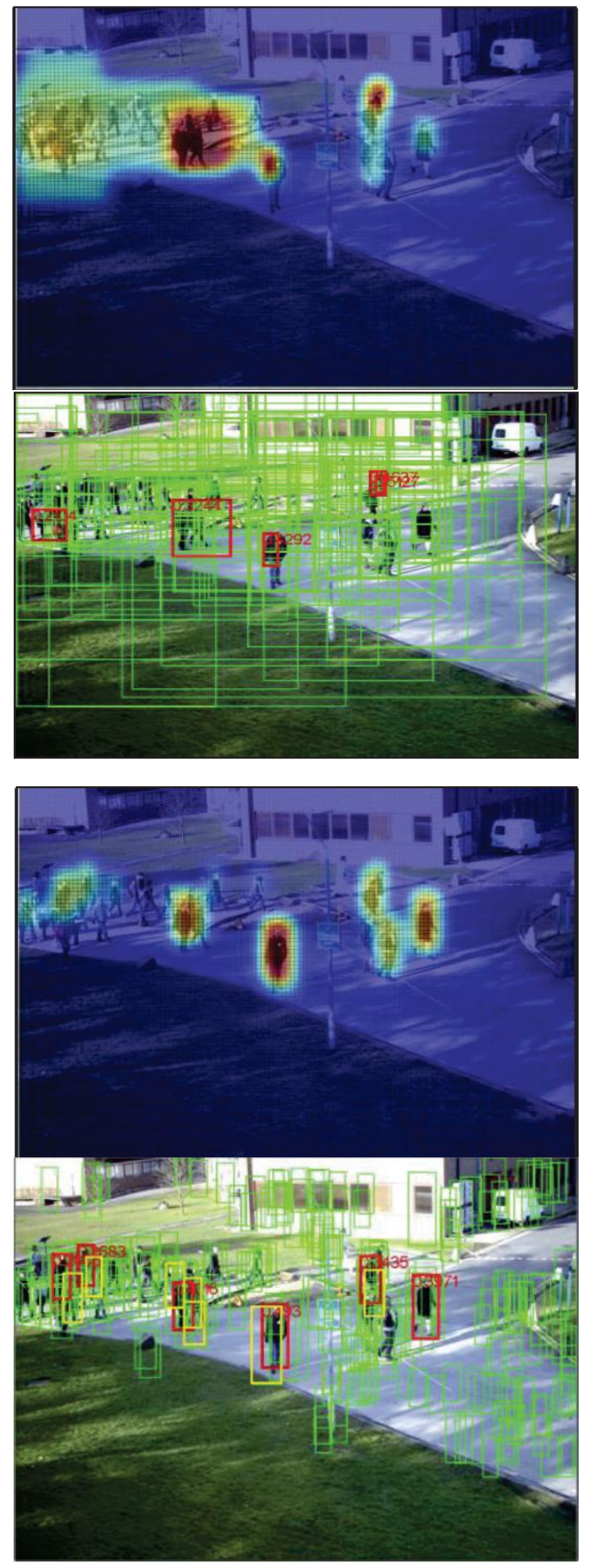
Object enforcement: Motivated by the success of hard negative mining, we propose using spatial regularization to enforce the localization of objects and the model.

$$\max_{\beta} \mathcal{F}_s(\beta) = \sum_{i=1}^{N} \sum_{\substack{h \in \mathcal{H}_i \\ h' \in \Omega_{\mathcal{H}_i, h}}} ||\beta^T \cdot (v(x_i, h) - v(x_i, h'))||^2$$

Label propagation: To mine multiple instances, using the graph-based label propagation for incremental learning

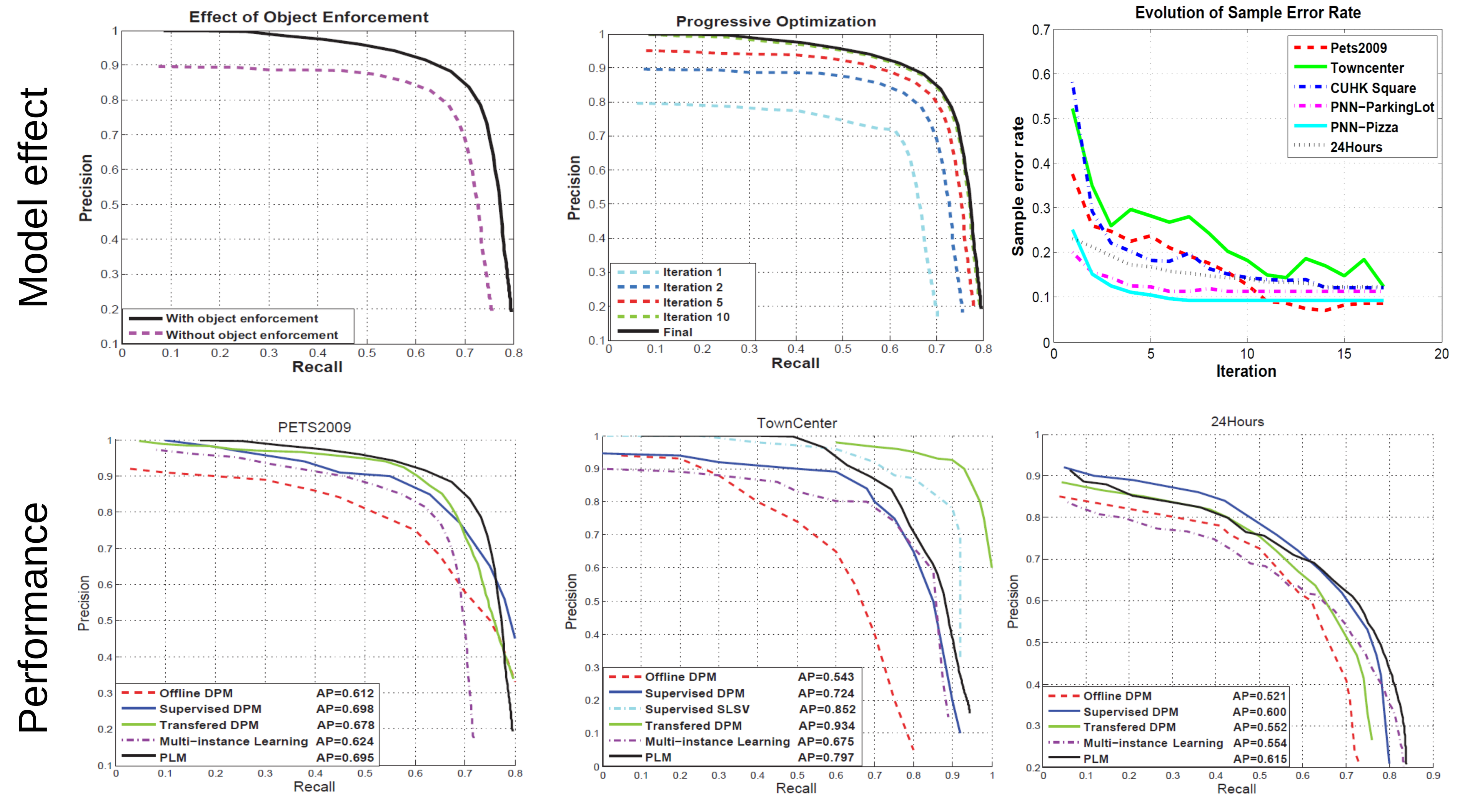
$$\min_{g(\beta,h)} \mathcal{F}_g(\beta, h) = \min_{g(\beta,h)} \sum_{i=1}^{l} \sum_{j=l}^{l+u} w_{ij} \left(g(\beta, h_i) - g(\beta, h_j)\right)^2$$

$$s.t. \quad g(\beta, h_i) = y_i, i = 1, ..., l$$

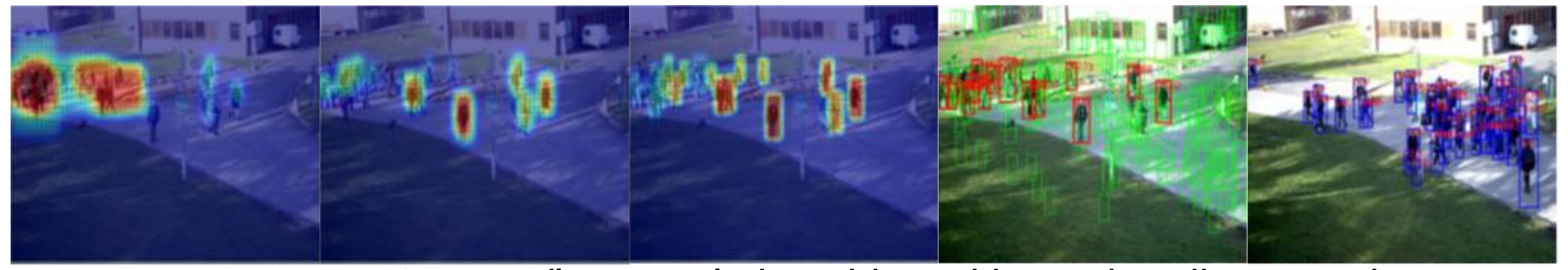


Optimization: With the difference of convex (DC) objective functions, PLM can be efficiently optimized with a concave-convex programming and thus guaranteeing the stability of self-learning.

Experimental Results



Pets2009 (crowd)



CUHK Square (low resolution video with moving distracters)

Conclusion

➢ A progressive latent model is proposed by incorporating discriminative and incremental functions.

➢ With the progressive latent model, self-learned detectors are comparable to supervised ones, taking a step towards self-learning cameras

https://www.ucassdl.cn qxye@ucas.as.cn