Combining ROI-base and Superpixel Segmentation for Pedestrian Detection

Ji Ma[1,2, a], Jingjiao Li[1], Zhenni Li[1] and Li Ma[2]

[1] College of Computer Science and Engineering, Northeastern University, Shenyang, China;

[2] College of Information, Liaoning University, Shenyang, China.

[a] Maji@lnu.edu.cn

Keywords: ROI, salient object detection, superpixel, DPM.

Abstract. Pedestrian Detection is a hot topic in recent years, which is attracting a large number of scholars. The detection models are developing from simple models to complex models and the detection accuracy has been greatly improved. DPM (deformable part model) become the best pedestrian detection model and also attracted many scholars to modify it. The biggest problem caused by complex models is low detection efficiency for the real-time application with the sliding windows framework. Meanwhile, the latent SVM algorithm in DPM mining parts information is greatly affected by the initialization of parts, and there is no exact solution. Aiming at the drawbacks of DPM, using the research achievement of salient object detection an background detection, we propose a novel pedestrian detection framework based on ROI and superpixel segmentation. Contrasting with DPM in experiments, our method have greatly improved in accuracy and efficiency. The proposed framework has the same reference to other complex models.

1. Introduction

Now a days, pedestrian is an important research direction over a large public areas, such as surveillance, image retrieval, robotics, and automotive assistant. In these situation, pedestrian detection is becoming a popular topic in the computer vision field. During recent years, a lot of effort [3,4, 5, 6, 7, 8,11] has been devoted to this field. In a general way, we can consider pedestrian detection as a part of visual search problem [10]. The method with the histograms of oriented gradients (HOG) features and linear SVM learning machine, proposed by Dalal in [8], has been proven as a effective method to detect pedestrians. Felzenszwalb[3] introduced the Deformable Part-based Model (DPM), which divides an object into a root block with several deformable parts, and could solve the problem of the appearance variations and occlusion to achieve the excellent performance.

Because the scale and position of the pedestrian are unknown, the methods usually adopt sliding window framework with residing the sliding window many times called multiple scales detection [9]. The precision and efficiency of methods are significantly affected by the choice of the step width and a proper scale of the sliding window. In the situation, we consider to select the candidate location of pedestrians by salient object detection firstly and detect the pedestrian further. As the Fig .1 shows that the salient object detection results for pedestrians and our detection framework. The red boxes are salient object region and the blue boxes are the pedestrian regions detected by our method.

In this paper, a novel ROI-base method combining the salient object detection model and the DPM[3] are proposed for detecting the pedestrians in images. Meanwhile, we improve the algorithm of sample mining in DPM to enhance the performance of pedestrian detetion. It is easy to see from Fig. 2 that our detection process focus on effective local regions of the image called ROI rather than the entire. We can get the ROI by combining salient object detection and HOG pyramid. And the feature of DPM [3] is extracted from the ROI to detect the pedestrian object rapidly.

Summarizing, our contribution is two-fold. (1) we make use of the ROI selection as the pedestrian regions rather than all of the sliding windows in images to reduce the false alarm rate and the computational complexity. (2) we improve the algorithm of sample mining in DPM by background detection and superpixel segmentation in the parts initialization.

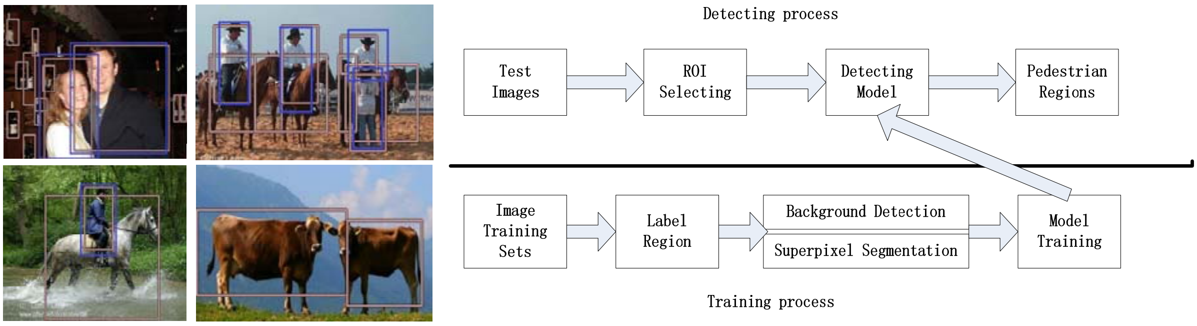

Fig. 1 ROI selection for pedestrian detection framework

2. Related works

2.1 Salient object detection.

ROI selection by salient object detection must be rapid as much as possible for pedestrian detection. Research from cognitive psychology and neurobiology suggest that humans have a natural ability to perceive objects before recognizing them. Considering the human reaction time with the biological signal transmission time and human attention theories, we hypothesize that the human vision system processes a few parts of an image in detail, while ignoring others. We further suggest that the human vision system have simple mechanisms to select possible object regions rapidly before recognizing objects.

The ability of perceiving objects before recognizing them is closely related to bottom up visual attention. The research developing in three aspects. Fixation prediction models pay attention to predicting saliency points of human eye movement[12]. Although this models have earned remarkable development, the prediction results include highlight edges and corners rather than the entire objects. So these models don't fit for object detection purpose. Salient object detection models segment the whole extent regions containing the most attention-grabbing objects in a scene [13]. However, these methods are not suitable for complicated images with presenting many rarely dominant objects. Objectness methods try to cover all objects in an image by proposing a small number of category-independent proposals [14]. However, these methods are expensive to compute.

In this paper, we introduce and improve the extremely simple and powerful feature BING[1] to help the search for ROI using objectness scores. The BING feature which requires only a few atomic CPU operations (i.e. ADD, BITWISE SHIFT, etc.) is simplicity, contrasting with recent state of the art techniques[14].

2.2 DPM model.

A lot of different pedestrian detection methods have been already evaluated using the PASCAL VOC challenge datasets[15]. DPM[3] is a detection method which achieving state-of the-art results in AP(Average Precision) for pedestrian detection on the PASCAL benchmarks.

All part-base models for pedestrian involve linear filters that are applied to dense feature maps. DPM[3] use the HOG features from [8]. A filter ***P***, is a rectangular template defined by an array of d-dimensional weight vectors with sizes . The response of a filter ***P*** at a position of feature map ***H*** is the "dot product" of the filter and a subwindow of the feature map ***H*** with the same size. A model for a pedestrian with n parts is formally defined by a n+2-tuple $(\boldsymbol{P_0},\boldsymbol{P_1},...,\boldsymbol{P_n},\boldsymbol{b})$, including a root filter, part filters and a real-valued bias term. The score of a hypothesis is given by the scores of each filter at their suitable locations, with subtracting a deformation cost dependant on the relative position between each part and the root, plus the bias, describing in formula (1).

$$score(\boldsymbol{p}_0,...,\boldsymbol{p}_n)=\sum_{i=0}^{n}\boldsymbol{P}_i\cdot\phi(\boldsymbol{H},\boldsymbol{p}_i)-\sum_{i=1}^{n}\boldsymbol{d}_i\cdot\phi_d(dx_i,dy_i)+b \tag{1}$$

Where gives the offset of the i-th part relative to its anchor position

$$(dx_i,dy_i)=(x_i,y_i)-(2(x_0,y_0)+v_i) \tag{2}$$

$$\phi_d(dx,dy) = (dx,dy,dx^2,dy^2) \tag{3}$$

The model training process have been presented in [3]. Note that this is a weakly labeled situation because there are not part labels or part locations with bounding boxes. Parameter learning is done by the latent SVM (LSVM) using the coordinate descent approach together with the data-mining and gradient descent algorithms.

3. ROI selection and superpixel learning

3.1 ROI selection.

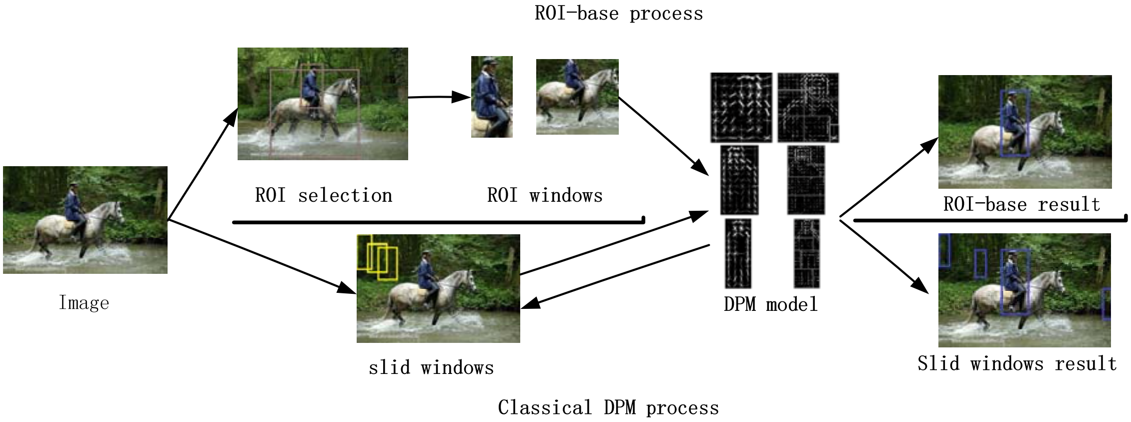

Fig. 2 the pedestrian detection processes contrasted between our method and classical DPM method

Pedestrian detection is a kind of object detection. Objects are independent things with specific closed boundaries and centers [14]. We use the method BING[1] to find candidate objects regions as ROI rapidly and then apply DPM method to detect pedestrian in detail. BING is self-adapting to translation, scale and aspect ratio. BING is both fastest and accurate with covering 96.2% true object windows in the PASCAL VOC2007 dataset.

Because DPM method generate HOG pyramid, we improve the BING by mix the HOG features with BING features to enhance the accuracy of ROI selection. Especially, at the situation of higher DR, the search window have been significantly reduced and it is in favor of the following pedestrian detection. Fig. 2 show the detection framework we proposal contrasting with classical DPM method and our method can remove some mistake in DPM.

3.2 Superpixel learning

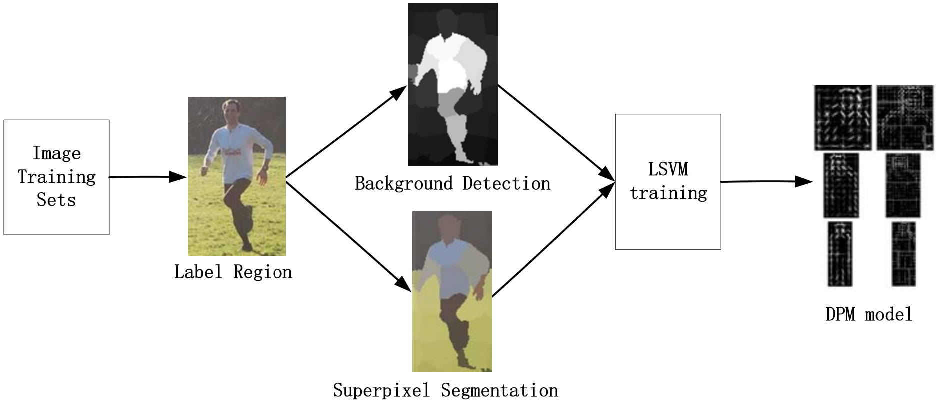

Fig. 3 DPM training process with background detection and superpixel segmentation

In training process, z called latent variable indicate the parts and deformable information. The definition in formula (4). In the training set, the parameter β is calculated by minimize the function (5). The detail description show in [3].

$$f_\beta(\boldsymbol{x}) = \max_{z\in Z(x)} \boldsymbol{\beta}\cdot\phi(\boldsymbol{x},\boldsymbol{z}) \tag{4}$$

$$L_D(\boldsymbol{\beta}) = \frac{1}{2}\|\boldsymbol{\beta}\|^2 + C\sum_{i=1}^{n} max(0,1-y_i f_\beta(x_i)) \tag{5}$$

The LSVM coordinate-descent algorithm as non-convex optimization is susceptible to local minima and thus sensitive to initialization. This is a common limitation with latent information for other methods as well. In DPM[3], initializing part filters uses a simple heuristic. By fixing six parts per component and using a small pool of rectangular part shapes, the method greedily place parts to cover high-energy regions of the root filter. Once a part is placed, the energy of the covered region of the root filter is set to zero. Then we search the next highest energy region, until six parts are chosen. The part filters initializing method is the uniform method for object detection and dosen't fit the situation of pedestrian detection as the hinge structure of body. Meanwhile, the rectangular label box in samples not only include pedestrian, but also contain background interference.

In order to resolve these problem, we introduce the background detection and superpixel segmentation in [2]. Fig.3 show our method vividly. We apply robust background detection for the image in rectangular label box and then extract the features as initializing parts in order to avoid interference of background. Meanwhile, for mining the structure information exactly, we apply the superpixel segmentation and make the parts fit the superpixel in part initialization. This progress costing vast time is done in training stage. So this method dosen't influence the efficiency in detecting stage. The experiments show that the exact model we trained make the detection accuracy increase.

4. Experiment

In this section, we used the PASCAL VOC2007 still image database to compare our approach against DPM[3].Why not compare to other methods? Since the ROI region selection is the preprocessing stage before main detection method, the effectiveness we proved in DPM also react on other method in slide window framework. Meanwhile, the benefits for the background detection and superpixel segmentation for training are suitable for other model. Fig.4 shows the PR curve and some real detection results. The red box indicate the ROI selection and the blue box indicate the pedestrian detection results. We can see our method is better than DPM in PR curve. We also perceive the relationship between the number of ROI search windows(#WIN) and the detection accuracy. Following the number of ROI search windows increasing, the accuracy of detection is enhanced with the complexity increasing. The relate concerns are described in [1]. Through comprehensive testing, the average detection time have been reduced to 20% of the original detection time.

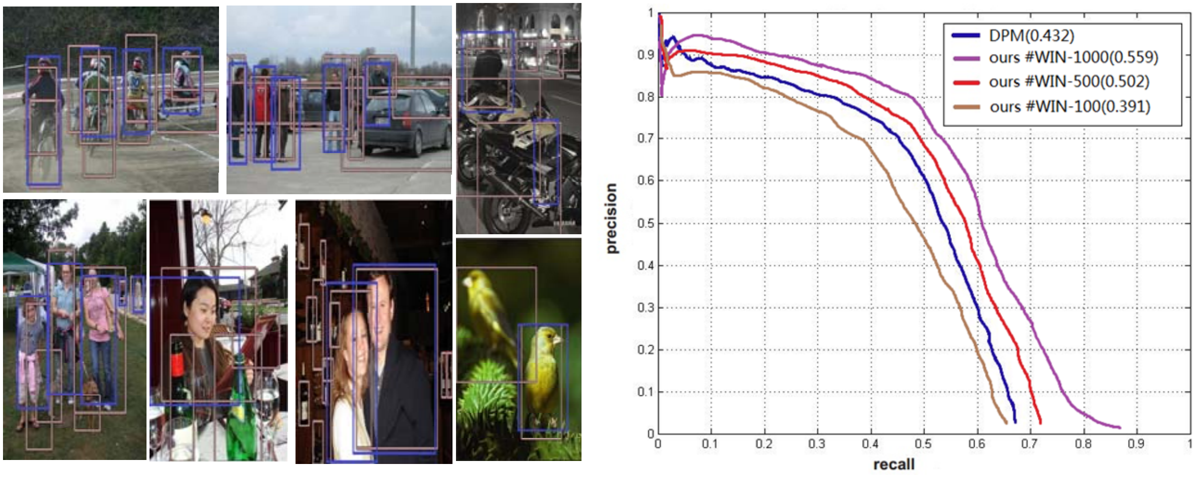

Fig. 4 PR curve and some detection results

5. Summary

The method we proposed make the accuracy and efficiency improvement. Meanwhile, this method also has the same benefits for other pedestrian detection model. In the future, we will study superpixel segmentation for model training deeply in order to improve the accuracy of model training.

References

[1]. Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, Philip Torr. BING: Binarized Normed Gradients for Objectness Estimation at 300fps. IEEE CVPR, 2014.

[2]. Wangjiang Zhu, Shuang Liang, Yichen Wei, and Jian Sun. Saliency Optimization from Robust Background Detection. IEEE CVPR, 2014.

[3]. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D. Object detection with discriminatively trained part-based models. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.32(2010)No.9,p.1627–1645.

[4]. Ouyang, W., Zeng, X., Wang, X. Modeling mutual visibility relationship in pedestrian detection. In: 2013 IEEE CVPR, 2013,p. 3222–3229.

[5]. Paisitkriangkrai, S., Shen, C., Hengel, A.V.D.: Efficient pedestrian detection by directly optimizing the partial area under the roc curve. IEEE ICCV, 2013,p. 1057–1064.

[6]. Yan, J., Lei, Z., Yi, D., Li, S.Z.: Multi-pedestrian detection in crowded scenes: A global view. IEEE CVPR), 2012,p. 3124–3129.

[7]. Zeng, X., Ouyang, W., Wang, X. Multi-stage contextual deep learning for pedestrian detection. IEEE ICCV, 2013,p. 121–128.

[8]. Dalal, N., Triggs, B. Histograms of oriented gradients for human detection. IEEE CVPR, 2005, p. 886–893.

[9]. Doll´ ar, P., Belongie, S., Perona, P. The fastest pedestrian detector in the west. BMVC, 2010, p. 7.

[10]. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. IEEE CVPR, 2012,p. 2288–2295.

[11]. Dollar, P., Wojek, C., Schiele, B., Perona, P. Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.34(2012)No.4, p .743–761.

[12]. A.Borji, D. Sihite, and L. Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. IEEE TIP, 2012.

[13]. Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille. The secrets of salient object segmentation. IEEE CVPR, 2014.

[14]. B. Alexe, T. Deselaers, and V. Ferrari. Measuring the objectness of image windows. IEEE TPAMI, Vol.34(2012)No.11, p.2189-2202.

[15]. M. Everingham, L. Van Gool, J. Winn and A. Zisserman, “The PASCAL Visual Object Classes (VOC) Challenge,” International Journal of Computer Vision. Vol. 88(2010)No.2, p. 303-338.