A Unified Framework for Concurrent Pedestrian and Cyclist Detection

Xiaofei Li, Lingxi Li, Fabian Flohr, Jianqiang Wang, Hui Xiong, Morys Bernhard, Shuyue Pan, Dariu M. Gavrila, and Keqiang Li

***Abstract*—Extensive research interest has been focused on protecting vulnerable road users in recent years, particularly pedestrians and cyclists, due to their attributes of vulnerability. However, comparatively little effort has been spent on detecting pedestrian and cyclist together, particularly when it concerns quantitative performance analysis on large datasets. In this paper, we present a unified framework for concurrent pedestrian and cyclist detection, which includes a novel detection proposal method (termed UB-MPR) to output a set of object candidates, a discriminative deep model based on Fast R-CNN for classification and localization, and a specific postprocessing step to further improve detection performance. Experiments are performed on a new pedestrian and cyclist dataset containing 30 490 annotated pedestrian and 26 771 cyclist instances in over 50 000 images, recorded from a moving vehicle in the urban traffic of Beijing. Experimental results indicate that the proposed method outperforms other state-of-the-art methods significantly.**

***Index Terms*—Multiple potential regions, pedestrian and cyclist detection, R-CNN, upper body detection.**

I. Introduction

SIGNIFICANT progress has been made over the past decade on improving driving safety with the development of Advanced Driver Assistance Systems (ADAS), such as pre-collision systems, crash imminent braking systems and others. In the last few years, however, extensive research interest has been focused on protecting vulnerable road users (VRUs), including pedestrians, cyclists and motorcyclists. According to the statistical data of WHO [1], half of the world's road traffic deaths occur among vulnerable road users. In some low- and middle-income countries, a much higher proportion of road users are pedestrians and cyclists. Among the vulnerable road users, pedestrians and cyclists are the weakest because there is no special protection device for them. Therefore, to make walking and cycling safer, detecting and protecting pedestrians and cyclists need to be paid more attention.

Many approaches based on different sensors are employed in vehicle environment perception systems, such as monocular camera, stereo camera, lidar and radar. Focusing on pedestrian and cyclist detection field, vision sensors are preferred, due to the possibility to capture a high-resolution perspective view of the scene with useful color and texture information, compared to active sensors [2]. Furthermore, vision technology is cost-effective and mature enough to handle many other tasks, such as lane detection and traffic sign detection.

Vision-based pedestrian detection has been studied for many years, but it is still a challenging problem due to the large variability in appearance, body pose, occlusion and cluttered backgrounds. Similar problems occur in the field of cyclist detection. In addition to the aforementioned problems, multiple viewpoints of cyclists bring more challenges to detect them, which is rarely taken into consideration in pedestrian detection. Cyclists can be viewed from a variety of possible orientations, which generates a problem to choose the detection window size as the aspect ratio of a cyclist differs from each orientation.

It's noted that traditional pedestrian or cyclist detection methods always consider pedestrians and cyclists separately [3], [4], although pedestrians and cyclists often appear in one picture. This often leads to scanning the input image several times and causing confused detection results, such as classifying cyclists as pedestrians, and vice versa, due to their similar appearance. In general, cyclists move faster than pedestrians, different attentions with pedestrians should be paid from ADAS or autonomous vehicles. Therefore, detecting pedestrians and cyclists concurrently and differentiating them clearly are urgently needed for the adaptive decision of ADAS and autonomous vehicles.

Generic object detection methods [5], [6], which are traditionally formulated as some detection proposals processed by a classifier, are good solutions to deal with the above issues. However, it's hard to learn a method to output a set of detection proposals that are likely to contain the objects, considering the large variability of pedestrian and cyclist instances. Whereas, the similarity between the two object classes is another challenging problem for pedestrian and cyclist classification.

To deal with these issues, a unified pedestrian and cyclist detector is presented in this paper, which can detect pedestrians

Manuscript received October 15, 2015; revised March 13, 2016 and May 5, 2016; accepted May 7, 2016. Date of publication July 7, 2016; date of current version February 1, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 51475254 and in part by the joint research project of Tsinghua University and Daimler. The Associate Editor for this paper was H. G. Jung. *(Corresponding author: Keqiang Li.)*

X. Li, J. Wang, H. Xiong, and K. Li are with the State Key Laboratory of Automotive Safety and Energy, Tsinghua University, Beijing 100084, China (e-mail: lixf11@mails.tsinghua.edu.cn; wjqlws@tsinghua.edu.cn; tomxiong@buaa.edu.cn; likq@tsinghua.edu.cn).

L. Li is with the Department of Electrical and Computer Engineering, Indiana University–Purdue University at Indianapolis, Indianapolis, IN 46202 USA (e-mail: ll7@iupui.edu).

F. Flohr and D. M. Gavrila are with the Environment Perception Department, Daimler Research and Development, 89081 Ulm, Germany, and also with the Intelligent Vehicles Section, TU Delft, 2628 CD Delft, The Netherlands (e-mail: fabian.flohr@daimler.com; dariu@gavrila.net).

M. Bernhard and S. Pan are with the Driver Assistance and Chassis Systems, Daimler Greater China Ltd., Beijing 100102, China (e-mail: bernhard.morys@daimler.com; shuyue.pan@daimler.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2567418

Fig. 1. Upper bodies of pedestrians and cyclists in different views.

and cyclists concurrently and differentiate them clearly. As the main contribution of this paper, the unified framework for concurrent pedestrian and cyclist detection involves a novel detection proposal method to output a set of object candidates, a discriminative deep model based on Fast R-CNN (FRCN) [7] for classification and localization, and a specific post-processing step.

The second contribution is a novel detection proposal method for pedestrian and cyclist detection, termed UB-MPR. We note that the upper bodies of pedestrians and cyclists are usually similar and visible, as shown in Fig. 1. So the upper body (UB) is utilized to extract object candidates where pedestrians or cyclists may appear. In order to propose potential object regions, multiple potential regions (MPR) around an upper body candidate that may cover the whole object (a pedestrian or a cyclist) are generated.

A FRCN-based deep model is deployed for concurrent pedestrian and cyclist detection task, followed by a specific post-processing step to further improve the detection performance, which is the third contribution of this paper.

Another contribution is the extensive comparative testing performed in this paper, on the large "Tsinghua-Daimler Cyclist Benchmark" [8] and on a sizable annotated pedestrian dataset.

The remainder of the paper is organized as follows. In Section II, the related work is presented, whereas Section III is an overview of our proposed system. In Section IV, the new pedestrian and cyclist detection dataset is introduced. In Section V, the evaluation protocol and system configurations are presented, and the performance of our approach is evaluated on the new dataset. The final conclusion and future work are given in Section VI.

II. Related Work

As mentioned above, vision-based pedestrian and cyclist detection is a challenging problem due to its practical use in the driving environment. Over the last decade, vision-based pedestrian detection has been extensively investigated, more than 60 methods are evaluated on the Caltech pedestrian detection benchmark until March 2016 [3]. Since an exhaustive survey of pedestrian detection is beyond the scope of this paper, interested readers are referred to some general surveys, such as [2], [9], [10] and the references therein, for an excellent review of pedestrian detection frameworks and benchmark datasets. Here five methods are mentioned as the representative landmarks. In 2003, Viola and Jones applied their VJ detector to the task of pedestrian detection [11]. Then Dalal and Triggs introduced the classical Histogram of Oriented Gradients (HOG) detector into the pedestrian detection task in 2005 [12]. Based on HOG detector, Deformable Part Model (DPM) was designed to weaken the deformation effect of non-rigid objects by Felzenswalb *et al.* in 2008 [13]. Another variant method ChnFtrs was applied to deploy multiple registered images channels for classification by Dollar *et al.* in 2009 [14]. In 2013, ConvNet model was introduced to yield competitive results on major pedestrian detection benchmarks by Sermanet *et al.* [15].

As opposed to pedestrian detection, very limited work has been undertaken in the domain of vision-based cyclist detection, although similar techniques are used for cyclist detection. Li [16] used HOG-LP features and linear SVM classifier to detect crossing cyclists, with the purpose of optimizing the time-consuming steps of HOG feature extraction. Chen [17] proposed a part-based bicycle and motorcycle detection for nighttime environments integrating appearance-based features and edge-based features. Cho [18] defined a mixture model of multiple viewpoints to detect cyclists, which was based on part-based representation, HOG feature and Support Vector Machine. In [19], a two-stage multi-model cyclist detection scheme was proposed for naturalistic driving video processing. An integral feature based detector was applied to filter out most of the negative windows, then the remaining potential windows were classified into cyclist or non-cyclist windows by three pre-learned view-specific detectors. In order to handle the multi-view problem of cyclists, the work proposed in [4] divided the cyclists into subcategories based on cyclists' orientation. For each orientation bin, they built a cascaded detector with HOG features trained based on the KITTI training dataset [20]. The work also explored the applications of geometric constraints to improve the detection performance.

A vision-based pedestrian and cyclist detection method was proposed by Fu [21], which is capable of recognizing the features of pedestrians and cyclists appeared in an image. The method harnessed the symmetry of objects, a two-wheeled recognition and plus a spatial relationship calculation between a cyclist and a vehicle, as a strategy to complete the detection process. Although this method could detect pedestrians and cyclists together, distributed processing was used to detect pedestrian and cyclist separately, then a cyclist confirmation step was processed according to a spatial relationship between the cyclist and the two-wheeled vehicle.

It's noted that most of the aforementioned works are capable of handling roughly rigid objects easily (such as pedestrians), but they have difficulty in detecting more deformable generic objects (such as cyclists). In the latter case, several view-specific detectors are required to deal with different aspect ratios. Since the resolutions of the object templates are fixed normally, an exhaustive sliding window search is required to

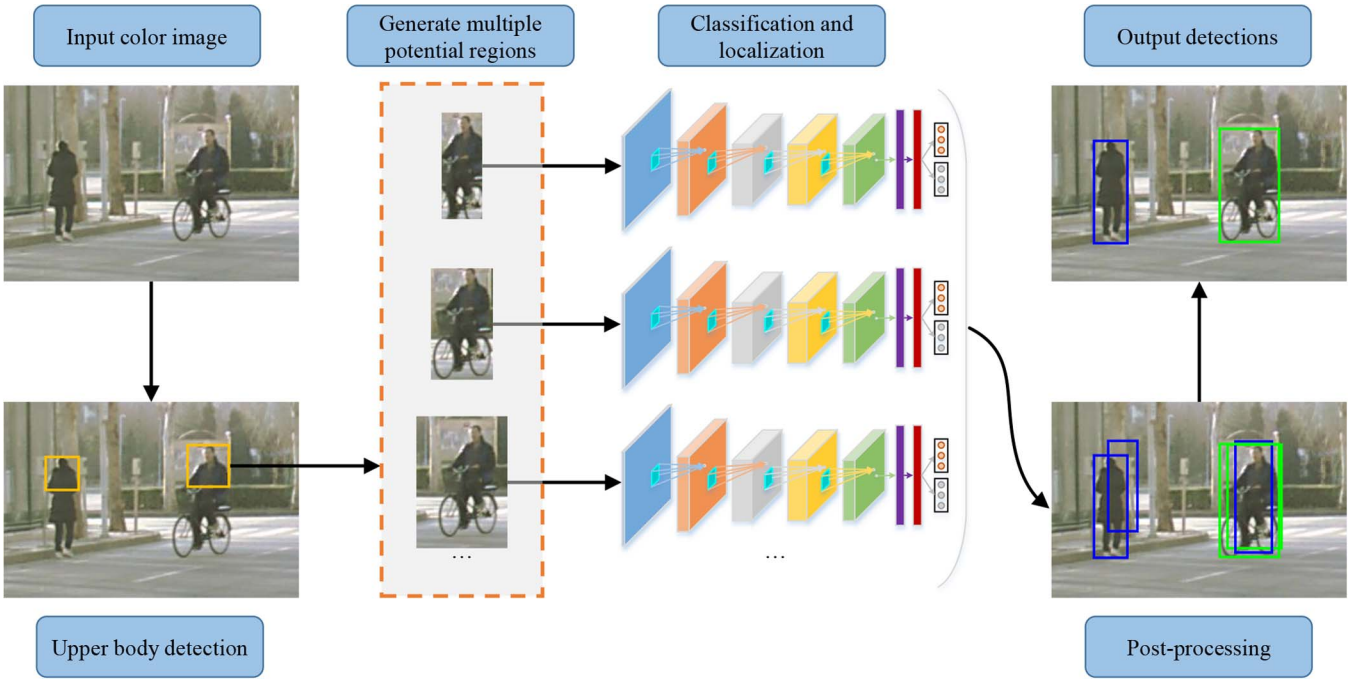

Fig. 2. Overview of the proposed pedestrian and cyclist detection method.

find objects at different scales with different aspect ratios, like DPM [13]. More recent approaches like [5], [22], may be alternative methods for pedestrian and cyclist detection, which utilize region proposal methods to generate potential bounding boxes in an image firstly, then classify these proposed boxes by a classifier, and finally use post-processing to refine the detection results. These methods can warp the proposal regions into fixed-size inputs regardless of the regions' aspect ratios and classify them concurrently, due to superior capabilities in learning a general object representation from large amounts of training data. However, it's hard to learn a method to output a set of detection proposals for pedestrians and cyclists, due to the large variability of pedestrian and cyclist instances.

We also note that most of the aforementioned methods detect pedestrian/cyclist separately, or cannot differentiate them clearly. For example, in the evaluation protocol of [3], ground-truth bounding boxes for cyclists were ignored and did not need to be matched, whereas, matches were not viewed as mistakes either. As a result, these methods always classified cyclists as positive pedestrian samples.

III. Proposed Methods

A. Overview

The proposed pedestrian and cyclist detector can be divided into three parts: 1) UB-MPR based detection proposal method, 2) Discriminative networks for object detection, and 3) Post-processing. The flowchart is shown in Fig. 2.

Firstly, an upper body detector is designed based on ACF framework [23] to extract upper body candidates where pedestrians or cyclists may appear. Around each upper body candidate, multiple potential regions that may cover the whole pedestrian or cyclist instance are generated. Then all the potential regions serve as inputs to a deep convolutional neural network to get classification probabilities and object localizations. Finally, a post-processing procedure specialized for the UB-MPR method is deployed to further enhance the detection performance.

B. UB-MPR Based Detection Proposal Method

1) Definition of Upper Body: The upper bodies of pedestrians and cyclists have similar appearance and pose, and can be visible in most cases, as shown in Fig. 1. A pedestrian or a cyclist may appear around the place where an upper body is detected, which is the basis of the unified framework for pedestrian and cyclist detection.

With respect to the division method of a body, several works can be referred to. For instance, Mogelmose *et al.* [23] and Liu *et al.* [25] divided the full body into two parts evenly, the upper half part was the upper body and the lower part was the lower body. Prioletti *et al.* [26] used two different compositions of body parts, two parts containing an upper body and a lower body, and three parts containing a head, a torso and legs. Zhang *et al.* [27] tested a three parts model and showed the head-shoulder area is more discriminative for pedestrian detection than other body parts.

Referring to the aforementioned works, the upper body containing the head and part of the torso is adopted in this work. The uppermost square of the *person* is chosen as the upper body of an object, whose side length is equal to half of the *person*'s height exactly, shown by the green bounding boxes in Fig. 3. Here, *person* indicates the pedestrian when the object is a pedestrian, and the rider when the object is a cyclist.

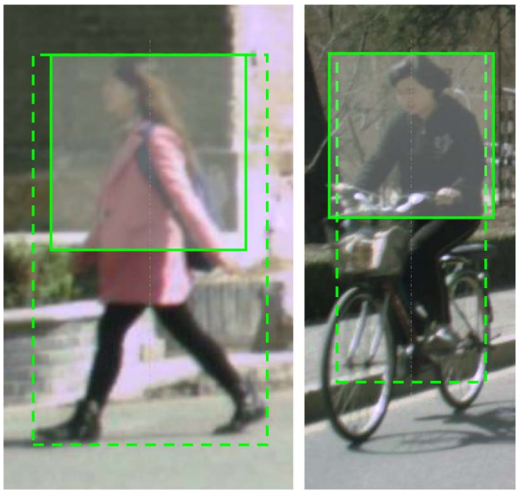

Fig. 3. The pedestrian or cyclist instance's upper body is defined as the uppermost square of the person, whose side length is equal to half of the person's height exactly. Green dashed and solid bounding boxes indicate the ground-truth bounding boxes of the person and the upper body, respectively.

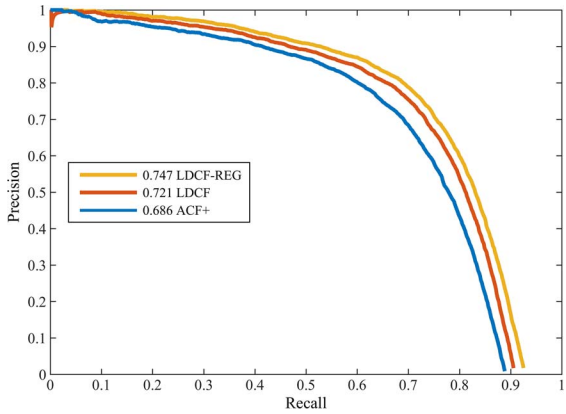

Fig. 4. Upper body detection performance in the new pedestrian and cyclist test dataset.

2) Upper Body Detector: The channel features detectors (firstly proposed by Dollar *et al.* [14], [23]) is utilized to detect the upper body in this work because the methods are conceptually straightforward and efficient. Specifically, for our work in this paper, Locally Decorrelated Channel Features (LDCF) variant [28] is employed due to the better performance.

Given an input image, LDCF computes several feature channels and removes correlations in local neighborhoods of feature channels, where each channel is a per-pixel feature map such that output pixels are computed from corresponding patches of input pixels. The same channel features as [28] are used: normalized gradient magnitude (1 channel), histogram of oriented gradients (6 channels), and LUV color channels (3 channels), 10 channels in total. We apply *RealBoost* with 4 rounds of bootstrapping to train upper body detector with 4096 depth-5 decision trees over the $h/2 \cdot w/2 \cdot 10$ aggregated features, where $h \times w$ is the input window and 2 is the down sample scale. To adapt to the size of upper body, the *modelDs* (model height and width without padding) is set to [20, 20] and *modelDsPad* (model height and width with padding) is set to [32, 32]. The upper body detector is trained using the new pedestrian and cyclist training dataset.

In order to generate a more competitive upper body detector, two variant detectors (ACF+ and LDCF [28]) were trained and evaluated in the new pedestrian and cyclist dataset (with moderate setting). Fig. 4 shows the upper body detection performance (the relationship between precision and recall rate) using the evaluation protocol developed by Dollar [3]. From the figure, we can see LDCF-based upper body detector outperforms ACF+-based method by 3.5% average precision. Therefore, LDCF is employed in the subsequent development.

3) Optimization of Upper Body Detections: Since MPR that may cover the whole pedestrian and cyclist are generated based on the upper body candidates, the localization accuracy of the upper body has a great influence on the performance of following procedures. In order to improve the localization accuracy of upper body candidates, a linear regression model is trained to predict an optimal window based on the channel features, inspired by the work in DPM [13] and R-CNN [5].

The target of this step is to map an upper body candidate $U(x_U^c, y_U^c, w_U, h_U)$ into the ground-truth of the upper body $U_G(x_{UG}^c, y_{UG}^c, w_{UG}, h_{UG})$ using a transformation, where (x_U^c, y_U^c) and (x_{UG}^c, y_{UG}^c) indicate the x, y coordinates of the central point of the upper body candidate and ground-truth, respectively, (w_U, h_U) and (w_{UG}, h_{UG}) indicate their widths and heights. The following transformation is deployed to transform an input bounding box U into a predicted bounding box $\hat{U}_G(\hat{x}_{UG}^c, \hat{y}_{UG}^c, \hat{w}_{UG}, \hat{h}_{UG})$:

$$\begin{cases} \hat{x}_{UG}^c = x_U^c + w_U d_x \\ \hat{y}_{UG}^c = y_U^c + h_U d_y \\ \hat{w}_{UG} = w_U \exp(d_w) \\ \hat{h}_{UG} = h_U \exp(d_h). \end{cases} \tag{1}$$

Here, each d_* is the transformation parameter modeled as a linear function of the vectorized aggregated channel features $f(U)$ of the input upper body candidate: $d_* = w_*^T f(U)$, where w_* is the model parameter to be learned. The model parameters can be learned as a standard regularized least squares problem by minimizing the loss function

$$\text{loss} = \frac{1}{2}\sum_i^N \left(\hat{U}_G^i - U_G^i\right)^2 + \frac{\lambda}{2}\|w_*\|^2. \tag{2}$$

Here, N is the number of training samples, λ is the regularization factor, which is set as 1000 in this paper. Only upper body candidates close to ground-truth samples are employed as training samples. The threshold of Intersection-over-Union (IoU) overlap threshold between upper body candidates and upper body ground-truths is chosen as 0.5 in this work.

The detection performance of LDCF-based upper body detector optimized by localization regression (LDCF-REG) is shown in Fig. 4, which shows that the introduced LDCF-REG outperforms LDCF by 2.6% improvement in average precision for upper body detection. It is worth mentioning that LDCF-REG can achieve 92.5% recall rate in the test dataset (with moderate setting) with an IoU overlap threshold as 0.5.

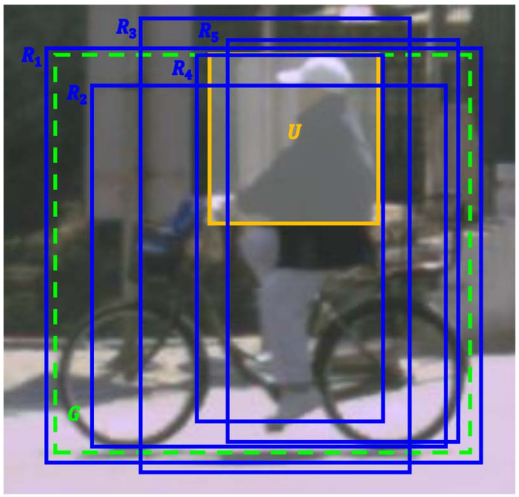

Fig. 5. Relationship among an upper body candidate, the corresponding object ground-truth, and MPR. Five potential regions are shown for example. Green, yellow, and blue bounding boxes indicate ground-truth bounding box of the cyclist, detected upper body candidate, and corresponding potential regions, respectively.

4) Definition of MPR: As we assumed before, each pedestrian or cyclist instance has a visible upper body. If we get the position of an upper body, we can estimate the rough location of the object bounding box. In this section, we introduce a novel idea to generate pedestrian and cyclist proposals, which can generate Multiple Potential Regions (MPR) around an upper body candidate, with the purpose of overlapping ground-truth bounding boxes as much as possible. The process of designing MPR is to select a group of representative transformation parameters, which can transform an input upper body candidate into M potential regions. Fig. 5 shows the relationship among an upper body candidate, the corresponding object ground-truth and MPR.

Firstly, we formulize the relationship between an upper body candidate U and the corresponding object ground-truth G

$$\begin{cases} x_G^c = x_U^c + \kappa_x w_U \\ y_G = y_U + \kappa_y h_U \\ w_G = \kappa_w w_U \\ h_G = \kappa_h h_U. \end{cases} \tag{3}$$

Because $x_G^c = x_G + w_G/2$ and $x_U^c = x_U + w_U/2$, so

$$\begin{cases} x_G = x_U + (\kappa_x - \kappa_w/2 + 1/2) w_U \\ y_G = y_U + \kappa_y h_U \\ w_G = \kappa_w w_U \\ h_G = \kappa_h h_U. \end{cases} \tag{4}$$

Here, κ_* indicates the factor; (x_U, y_U) indicate the x, y coordinates of the left-top point of the upper body, w_U and h_U indicate its width and height; (x_G, y_G) indicate the x, y coordinates of the left-top point of the object ground-truth, w_G and h_G indicate its width and height; x_U^c and x_G^c indicate the x coordinates of the central point of the upper body and object ground-truth, respectively.

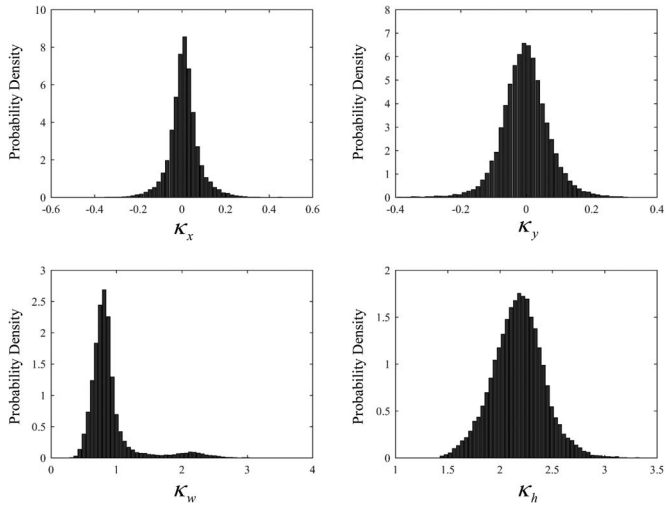

Fig. 6. Histograms of κ_*.

Based on the above equations, we can calculate κ_* by the following formula:

$$\begin{cases} \kappa_x = \dfrac{(x_G + w_G/2 - x_U - w_U/2)}{w_U} \\ \kappa_y = \dfrac{(y_G - y_U)}{h_U} \\ \kappa_w = \dfrac{w_G}{w_U} \\ \kappa_h = \dfrac{h_G}{h_U}. \end{cases} \tag{5}$$

In order to observe the distribution of κ_* directly, the histograms of κ_* calculated by coupled upper body candidates and object ground-truths $\{U^i, G^i\}$ are shown in Fig. 6. From the histograms, it can be seen that the distributions of κ_* match with normal distributions approximately, which illustrates the relationship between upper body candidates and the object ground-truths is formulized properly. It is worth noting that only upper body candidates close to ground-truth samples are employed to calculate κ_* and optimize MPR parameters. The threshold of IoU overlap between upper body candidates and upper body ground-truths is chosen as 0.5 in this work.

If a group of representative transformation parameters $\mathcal{K}$ $(\kappa_*^1, \ldots \kappa_*^m, \ldots, \kappa_*^M)$ are chosen, we can get MPR to estimate the position of the object ground-truth by the following formula:

$$\begin{cases} x_R^m = x_U + (\kappa_x^m - \kappa_w^m/2 + 1/2)\, w_U \\ y_R^m = y_U + \kappa_y^m h_U \\ w_R^m = \kappa_w^m w_U \\ h_R^m = \kappa_h^m h_U. \end{cases} \tag{6}$$

Here, (x_R^m, y_R^m) indicates the x, y coordinates of the left-top point of the m-th potential region, w_R^m and h_R^m indicate its width and height, respectively.

5) Optimizing MPR Parameters: Intuitively, we can sample MPR parameters based on some importance sampling methods based on the joint probability distribution of κ_*. However, the diversity of MPR parameters is hard to be guaranteed because most of the parameters may concentrate in the central region. In

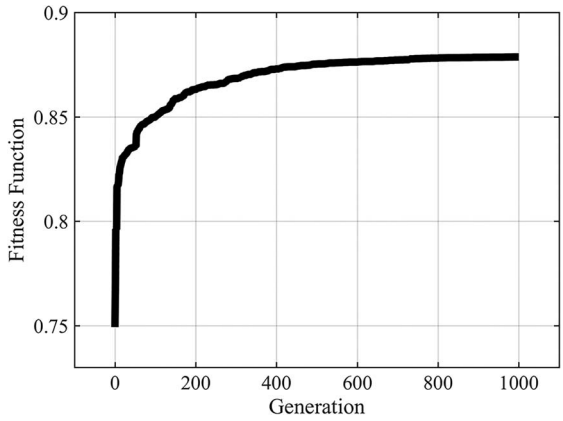

Fig. 7. Evolution of the fitness function.

order to choose MPR parameters optimally, Genetic Algorithm (GA) [29] is deployed to search an ideal solution.

Genetic Algorithm (GA) is an adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. The algorithm usually iterates many generations, starting from a population of randomly generated individuals. In each generation, the fitness of each individual in the population is calculated, which will be the clue to select fit individuals using a fitness proportionate selection method. Selected individuals generate a new generation population after being modified by crossover and mutation operators. The algorithm usually terminates when either a maximum number of generations has been produced or other conditions have been reached.

A group of MPR parameters $\mathcal{K}$ $(\kappa_*^1, \cdots \kappa_*^m, \cdots, \kappa_*^M)$ is represented as an individual. M indicates M potential regions are chosen around one upper body candidate, we set $M = 40$ in this paper. So the variable number of an individual is $4 \times M$. The objective of optimizing the population is to find the best individual which can overlap ground-truth bounding boxes from the training dataset as much as possible. For each object ground-truth, it just requires one potential region which can match it well. Therefore, the Intersection-over-Union (IoU) overlap rate of R^m and G is calculated as the objective function

$$\text{fitness} = \sum_{i=1}^{I} \max_{m \in [1,M]} \left(\text{overlap}\left(R_m^i, G^i\right)\right). \tag{7}$$

where, overlap$(\cdot)$ is a function to compute the IoU overlap rate, I is the number of all the coupled upper body candidates and object ground-truths. Besides, the roulette wheel selection method is deployed to select individuals, the size of population is 100, the generation is 1000, the crossover probability is 0.8 and the mutation probability is 0.2. After 1000 generations, the evolution of the fitness function can be found in Fig. 7. It is worth mentioning that MPR with optimized parameters can achieve 96.5% recall rate in the test dataset (with moderate setting) with an IoU overlap threshold 0.5 if maximal 50 upper body candidates per image are considered, which means the UB-MPR proposal method is very effective.

Algorithm 1 provides the detailed training procedure of the UB-MPR proposal method, which mainly consists of training the upper body detector, optimizing upper body detections and optimizing MPR parameters.

Algorithm 1: Training the UB-MPR detection proposal method

Input: The new pedestrian and cyclist dataset
Output: Upper body detector $\mathcal{D}$, upper body localization regression $\mathcal{R}$, MPR parameter vector $\mathcal{K}$

Initialize upper body samples: positive samples $\mathcal{P} \leftarrow$ training set 1 and 3, negative samples $\mathcal{N} \leftarrow \phi$
for $n = 1$ to num-rounds **do**
  Mine (hard) negatives $\mathcal{N}_n$ from training set 2 and non-VRU set
  Append $\mathcal{N}_n$ to $\mathcal{N}$
  $\mathcal{D} = train{-}detector(\mathcal{P}, \mathcal{N})$
end
for $i = 1$ to num-images **do**
  $dt_i \leftarrow$ detect upper body in i-th image
  $m{-}dt_i$, $m{-}gt_i \leftarrow$ match dt_i with i-th ground-truth $upper{-}gt_i$
  Append $m{-}dt_i$ to $m{-}dt$, $m{-}gt_i$ to $m{-}gt$
end
$\mathcal{R} = train{-}regressor\ (m{-}dt, m{-}gt)$
for $i = 1$ to num-images **do**
  $r - dt_i \leftarrow$ optimize dt_i by $\mathcal{R}$
  $mr{-}dt_i$, $mr{-}gt_i \leftarrow$ match $r{-}dt_i$ with i-th ground-truth $upper{-}gt_i$
  Append $mr{-}dt_i$ to $mr{-}dt$, $mr{-}gt_i$ to $mr{-}gt$
end
Initialize population $\mathcal{K}_0$ for genetic algorithm
$\mathcal{K} = genetic{-}algorithm\ (\mathcal{K}_0, mr{-}dt, mr{-}gt)$
return $\mathcal{D}$, $\mathcal{R}$ and $\mathcal{K}$

C. Discriminative Networks for Object Detection

It has been proved that deep network models are potentially powerful in handling complex tasks, such as pedestrian detection [15], [30]. Recent advances in object detection are driven by the success of R-CNN, which involves a category-independent region proposal method to extract a set of candidate detections, a large convolutional neural network to extract object feature vectors and a linear SVM to classify object classes.

As an upgraded version of R-CNN [5], FRCN [7] is deployed to detect pedestrians and cyclists in this paper. Unlike original R-CNN, FRCN trains a single-stage multi-task loss network. The inputs of the network are the whole image and a set of proposals. After several convolutional and max pooling layers, a region of interest pooling layer extracts a fixed-length feature vector from the feature map. Finally, two sibling output layers, which produce classification probability and bounding box regression, are connected after a sequence of fully connected layers.

Unlike the original method, which either uses Selective Search [31] or a Region Proposal Network (Faster R-CNN [32]) for extracting relevant proposals, we utilize UB-MPR method for proposal generation that is described in the previous section.

Fig. 8. Overview of the new pedestrian and cyclist detection dataset. (a) Pedestrian samples. (b) Cyclist samples. (c) Test images with annotations: Blue, green, and yellow bounding boxes indicate pedestrians, cyclists, and other riders, respectively.

D. Post-Processing

After the aforementioned procedure, classification probabilities and bounding boxes of $M \times N$ proposals are produced, where N is the number of considered upper body candidates per image. Intuitively, the bounding box with the highest classification probability may be the object that we want to find. However, this will lead to ambiguous situations sometimes. Take a cyclist instance for example, the region covering the cyclist may get a high classification score, but another region covering the rider may get a high score too. As a result, the object may be recognized as a pedestrian, which represents a confused case. Since the object category corresponding to an upper body candidate can be indicated by its MPR, we aggregate all the classification probabilities of the MPR and classify them using linear Support Vector Machine (SVM) [33]. The aggregated classification probability feature can be represented as

$$f = \left(s_1^1, \ldots, s_1^M, s_2^1, \ldots, s_2^M, s_3^1, \ldots, s_3^M\right). \tag{8}$$

Here s_1, s_2, and s_3 indicate the classification probabilities of pedestrian, cyclist and background, respectively. Once all upper body candidates are classified correctly, we can ignore the influence of interference from other categories easily. Here, only upper body candidates close to ground-truth samples (IoU overlap higher than 0.5) in the training dataset are employed to train the classifier.

We usually get multiple overlapping detections for each object. Thus, a greedy procedure via non-maximum suppression (NMS) is used to eliminate repeated detections. Since an upper body candidate can generate a proposal group with M proposals, among which only one proposal is valid. After NMS procedure, several proposals from one group may be retained. Thus we only reserve the proposal with the highest score from one group.

TABLE I
STATISTICS OF THE NEW PEDESTRIAN AND CYCLIST DATASET

	Training				Test	Total
	Set 1	Set 2	Non-VRU	Set 3	Set	
Total Frames	9741	5095	1000	22780	14570	53186
Labeled Frames	9741	1019	1000	4556	2914	19230
Total BBs	16202	3016	0	28732	13143	61093
Cyclist BBs	16202	1301	0	4610	4658	26771
Pedestrian BBs	0	1539	0	21571	7380	30490
Other rider BBs	0	176	0	2551	1105	3832

IV. A NEW PEDESTRIAN AND CYCLIST DATASET

Challenging datasets have promoted technological progress in computer vision. There are already some publicly available pedestrian datasets, such as the INRIA [12], Caltech [3] and Daimler [9], [34] pedestrian detection datasets, which promote the development of pedestrian detection. Although cyclists are often encountered in traffic accidents, there is no challenging cyclist dataset publicly available yet, except the KITTI object detection benchmark [20]. However, there are very limited cyclist instances (no more than 2000) in the training set, which might not be sufficient for cyclist detection and evaluation. Therefore, our group introduced a public "Tsinghua-Daimler Cyclist Benchmark" recently [8], which contained plenty of annotated cyclists. In order to train and evaluate the proposed method in this work, we present a new pedestrian and cyclist detection dataset, which supplements the public cyclist dataset with a richly annotated pedestrian dataset.

An excerpt from the new pedestrian and cyclist detection dataset is shown in Fig. 8. We add a fully labeled dataset, termed Training Set 3, into the cyclist dataset to supplement plenty of pedestrian instances. Detailed labeling rules are the same as the description in the previous work [8]. Statistics about the new dataset can be found in Table I.

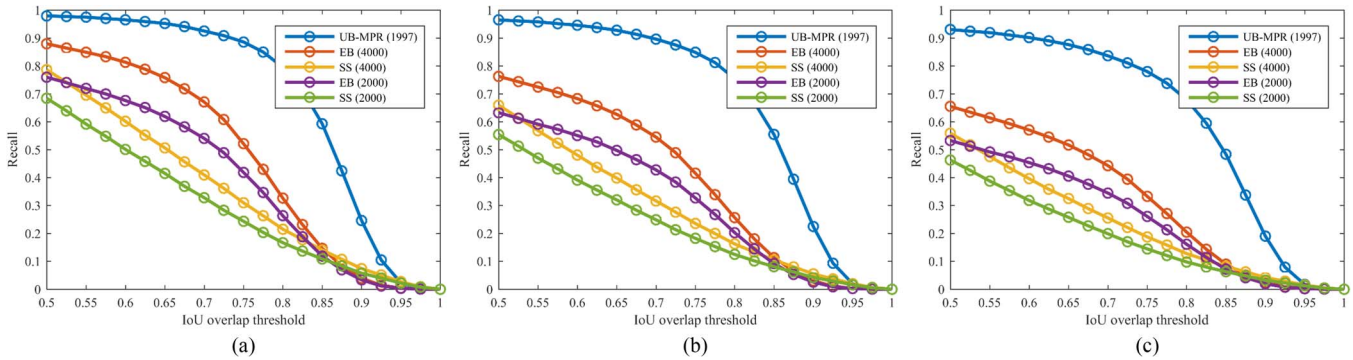

Fig. 9. Recall rate versus IoU threshold of different proposal methods shown for the new pedestrian and cyclist test dataset with different settings. UB-MPR, EB, and SS indicate our proposal method, Edge Boxes, and Selective Search, respectively. The number of proposals is also listed following each method's name. (a) Easy. (b) Moderate. (c) Hard.

V. Experiments

In this section, evaluation parameters and detailed settings of different methods are described. The performance evaluation of different proposal methods and detectors on the new pedestrian and cyclist test dataset is compared and discussed.

A. Evaluation Protocol

As we used in [8], the well-established methodology used in the PASCAL object detection challenges [35] is utilized to show the relationship between precision and recall rate. Meanwhile, the average precision (AP) is used to summarize the performance of precision/recall curve. To assign the output detections to ground-truth objects, the PASCAL measure is employed, which states that the area of IoU overlap must exceed the threshold of 0.5.

In order to evaluate performance in various subsets of the new pedestrian and cyclist dataset, we define three difficulty levels as follows:

— Easy: pedestrians and cyclists with bounding boxes higher than 60 pixels and fully visible.
— Moderate: pedestrians and cyclists with bounding boxes higher than 45 pixels and less than 40% occlusion.
— Hard: pedestrians and cyclists with bounding boxes higher than 30 pixels and less than 80% occlusion.

During evaluation in a subset, the objects not included in the subset are ignored instead of discarded directly, which need not to be matched with detections. Besides, we also want to evaluate the capacity of the detectors to differentiate pedestrians and cyclists. Therefore we also compare the detection performance between ignoring and discarding the other objects in the following experimental sections.

B. Parameter Configuration

The latest version of Dollar's Computer Vision MATLAB Toolbox [23] was applied in this work to train the upper body detector. The pedestrian and cyclist instances were extracted from training set 1 and set 3 with moderate setting, and the negative samples were sampled from training set 2 and non-VRU set. Only upper body candidates close to ground-truth samples (IoU overlap higher than 0.5) are employed to calculate and optimize MPR parameters. The number of considered upper body candidates per image is limited to 50 in this paper.

For training the deep networks with UB-MPR proposal methods, the open source of FRCN [7] with pre-trained ZF-nets was applied in this work. During the fine-tuning procedure, the final sibling layers were adapted to this task. Each SGD mini-batch was constructed by 2 images. The input image size was set to 2048×1024. The first image of a batch was chosen from training set 1 or set 3, and the second image was chosen from training set 2 or non-VRU set. Both of the images were chosen uniformly at random from corresponding dataset. We used mini-batches of size 128, sampling positive samples (max 25% of batch size) from the first image with a minimum IoU overlap of 0.5 to a ground-truth bounding box. The left negative samples were sampled from all the images of the training dataset with a maximum IoU overlap of 0.5: negatives around positive samples could be extracted from the first image; hard negatives and additional normal negatives might be extracted from the second image. We did bootstrapping every 10 000 iterations to mine hard negative samples. We used a learning rate of 0.001 for 40 000 iterations, and 0.0001 for the next 20 000 iterations. Other network's configurations and parameters were the same as the original paper [7].

Besides, in order to compare different proposal methods, two state-of-the-art proposal methods, Selective Search (SS) [31] and Edge Boxes (EB) [36] were also considered for training pedestrian and cyclist detectors. For Selective Search, we followed the same settings that were used in R-CNN [5], and we got about 6000 proposals per image in the training dataset. For Edge Boxes, we used the default parameters the same as the original paper, and got 4000 proposals per image in the training dataset.

In addition, ACF-based and LDCF-based pedestrian and cyclist detectors were also considered for comparisons. During training ACF and LDCF detectors, we extracted pedestrian and cyclist instances from training set 1 and set 3, and extracted negative samples from training set 2 and non-VRU set. We

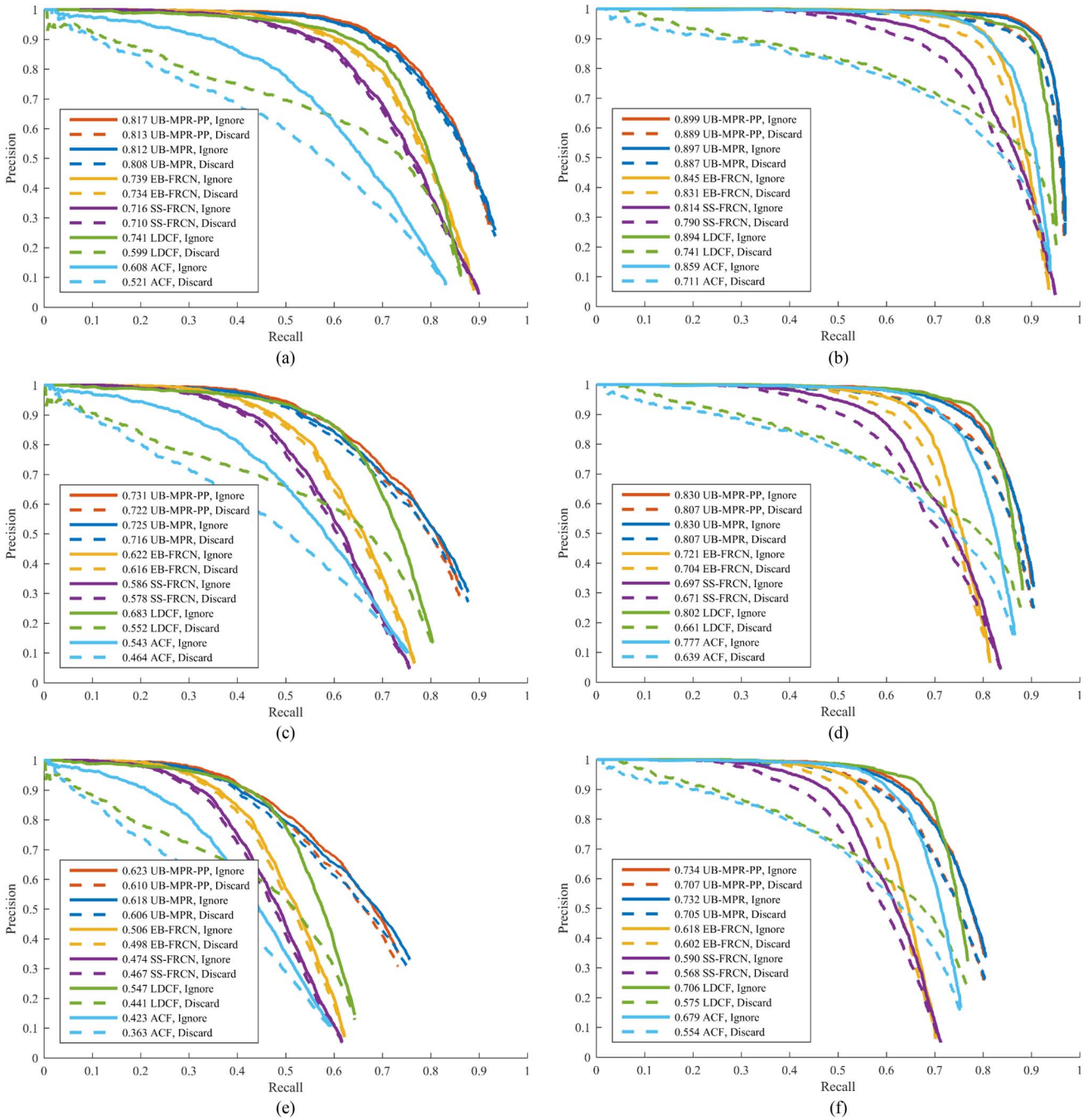

Fig. 10. Precision versus recall curves of various detectors shown for the new pedestrian and cyclist test dataset with different settings. The AP is listed before the name of each method. "Ignore" indicates ignoring cyclist (pedestrian) instances when evaluating the performance of pedestrian (cyclist) detection, and "Discard" indicates discarding cyclist (pedestrian) instances directly, which are plotted in solid and dashed lines, respectively. UB-MPR-FRCN-PP and UB-MPR-FRCN are shorted for UB-MPR-PP and UB-MPR, respectively. (a) Pedestrian, easy. (b) Cyclist, easy, (c) Pedestrian, moderate. (d) Cyclist, moderate. (e) Pedestrian, hard. (f) Cyclist, hard.

divided the positive cyclist instances into three classes to train three cyclist detectors, just like we did in [8]. Meanwhile, another detector for pedestrian detection was also trained in the same way. The same parameters used in the original application [23] were utilized to train the pedestrian and cyclist detector.

During the test phase, all the detectors mentioned above deployed a greedy fashion of non-maximum suppression to suppress bounding boxes with lower scores, just like the methods used in ACF [23].

C. Comparisons With Other Proposal Methods

In order to validate the performance of the presented proposal method, we compute the recall rate of different proposals at different IoU ratios with ground-truth bounding boxes, shown in Fig. 9. SS and EB methods are utilized with default parameters. The N proposals are the top-N ranked ones based on their confidences. We consider different numbers (2000 and 4000) for SS and EB.

TABLE II
PEDESTRIAN DETECTION AVERAGE PRECISION IN THE TEST DATASET

method	AP (%) easy		moderate		hard	
	ignore	discard	ignore	discard	ignore	discard
UB-MPR-PP	**81.7**	**81.3**	**73.1**	**72.2**	**62.3**	**61.0**
UB-MPR	81.2	80.8	72.5	71.6	61.8	60.6
EB-FRCN	73.9	73.4	62.2	61.6	50.6	49.8
SS-FRCN	71.6	71.0	58.6	57.8	47.4	46.7
LDCF	74.1	59.9	68.3	55.2	54.7	44.1
ACF	60.8	52.1	54.3	46.4	42.3	36.3

TABLE III
CYCLIST DETECTION AVERAGE PRECISION IN THE TEST DATASET

method	AP (%) easy		moderate		hard	
	ignore	discard	ignore	discard	ignore	discard
UB-MPR-PP	**89.9**	**88.9**	**83.0**	**80.7**	**73.4**	**70.7**
UB-MPR	89.7	88.7	**83.0**	**80.7**	73.2	70.5
EB-FRCN	84.5	83.1	72.1	70.4	61.8	60.2
SS-FRCN	81.4	79.0	69.7	67.1	59.0	56.8
LDCF	89.4	74.1	80.2	66.1	70.6	57.5
ACF	85.9	71.1	77.7	63.9	67.9	55.4

The results show that the UB-MPR proposal method outperforms the other compared methods significantly, even with less proposed bounding boxes. Take the moderate subset as an example, when IoU overlap is 0.5, our UB-MPR method achieves 96.5% recall rate, which outperforms SS and EB (with average 2000 proposals per image) by 41.1% and 33.2%, respectively. Even when SS and EB use 4000 proposals, the UB-MPR method outperforms them by 30.5% and 20.2%, respectively. When IoU overlap is 0.75, the UB-MPR method achieves 84.8% recall rate, which outperforms SS (2000), EB (2000), SS (4000), and EB (4000) by 66.6%, 52.2%, 61.3%, and 43.3%, respectively.

Besides, we also found two trends from the comparative figures: with the test subsets becoming harder, the advantage of the UB-MPR method is increasingly obvious; when the IoU overlap is less than 0.9, the higher the IoU overlap is, the more obvious improvement of our method compared to the other methods.

D. Comparisons With Other Detectors

In this section, we compare the performance of our proposed method to other representative methods using the experimental protocol explained above. Fig. 10 illustrates the overall detection performance of all the detectors in the new pedestrian and cyclist test dataset with different settings. In order to compare different detectors directly, we also provide two summary tables in Tables II and III. From the figure and two summary tables, we can find that all the selected methods can get reasonable performances in different subsets. Among them, the proposed methods (UB-MPR-FRCN-PP and UB-MPR-FRCN) outperform the others under any conditions, which illustrates the effectiveness of our unified framework for pedestrian and cyclist detection.

When compared to other FRCN-based methods in the moderate subsets, UB-MPR-FRCN based pedestrian detector outperforms SS-FRCN and EB-FRCN by 13.9% and 10.3% AP, respectively, meanwhile UB-MPR-FRCN based cyclist detector outperforms SS-FRCN and EB-FRCN by 13.3% and 10.9% AP, respectively. The improvement of the performance is brought by the UB-MPR proposal method, which demonstrates the benefit of the new proposal method.

We also find LDCF and ACF based pedestrian detectors and cyclist detectors can get competitive results when ignoring cyclist and pedestrian instances, respectively. But with "discard" settings (discarding cyclists when evaluating pedestrian detectors, or discarding pedestrians when evaluating cyclist detectors), their performances drop significantly. Meanwhile, the performance of FRCN-based methods do not change a lot. This is because LDCF and ACF based detectors train pedestrian and cyclist detectors separately, thus they cannot differentiate them clearly. Therefore, from this point, FRCN-based framework for pedestrian and cyclist detection has obvious advantages.

When the specific post-processing (PP) for UB-MPR-FRCN method is deployed, the performance in almost all subsets can be slightly improved, which shows the post-processing step described in Section V-D can further enhance the detection performance.

E. Discussion

The above experimental results show that our proposed method UB-MPR-FRCN-PP outperforms other state-of-the-art detectors significantly. Some qualitative detection results of the proposed method under different scenarios from the new pedestrian and cyclist dataset can be found in Fig. 11. However, several important issues about the experiments need to be discussed and explained.

It is noted that the performance of all pedestrian detectors are not as good as cyclist detectors. This is because the unbalanced training samples are applied during the training procedure. Although plenty of pedestrian instances have been supplemented into the training set 3, quite a number of them are over occluded or too small, which are ignored during training.

We also find that, with the test subset setting becoming harder, average precisions of all detectors decrease gradually, because many pedestrian and cyclist instances are with lower resolution and under partial occlusion. Thus, there is still a big room to improve in the moderate and hard subsets, more work needs to be followed up in the new dataset.

We only evaluate the proposed detector in our new pedestrian and cyclist dataset, because no relevant pedestrian and cyclist dataset is available for training and evaluating the detector. The KITTI object detection dataset [20] is considered as a difficult dataset including annotated cars, pedestrians and cyclists. But very limited pedestrian and cyclist instances are involved in this dataset. Moreover the rider of cyclist instances is not annotated. Thus the proposed method cannot be validated in this dataset.

In this paper, we focuses on detection performance rather than processing speed. Our proposed method, running on a 3.3-GHz i7 Central Processing Unit (CPU) processor and a TITIAN X Graphics Processing Unit (GPU) processor, needs about 1.9s per image (2048×1024), which is almost equal to EB-FRCN ($\sim$1.8 s), but more efficient than SS-FRCN ($\sim$23 s). However, with the development of computer hardware and

Fig. 11. Detection examples of our detector under different scenarios from the pedestrian and cyclist dataset. Blue and green bounding boxes indicate detected pedestrians and cyclists, respectively.

GPU optimization, processing speed in object detection has seen great progress recently, and the complex models like convolutional neural networks will be released real-time.

VI. Conclusion and Future Work

In this paper, a unified framework for concurrent pedestrian and cyclist detection is presented, which consists of an UB-MPR based detection proposal method, a FRCN-based model for classification and localization, and a specific post-processing step. The proposed method can detect pedestrians and cyclists concurrently and differentiate them clearly, both of which are needed for the decision of intelligent vehicles.

Experimental results demonstrate that our UB-MPR proposal method outperforms the other compared methods significantly, even with less proposed bounding boxes. And our proposed method UB-MPR-FRCN-PP outperforms the others almost under any conditions. The proposed method achieves more than 10% AP improvements in the moderate subset compared to FRCN-based methods, due to the use of UB-MPR proposal. It also outperforms ACF and LDCF based detectors, especially when using the "Discard" setting, which demonstrates the benefit of the discriminative networks.

In order to make walking and cycling safer, the temporal and orientation information [37] of pedestrians and cyclists could help to improve risk assessment. Therefore, we are planning to extend our work to explore multiple object tracking and orientation estimation for pedestrians and cyclists.

References

[1] "WHO global status report on road safety," World Health Organization, Geneva, Switzerland, 2015.
[2] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1239–1258, Dec. 2010.
[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012.
[4] W. Tian and M. Lauer, "Fast cyclist detection by cascaded detector and geometric constraint," in *Proc. IEEE ITSC*, 2015, pp. 1286–1291.
[5] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE CVPR*, 2014, pp. 580–587.

[6] X. Wang, M. Yang, S. Zhu, and Y. Lin, "Regionlets for generic object detection," in *Proc. IEEE ICCV*, 2013, pp. 17–24.
[7] R. B. Girshick, "Fast R-CNN," in *Proc. IEEE ICCV*, 2015, pp. 1440–1448.
[8] X. Li *et al.*, "A new benchmark for vision-based cyclist detection," in *Proc. 4th IEEE*, 2016.
[9] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2179–2195, Dec. 2009.
[10] R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Ten years of pedestrian detection, what have we learned?" in *Proc. ECCV*, 2014, pp. 613–627.
[11] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proc. IEEE ICCV*, 2003, pp. 734–741.
[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE CVPR*, 2005, pp. 886–893.
[13] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[14] P. Dollár, Z. Tu, P. Perona, and S. Belongie, "Integral channel features," in *Proc. BMVC*, 2009, pp. 1–11.
[15] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning," in *Proc. IEEE CVPR*, 2013, pp. 3626–3633.
[16] T. Li, X. Cao, and Y. Xu, "An effective crossing cyclist detection on a moving vehicle," in *Proc. IEEE WCICA*, 2010, pp. 368–372.
[17] H. Chen *et al.*, "Integrating appearance and edge features for on-road bicycle and motorcycle detection in the nighttime," in *Proc. IEEE ITSC*, 2014, pp. 354–359.
[18] H. Cho, P. E. Rybski, and W. Zhang, "Vision-based bicyclist detection and tracking for intelligent vehicles," in *Proc. 4th IEEE*, 2010, pp. 454–461.
[19] K. Yang, C. Liu, J. Zheng, L. Christopher, and Y. Chen, "Bicyclist detection in large scale naturalistic driving video," in *Proc. IEEE ITSC*, 2014, pp. 1638–1643.
[20] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI Vision Benchmark Suite," in *Proc. IEEE CVPR*, 2012, pp. 3354–3361.
[21] L. Fu, P. Hsiao, C. Wu, Y. Chan, and S. Hu, "Vision based pedestrian and cyclist detection method," U.S. Patent 9 087 263, Jul. 21, 2015.
[22] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1915, Sep. 2015.
[23] P. Dollár, R. Appel, S. Belongie, and P. Perona, "Fast feature pyramids for object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1532–1545, Jul. 2014.
[24] A. Mogelmose, A. Prioletti, M. M. Trivedi, A. Broggi, and T. B. Moeslund, "Two-stage part-based pedestrian detection," in *Proc. IEEE ITSC*, 2012, pp. 73–77.
[25] W. Liu *et al.*, "A pedestrian-detection method based on heterogeneous features and ensemble of multi-view-pose parts," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 813–824, Apr. 2015.
[26] A. Prioletti *et al.*, "Part-based pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1346–1359, Sep. 2013.
[27] S. Zhang, C. Bauckhage, and A. B. Cremers, "Efficient pedestrian detection via rectangular features based on a statistical shape model," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 763–775, Apr. 2015.
[28] W. Nam, P. Dollár, and J. H. Han, "Local decorrelation for improved pedestrian detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 424–432.
[29] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY, USA: Wiley, 2012, pp. 161–174.
[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[31] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, no. 2, pp. 154–171, 2013.
[32] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[33] R. Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, 2008.
[34] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *Proc. IEEE IV*, 2011, pp. 691–696.
[35] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010.
[36] C. L. Zitnick and P. Dollar, "Edge boxes: Locating object proposals from edges," in *Proc. ECCV*, 2014, pp. 391–405.
[37] F. Flohr, M. Dumitru-Guzu, J. F. P. Kooij, and D. M. Gavrila, "A probabilistic framework for joint pedestrian head and body orientation estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1872–1882, Aug. 2015.

Xiaofei Li received the B.E. degree in 2011 from Tsinghua University, Beijing, China, where he is currently working toward the Ph.D. degree in the Department of Automotive Engineering.

His research interests include detection and tracking of road users based on vision, radar, and sensor fusion methods for the intelligent vehicle, particularly pedestrian and cyclist detection based on mono-camera.

Lingxi Li received the B.E. degree in automation from Tsinghua University, Beijing, China, in 2000; the M.S. degree in control theory and control engineering from Chinese Academy of Sciences, Beijing, in 2003; and the Ph.D. degree in electrical and computer engineering from University of Illinois at Urbana-Champaign, IL, USA, in 2008.

Since August 2008, he has been with Indiana University–Purdue University Indianapolis, Indianapolis, IN, USA, where he is currently an Associate Professor of electrical and computer engineering. His research interests include modeling, analysis, control, and optimization of complex systems, intelligent transportation systems and intelligent vehicles, discrete event systems, active safety systems, and human factors.

Dr. Li served as the Program Chair for the 2011 and 2013 IEEE International Conference on Vehicular Electronics and Safety and has been serving as an Associate Editor of IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS since 2009. He was a recipient of the Indiana University Trustees Teaching Award in 2012, the Outstanding Editorial Service Award for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS in 2012, and the IUPUI Prestigious External Awards Recognition (PEAR) in 2013.

Fabian Flohr received the M.Sc. degree in computer science from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2012. He is currently working toward the Ph.D. degree at TU Delft, Delft, The Netherlands.

He is also currently with Daimler Research and Development, Ulm, Germany. His research interests include machine learning and video analysis for intelligent vehicles, with a focus on pedestrian detection, segmentation, and tracking.

Jianqiang Wang received the B.Tech., M.S., and Ph.D. degrees from Jilin University of Technology, Jilin, China, in 1994, 1997, and 2002, respectively.

He is an Associate Professor with the Department of Automotive Engineering, Tsinghua University, Beijing, China. He has authored more than 40 journal papers. He is the holder of 20 patent applications. He has engaged in more than ten sponsored projects. His research interests include intelligent vehicles, driving assistance systems, and driver behavior.

Hui Xiong received the B.E. degree from Fujian University of Technology, Fujian, China, in 2012. He is working toward the M.E. degree in the School of Software, Beihang University, Beijing, China.

He is also with the Department of Automotive Engineering, Tsinghua University, Beijing. His research interests include object detection and tracking for onboard visual technology, and data mining.

Morys Bernhard received the Ph.D. degree from University of Karlsruhe, Karlsruhe, Germany, in 1998, with a focus on simulation of dynamical behavior of trains and wear of wheels.

From 1998 until 2014, he was with Daimler AG, Germany, as a Product Manager for Strategy Mercedes-Benz Cars, Innovation Management Mercedes-Benz Cars, and the Manager of Operations Daimler Driving Simulator. Since 2014, he has been the Head of the Driver Assistance and Chassis Systems, Daimler Greater China Ltd, Beijing, China.

Shuyue Pan received the B.E. and M.E degrees from Dalian Technology and University, Dalian, China, in 2007 and 2009, respectively, and the Ph.D. degree from Technology University of Braunschweig, Braunschweig, Germany, in 2014, with a focus on the control algorithm for the adaptive cruise control system.

Since 2015, she has been a Senior Engineer with the Driver Assistance and Chassis Systems, Daimler Greater China Ltd., Beijing, China.

Dariu M. Gavrila received the Ph.D. degree in computer science from University of Maryland, College Park, MD, USA, in 1996.

From 1997 until 2016, he was with Daimler Research and Development, Ulm, Germany, where he became a Distinguished Scientist. He has led the multiyear pedestrian detection research effort at Daimler, which was incorporated in the Mercedes-Benz S-, E-, and C-Class models (2013–2014). In 2003, he also became a part-time Professor with University of Amsterdam, Amsterdam, The Netherlands, in the area of intelligent perception systems. Recently in 2016, he became a Full Professor of intelligent vehicles at TU Delft, Delft, The Netherlands. Over the past 20 years, he has focused on visual systems for detecting human presence and activity, with application to intelligent vehicles, smart surveillance, and social robotics.

Prof. Gavrila was a recipient of the I/O Award 2007 from the Dutch Science Foundation (NWO) and the Outstanding Application Award 2014 from the IEEE Intelligent Transportation Systems Society.

Keqiang Li received the B.Tech. degree from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees from Chongqing University, Chongqing, China, in 1988 and 1995, respectively.

He is a Professor with the Department of Automotive Engineering, Tsinghua University. He has authored more than 170 papers. He holds 50 patents in China and Japan. His research interests include vehicle dynamics and control for driving assistance systems, as well as hybrid electrical vehicles.

Dr. Li has served as a Senior Member of the Society of Automotive Engineers of China and on the Editorial Boards of *International Journal of Intelligent Transportation Systems Research* and *International Journal of Vehicle Autonomous Systems*. He was a recipient of the "Changjiang Scholar Program Professor" and some awards from public agencies and academic institutions of China.