Detection of pedestrian actions based on deep learning approach

Danut Ovidiu Pop

DETECTION OF PEDESTRIAN ACTIONS BASED ON DEEP LEARNING APPROACH

DĂNUŢ OVIDIU POP[1,2,3]

Abstract. The pedestrian detection has attracted considerable attention from research due to its vast applicability in the field of autonomous vehicles. In the last decade, various investigations were made to find an optimal solution to detect the pedestrians, but less of them were focused on detecting and recognition the pedestrian's action. In this paper, we converge on both issues: pedestrian detection and pedestrian action recognize at the current detection time (T=0) based on the JAAD dataset, employing deep learning approaches. We propose a pedestrian detection component based on Faster R-CNN able to detect the pedestrian and also recognize if the pedestrian is crossing the street in the detecting time. The method is in contrast with the commonly pedestrian detection systems, which only discriminate between pedestrians and non-pedestrians among other road users.

1. INTRODUCTION

Pedestrian detection is a significant problem for computer vision, which involves several applications, including robotics, surveillance, and the automotive industry. It is one of the main interests of transport safety since it implies reducing the number of traffic collisions and the protection of pedestrians (i.e., children and seniors), who are the most vulnerable road users.

There are almost 1.3 million persons die in road traffic collisions each year, and nearly 20-50 million are injured or disabled due to human errors inherited in the usual road traffic. Moreover, the clashes between cars and pedestrians are the leading cause of death among young people, and it could be effectively reduced if such human errors were eliminated by employing an Advanced Driver Assistance System (ADAS) for pedestrian detection.

Over the last decade, the scientific community and the automobile industry have contributed to the development of different types of ADAS systems in order to improve traffic safety. The Nissan company has developed a system

Key words and phrases. Pedestrian Detection, Pedestrian Action Recognition, Deep Learning.

which recognizes the vehicle's environment, including pedestrians, other vehicles, and the road. Lexus RX 2017 has a self-driving system which is linked up to a pedestrian detection system. More recently, the Audi ADAS system accumulates the data of the camera and/or radar sensor to determine the possibility of a collision by detecting pedestrians or cyclists and warns the driver with visual, acoustic and haptic alerts if a crash is coming.

These current ADAS systems still have difficulty distinguishing between human beings and nearby objects. In recent research investigations, deep learning neural networks have frequently improved detection performance. The impediment for those patterns is that they require a large amount of annotated data.

This paper proposes a pedestrian detection system which not only discriminates the pedestrians among other road users but also able to recognize the pedestrian actions at the current detection time (T=0).

The contribution of this paper concerns detecting and classifying pedestrian actions. To do so, we develop the following methodology relying on a deep learning approach:

- Train, all pedestrian samples with the Faster R-CNN-Inception version 2 for pedestrian detection, proposes [1];
- Train all pedestrian samples also using the pedestrian actions tags (cross/not cross) with the CNN as mentioned above for detection and action recognition based on the Joint Attention for Autonomous Driving (JAAD) [2] dataset;

The paper is organized as follows: Section 2 outlines sever existing approaches from the literature and supplies our main contribution. Section 3 presents an overview of our system. Section 4 describes the experiments and the results on the JAAD dataset. Finally, Section 5 presents our conclusions.

2. RELATED WORK

A wide variety of methodologies have been proposed with optimization in performance, resulting in the development of detection methods using deep learning approaches [3, 4, 5] or combination of features followed by a trainable classifier [4, 6].

A deep, unified pattern that conjointly learns feature extraction, deformation handling, occlusion handling, and classification evaluated on the Caltech and the ETH datasets for pedestrian detection was proposed in [7]. An investigation focused on the detection of small scale pedestrians on the Caltech data set connected with a CNN learning of features with an end-to-end approach was presented in [8].

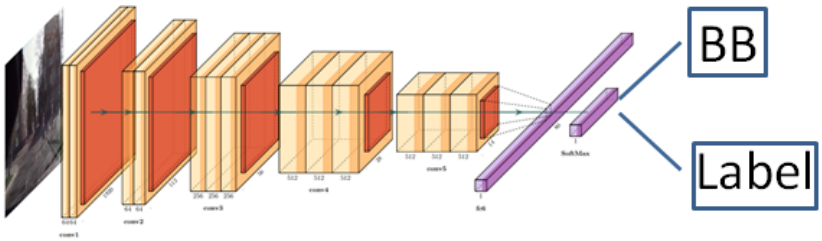

FIGURE 1. Pedestrian detection system architecture. Label represents the pedestrian tags which could be pedestrian, pedestrian crossing; BB cord represents the bounding box coordinates.

A combination of three CNNs to detect pedestrians at various scales was introduced on the same monocular vision data set [9]. A cascade Aggregated Channel Features detector is used in [10] to generate candidate pedestrian windows followed by a CNN-based classifier for verification purposes on monocular Caltech and stereo ETH data sets.

In [11] is presented a pedestrian detection based on a variation of YOLO network model, (three layers were added to the original one) in order to join the shallow layer pedestrian features to the deep layer pedestrian features and connect the high, and low-resolution pedestrian features.

A mixture of CNN-based pedestrian detection, tracking, and pose estimation to predict the crossing pedestrian actions based on JAAD dataset is addressed in [12]. The authors utilize the Faster R-CNN object detector based on VGG16 CNN architecture for the classification task, use a multi-object tracking algorithm based on Kalman filter, apply the pose estimation pattern on the bonding box predicted by the tracking system and finally handle the SVM/Random Forest to classify the pedestrian actions (Crossing /Not Crossing).

This approaches mentioned above use standard pedestrian detection pattern which only discriminates the pedestrian and non-pedestrian among other road users. To our knowledge, there is no prior research linked to pedestrian detection, which involves various pedestrian behavior tags.

We investigate the pedestrian detection issue in two ways, using the pedestrian and non-pedestrian tags, we call the Classical Pedestrian Detection Method (CPDM), and various pedestrian tags the: pedestrian, pedestrian crossing which we call Crossing Pedestrian Detection Method (CrPDM).

3. PEDESTRIAN DETECTION METHOD

This paper concerns solve the problem of pedestrian detection and pedestrian recognition actions using deep learning approach.

For developing a pedestrian detection system is mandatory to take into account three main components: the sensors employed to capture the visual data, the modality image processing elements, and the classification parts. In general, all these components are synchronically developed to achieve a high detection performance, but seldom specific item could be investigated independently according to the target application. We concurrently exam in the detection part by applying a generic object detector based on the public Faster R-CNN [13]. We handle the Inception CNN architecture versions 2 for the classification task with the TensorFlow public open-source implementation described in [1]. All the training process is based on JAAD [2] dataset. This dataset has different pedestrian tags. It has an annotation of pedestrians with behavioral tags and pedestrians without behavior tags.

4. EXPERIMENTS

This section presents the set of experiments, including setups and performance assessment of our approaches.

4.1. **Data Setup.** The experiments are completed on the JAAD dataset [2] because of its data set in typical urban road traffic environments for various locations, times of the day, road and weather conditions. This dataset supplies pedestrian bounding boxes (BB) for pedestrian detection, tagged as non-occluded, partially occluded and heavily occluded BBs. Moreover, it includes the pedestrian actions annotations for several of them, the pedestrian attributes for estimating the pedestrian behavior and traffic scene elements.

The experiment employs all the pedestrian samples, including the partially and heavily occluded pedestrians for all training and testing process.

4.2. **Training, Testing and Evaluation Protocol.** We used the first 70% of samples from the whole JAAD dataset for the learning process. The training set consists of first 190 video sequence training samples, and includes even the partially occluded and heavily occluded BBs, in contrast with [2] where the authors used just a part of the dataset, omitting samples with partially and heavy occlusion. Moreover, the authors do not give a detailed explanation of how the datasets were divided into training and testing sets; hence, it does not allow a fair comparison.

The validation set represents 10% of the learning set. We used the holdout validation method, which held back from training the model. The evaluation of a model skill on the training dataset would result in a biased score. Therefore

FIGURE 2. Pedestrian detection using the pedestrian tag for all pedestrians.

FIGURE 3. Pedestrian detection using multiple tags.

the model is evaluated on the holdout sample to give an unbiased estimate of model skill.

The JAAD dataset has three main pedestrian actions tags:

TABLE 1. Our detection performances using one or multiple output labels. One label represents that all samples are tagged as a pedestrian. Multiple labels represent: P=Pedestrian, PCr=Pedestrian Crossing.

Method	Train on	Output	mAP% ± CI
Faster R-CNN Inception v2	All pedestrian samples tagged as P	Pedestrian BB Label	70.91 ± 1.61
Faster R-CNN Inception v2	All Pedestrian with P and PCr Tags	Pedestrian BB+ Action Label	64.31 ± 1.70
SSD Fusion Inception [14]	RGB, Lidar, Distance	Pedestrian BB Label	51.88

- pedestrian completes crossing the street;
- pedestrian does not cross the street;
- pedestrian does not have any intention of crossing.

The first pedestrian action we tag as pedestrian cross and others as pedestrian considering his/her intention is ambiguous or does not cross the street.

We adopt two approaches for the training stage (the main architecture is described in Fig 1):

- using the Classical Pedestrian Detection Method (CPDM) where we consider all pedestrian samples without any specific tag (see Fig 2);
- using the Crossing Pedestrian Detection Method (CrPDM) where we use various pedestrian tags: Pedestrian is crossing the street (PCr), and Pedestrian (P) for all other pedestrians who do not cross the street or their action is obscure (see Fig 3).

We perform the CNN training process on 200000 iterations, using an initial learning rate value to 0.00063 with ADAM learning algorithm and momentum at 0.9 [2]. We used the pre-trained weights from COCO dataset with the default Faster RCNN loss function which is optimized for a multi-task loss function.

The testing set used to assess the CNN model performance is independent of the training dataset. It contains 110 video samples, the last 30% of video samples from the whole JAAD dataset.

The evaluation process for all the CNN models is performed with Tensorflow Deep Neural Network Framework. The performances are assessed by the mean average precision (mAP) for the detection part using the TensorFlow metrics tool.

We calculate the Confidence Interval (CI) to evaluate whether one model is statistically better than another one.

$$CI = 2 * 1.96\sqrt{\frac{P(100-P)}{N}}\%. \tag{1}$$

In this formulation, P represents the performance system (e.g., mAP) and N represents the number of video testing.

4.3. **Results and Discussions.** The detection results are presented in Table 1. We observed that the detection performance achieved with the CPDM method (using all samples as pedestrians) a good performs on Jaad dataset since it has to identify the pedestrian among other road users. The CrPDM approach (using multiple pedestrian tags), although it detects the pedestrians, cannot be associated with the first method because it also instantly classifier the action of the pedestrians during the detection step. Therefore its performance is less than the first detection approach. On the other hand, pedestrian detection using the multiple tags approach could be a start point for a deep investigation. This approach estimates the pedestrian actions on the current time (T=0) and could be beneficial for developing a pedestrian prediction system. We can not compare our detection models with JAAD approaches [2] as our results are not directly comparable since the authors made a classification for a specific pedestrian action based on pedestrian attention information and only used the non-occluded pedestrian samples [2]. Their approach is based on a variation of AlexNet-imagenet CNN where the input data are cropped beforehand.

Contrariwise we should evaluate our model using approximately the same non-occluded pedestrian training samples as in [2] for a fair comparison, but we did not do that in this previous work since we assume that in the all traffic congestion even exist pedestrians who are partially and heavily occluded.

Nevertheless, we compare our detection models with another method [14] which is close to our first one. This approach is based on a variation of SSD-Inception CNN based on SvDPed dataset. The method merged the RGB images, low-resolution Lidar and the distance between the camera and object detected. Notwithstanding the [14] approach used many sensors for pedestrian detection, our methods outperforms this approach significantly (please see Table 1).

5. CONCLUSIONS

This paper presents a classical pedestrian detection system and a pedestrian detection system able to recognize even the pedestrian actions based on deep learning approaches using JAAD dataset.

We evaluated the pedestrian detection approach (we called Classical Pedestrian Detection Method (CPDM)), where all sample are tagged as pedestrian and not pedestrian and a pedestrian detection approach using multiple tags (pedestrian, pedestrian cross), which we call Crossing Pedestrian Detection Method (CrPDM). The first method achieved better performance since it has only to distinguish the pedestrians among other road users in contrast with the second one who has to recognize even the pedestrian actions. The second detection approach returned a weaker performance than the classical one. Contrariwise, we deem this approach could be a start point for a deep investigation. The experiments were carried out on the common object detector based on the public Faster R-CNN merged with Inception version 2 architecture for the classification part.

Future work will be concerned with improving and benchmarking of pedestrian detection using multiple pedestrian action tags and extending the method to a pedestrian prediction system.

References

[1] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors. *CoRR*, abs/1611.10012, 2016.

[2] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior. In *The IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017.

[3] Dănuţ Ovidiu Pop, Alexandrina Rogozan, Fawzi Nashashibi, and Abdelaziz Bensrhair. Incremental cross-modality deep learning for pedestrian recognition. *In 28th IEEE Intelligent Vehicles Symposium (IV)*, pages 523–528, June 2017.

[4] Shanshan Zhang, Rodrigo Benenson, Mohamed Omran, Jan Hendrik Hosang, and Bernt Schiele. How far are we from solving pedestrian detection? *CoRR*, abs/1602.01237, 2016.

[5] J. Schlosser, C. K. Chow, and Z. Kira. Fusing lidar and images for pedestrian detection using convolutional neural networks. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2198–2205, May 2016.

[6] Rodrigo Benenson, Mohamed Omran, Jan Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? In Lourdes Agapito, Michael M. Bronstein, and Carsten Rother, editors, *Computer Vision - ECCV 2014 Workshops*, pages 613–627, Cham, 2015. Springer International Publishing.

[7] W. Ouyang, H. Zhou, H. Li, Q. Li, J. Yan, and X. Wang. Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2017.

[8] R. Bunel, F. Davoine, and Philippe Xu. Detection of pedestrians at far distance. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2326–2331, May 2016.

[9] M. Eisenbach, D. Seichter, T. Wengefeld, and H. M. Gross. Cooperative multi-scale convolutional neural networks for person detection. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 267–276, July 2016.

[10] Xiaogang Chen, Pengxu Wei, Wei Ke, Qixiang Ye, and Jianbin Jiao. *Pedestrian Detection with Deep Convolutional Neural Network*, pages 354–365. Springer International Publishing, Cham, 2015.

[11] W. Lan, J. Dang, Y. Wang, and S. Wang. Pedestrian detection based on yolo network model. In *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1547–1551, Aug 2018.

[12] Z. Fang and A. M. Lpez. Is the pedestrian going to cross? answering by 2d pose estimation. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1271–1276, June 2018.

[13] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[14] T. Kim, M. Motro, P. Lavieri, S. S. Oza, J. Ghosh, and C. Bhat. Pedestrian detection with simplified depth prediction. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2712–2717, Nov 2018.

(1) Babeş-Bolyai University, Department of Computer Science, 1 M. Kogălniceanu Street, 400084 Cluj-Napoca, Romania

(2) INRIA Paris, RITS team, 2 Rue Simone IFF, 75012 Paris, France

(3) INSA Rouen, LITIS, 685 Avenue de l'Universit, 76800 Saint-tienne-du-Rouvray, France

E-mail address: danutpop@cs.ubbcluj.ro