School of Informatics

Informatics Research Review
Deep Learning Based Pedestrian Detection for Autonomous Driving

[redacted]

January 2021

Abstract

Autonomous driving is an important application of pedestrian detection, which can be used to avoid traffic accidents and protect the safety of road users. The challenges of pedestrian detection for automated driving lie in occlusion, distant objects, posture variants, and different environments. Traditional vision based and machine learning methods have achieved high detection accuracy, but some problems remain. The development of deep learning recently sees further improvements on pedestrian detection with less false and miss detection. This literature review lists common deep learning approaches for pedestrian detection which mitigates the limitations of automated driving.

Date: Friday 22$^{\text{nd}}$ January, 2021

Supervisor: [redacted]

1 Introduction

Every year, more than 1.2 million people die in traffic crashes. In 2015, more than 5,000 pedestrians died in traffic accidents in the United States [1]: one pedestrian died every 1.6 hours due to traffic crashes. In addition, pedestrians are 1.5 times more likely to die in car accidents than passengers in vehicles [1]. The automotive industry generally believes that vehicles must avoid accidents instead of trying to survive them to improve safety since the advantages of passive safety systems such as seat belts and airbags have reached a stable level [2]. In addition to the advantage of safety, autonomous vehicles provide benefits in society and economy. The large amount of time spent in commuting can be used to do other more important things instead. Autonomous vehicles are especially helpful for the aged and the disabled to interact with family, friends and the community [2]. Self-driving cars can improve fuel efficiency by adjusting the acceleration and deceleration curves, thus reducing the impact on the environment [2]. For these reasons, autonomous driving has become an important application in pedestrian detection.

Pedestrian detection is a rapidly developing area in computer vision, which can replace human eyes to detect human-like objects from images or videos. This technology is rather essential in applications such as autonomous driving, video surveillance, and robotics. Past pedestrian detection approaches are mostly vision-based and features-based [3]. The procedure of pedestrian detection based on vision can be divided into three primary steps: (i) Image Acquisition, (ii) Feature Extraction and (iii) Classification [3]. Recent work in this area is mainly based on monocular detectors with large-scale annotations, which develop significantly with deep learning techniques. However, monocular detectors have limitations in crowded scenes, resulting in severe occlusions and information loss of pedestrians. Accordingly, pedestrian detection is emerging in multi-camera views with overlapping fields of view to alleviate the problem of occlusion in single camera view. The deep learning architectures differ from traditional pedestrian detection methods by removing the extraction of features [3]. More specifically, deep classifier automatically extracts and learns features of input images through several processing layers, which obtains accurate classifications and reduces the running time.

False and miss detection are urgent problems to be solved, especially in autonomous driving, which can lead to serious accidents in the traffic system. When the detector recognizes that something else is a pedestrian, false detection occurs. Miss detection refers to the loss of pedestrians in crowded scenes due to severe occlusions, especially in monocular images. This literature review aims to evaluate how deep learning methods can improve the accuracy of pedestrian detection, thus addressing the above problems in autonomous driving. In addition, how to improve the performance of pedestrian detection under specific conditions using deep learning is also mentioned.

This IRR will focus on the most widely used deep learning method Convolutional Neural Network (CNN) for pedestrian detection. Although automatic driving is applied outdoors, the pedestrian detection approaches mentioned in this IRR are generally applicable to both indoor and outdoor environments. However, this review will not discuss pedestrian detection in 3D vision systems which is not common in most research. Most of the reviewed papers in this literature reviews are published in authoritative journals and cited by other related papers.

2 Literature Review

The organization of the literature review is as follows. First, the process of pedestrian detection is shown in several important stages. Annotated pedestrian datasets are introduced togeth-

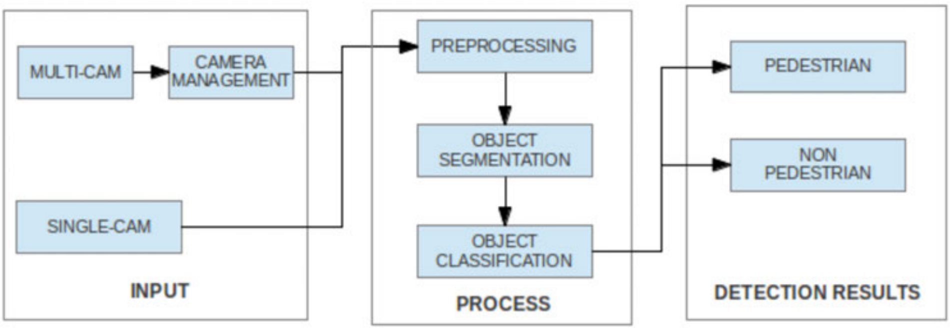

Figure 1: Pedestrian detection process.

er with image acquisition equipment. The fundamental detection procedure is also described. Second, some common challenges for pedestrian detection in autonomous driving, including occlusion and distant instances, are expounded, and the demand of deep learning architecture is also explained. Then, the most common deep learning method CNN is introduced. CNN can be divided into two architectures, and a typical method of each architecture is reviewed, respectively. The structures and principles of these methods are described, and their performance is evaluated with experimental data. Finally, the application of these deep learning approaches in some specific light or weather conditions is mentioned.

2.1 Pedestrian Detection

In [4], pedestrian detection process is divided into three major stages: input, process, and output. The whole process is shown in Figure 1 [4]. In the input process, the input data and input devices can influence the selection of detection methods. Once the received data is preprocessed, specific techniques and algorithms are applied to image segmentation and classification. The final task is to judge whether an object is a pedestrian or not.

A variety of datasets are available for training models and testing the performance of some pedestrian detection approaches. These datasets are different in resolutions, the number of camera views, state (mobile or static), overlapping or non-overlapping, camera height, and environment (indoor or outdoor), and so on. Table 1 lists some commonly used pedestrian datasets, most of which are captured from moving car. INRIA [5], ETH [6], and TUD-Brussels [7] are widely used in early research for pedestrian detection, containing thousands of images captured in daytime. The Daimler dataset [8] is larger in scale and can be even expanded using stereo image pairs. However, images in Daimler set are grey scale, which is likely to degrade the evaluation. Caltech has been widely used since it was publised in 2009 [9]. Pedestrians are assigned using three labels in this dataset: 'Person' represents individual pedestrian; 'People' for large crowds of pedestrians; and 'Person?' refers to ambiguous pedestrians who are easy to be labeled incorrectly. The KITTI dataset [10] is collected based on an autonomous driving platform in rural areas and on highways. It is suitable for multi-class detection and traffic scenes.

Input devices also play an important role in the input stage [4]. Single laser scanner sensors and multiple laser sensors are commonly used in many researches. At the same time, research can also be based on the far infrared sensor which can detect low-resolution objects with blurred colour and texture [11]. For poor light environments, stereo cameras with night vision functions are usually used [12]. Pan-tilt-zoom (PTZ) camera is usually applied in object detection with

Dataset	Camera numbers	Resolution	Camera height/position	Scene
INRIA	n/a	high	n/a	n/a
ETH	1	640×480	approx. 90 cm above ground	outdoor
TUD-Brussels	1	640×480	from a driving car	inner city
Daimler	1	640×480	from a moving vehicle	canal area of city
Caltech	1	640×480	from a vehicle	urban environment
KITTI	4	1240×376	by driving	rural areas/highways

Table 1: Commonly used pedestrian datasets.

changing position, direction and size [13]. One of the most commonly used devices in pedestrian detection research is the video camera due to its low price and availability [4].

In the detection process, after preprocessing the received data from input devices, extracting region of interest (RoI) and segmenting objects are two essential steps. Moreover, object classification also plays an important role in pedestrian detection. The commonly used methods in this step are Support Vector Machine (SVM) and neural networks. The preprocessing of the video data from cameras is used for normalization and calibration. The detection process can be divided into offline and real-time detection processes. Offline detection manually processes the input data which is a video or a sequence of frames, while in real-time detection, input data is directly and in real-time captured from cameras, CCTV, or other sensors. Real-time detection is more challenging because it requires fast and accurate detection.

After pre-processing, the most important step in pedestrian detection is object segmentation, which is associated with the segmentation of RoI (Region of Interest) from background. Background subtraction is the simplest and fastest segmentation method. However, it is not suitable for the unpredictable dynamic environment where the background can suddenly change. This problem can be addressed by adaptive background and environmental conditions. In addition to the separation of background, object segmentation and classification can also be implemented through extraction of features in images such as Histograms of Oriented Gradient (HOG) and optical flow. HOG-based methods are usually used because they are fast and can provide accurate detection results. Other features include the trajectory of the object, the shape of the object, edge, and ABM (Active Base Model) feature. Next, classification algorithms will classify objects into different categories, which is a very important stage in the detection process.

According to the method results summarized in [4], SVM and its derivatives are most commonly used traditional pedestrian detection methods, while most recent detection methods are based on deep learning, because neural networks can automatically learn features and have more powerful generalization ability to overcome various challenges in pedestrian detection.

Finally, the detection results are obtained after the detection process, and can be evaluated according to the pedestrian annotations.

2.2 Challenges of Automated Driving

The enhancement of advanced auxiliary systems for drivers and self-driving vehicles requires significant improvements in pedestrian detection, especially in complicated and crowded scenes. The perception of the environment by Radar, Lidar sensors, or cameras plays an important

Figure 2: Typical challenges to pedestrian detection for automated driving.

role in autonomous driving which aims to extract the environmental information [14]. Highly accurate sensor systems reliably and quickly warn the driver of vulnerable road users, automatically performs emergency steering to prevent traffic crashes, and plans action and driving strategies [14]. As both false and miss detections can cause harmful effects, machine detection performance is expected to reach or even exceed human performance [14]. Challenges of high performance of detection for automated vehicles include accurate object localization, distant objects with low resolution, sudden appearance of pedestrians, corner situations with few or no training examples, and most importantly, occlusion. Figure 2 [14] illustrates some possible and typical challenges of pedestrian detection in urban roads. In the scene, it is noticeable that there are pedestrians occluded by each other in a crowd, pedestrians hidden behind cars who are likely to suddenly appear. Accordingly, reliable pedestrian detection for autonomous driving has the following requirements [14]:

1. Reliable pedestrian detection can function at various lighting conditions, especially at night with low illumination.

2. Reliable detection can recognize both nearby and distant pedestrians without any information loss.

3. Robust detection can deal with different pedestrian sizes, postures and appearances.

4. Robust detection can address the problems in complex traffic scenes and different weather conditions.

5. Reliable detection can deal with the problem of occlusion, that is, pedestrians are blocked by vehicles or other road users.

Traditional machine learning strategies, such as HOG+SVM, have made some achievements in high performance of pedestrian detection, but there are still some problems under challenging environment and weather conditions. Deep learning helps improve the machine detection and classification performance in autonomous driving to guarantee safety in any situations.

2.3 Convolutional Neural Network

Deep learning strategies have been developed in pedestrian detection, which can be used to automatically segmentate, detect, and classify objects in images and videos. The most well-known deep learning architectures are Convolutional Neural Networks (CNN) which can be used to automatically extract and learn features of images through several hidden layers with nonlinearity and classify images into different classes. Hand-crafted features include low-level features like corners and gradients, medium-level features are some basic shapes such as lines, curves, and circles, and high-level features which represent parts of objects [14].

In 2005, Szarvas et al. [15] used CNN to detect pedestrians on a dataset with no restrictions on posture, movement, background and lighting conditions. The first layers of CNN acts as a feature extractor with shared weights. These weight parameters are optimized by gradient descent to minimize the generalization error. Each node in these feature extraction layers (feature maps) is connected to a node at another region of the previous layer. The last few layers in the network are fully connected, and classification is performed based on features extracted in shallow layers. The authors compared their approach with the classical SVM approach. The false positive rate (FPR) of the proposed method is less than 1/5 of that of the Support Vector Machine (SVM) classifier. The detection rate of the CNN approach reached approximately 90%. Furthermore, the computational cost of the CNN is about 40 times less than that of the SVM. CNN has made remarkable progress in accuracy and computation.

Convolutional neural networks can generally be divided into two structures: single-stage and two-stage approaches [14]. After feature extraction, the single-stage framework can directly perform multi-class classification and localization. Two-stage frameworks rely on two networks a region proposal network (RPN) for localization and binary classification, and the other classification network for multi-class classification. Single-stage frameworks include YOLO network [16], and two-stage frameworks include network models such as Fast R-CNN [17] and Faster R-CNN [18]. The Region CNN (R-CNN) architectures generate object region proposals and extract features from region candidates. Single-stage frameworks are faster in computation, while two-stage frameworks are more accurate in detection. The following two subsections focus on the rationality of two recently used deep learning methods for pedestrian detection, one for each type of framework.

2.4 Faster R-CNN

Faster R-CNN relies on two parts: a Region Proposal Network (RPN) for predicting pedestrian candidates, and a classification network containing a set of fully convolutional layers. RPN is responsible to identify object proposals in images and feeds those potential candidates with high probability to the classification network. Faster R-CNN [18] has proved to be a reliable method for general object detection without using hand-crafted features. As pedestrian detection is considered to be a special domain beyond general object detection, Zhang et al. [19] justified whether Faster R-CNN was suitable for detecting pedestrians as an extension of Fast R-CNN [17]. Surprisingly, the combination of RPN and Fast R-CNN classifier as a pedestrian detector reduces the accuracy in comparison to the RPN as a stand-alone pedestrian detector. There are two possible causes for this issue.

First of all, pedestrian detection for automatic driving or video surveillance usually makes the pedestrian volume smaller visually. For small objects, the Fast R-CNN classifier is degraded because the resulting features from low-resolution feature map cannot be discriminated in small regions. In [19], this problem is addressed by the hole-algorithm for increasing the size of feature map and by the sampling features from shallow layers but with high-resolution. Second, the false prediction in pedestrian detection is usually caused by the confusion of hard background instances. Zhang et al. [19] applied cascaded Boosted Forest (BF) [20] for mitigating hard negative examples to classify RPN candidates.

To summarize, the improved system for pedestrian detection using Faster R-CNN consists of two cascaded stages: the first is to predict potential image regions containing a person; the second one is the Boosted Forest classifier for classification. The region proposals generated by RPN are forwarded to train the Boosted Forest classifier. To address the hard negatives, the training set contains both positive examples and randomly selected negative examples. Furthermore, additional negative examples are added to the training set after each bootstrapping stage.

The Miss Rate (MR) of the proposed method was 9.6% on Caltech dataset [9]. On the INRIA dataset [5], the approach reached an MR of 6.9%, which exceeded the leading competitors 11.2% in all experiments. The miss rate on ETH dataset [6] achieved 30.2%, which was better than previous optimal method by about 5%. The performances on KITTI [10] were evaluated using Mean Average Precision (MAP) at three different levels. The improved approach has faster speed and higher accuracy. The Log Average Miss Rate on False Positive Per Image (FPPI) of the developed Faster R-CNN method is also superior to the combination of RPN and Fast R-CNN.

This paper justified the effectiveness of Faster R-CNN for detecting pedestrians with sufficient evidence and analysis. The improvements proposed in this paper were based on the limitations of previous pedestrian detection methods, which indicates that developments of deep learning approaches for specific tasks is also necessary. The paper compared performances internally and externally, however, the evaluation metric was not consistent for all the experiments without any explanation. Additionally, different experimental results were illustrated on one line chart and represented by curves of different colours. As the results of previous detection approaches were mostly similar, the corresponding curves were easy to overlap and difficult to distinguish.

2.5 YOLO Networks

In contrast to Faster R-CNN, YOLO network model [16] designs object detection as a regression problem, which is beneficial for real-time object detection. This model can directly predict the object category and the corresponding probabilities of bounding boxes from the input image. It can also can predict whether an object exists or not and its position from the entire image at once, so it is called *You Only Look Once*. The pedestrian detection procedure using YOLO network is as follows [16]:

1) The input image is first divided into $S \times S$ cells. If one cell contains the centre point of an object, then that cell is used to detect the object in it. Each cell predicts B possible bounding boxes and corresponding probability.

2) There are five parameters for the prediction of each bounding box: (x, y, w, h, c). The (x, y) is the centre point of the bounding box relative to the cell boundary, w and h are the width and height relative to the entire image, and c represents the confidence of the bounding box.

3) The confidence of each bounding box consists of two aspects: the probability that the grid contains the object, and the accuracy of the detection box. The probability is noted as $Pr(object)$, which indicates whether the bounding box is the background ($Pr(object) = 0$) or it contains the target ($Pr(object) = 1$). The accuracy can be evaluated by IOU (intersection over union) between the ground truth box and the predicted box, which is denoted as iou_{pred}^{truth}. Therefore, the confidence can be expressed as $Pr(object) \times iou_{pred}^{truth}$.

4) For classification, each grid predicts C conditional probabilities $Pr(class|object)$, given that the target grid cell contains an object. The conditional probability is multiplied

by the predicted confidence value to obtain the class-specific confidence scores of each individual box during detection.

There will be some information loss through deep networks, which leads to the vanishing gradients and inaccurate pedestrian detection, Lan et al. [21] proposed an improved YOLO network architecture YOLO-R with three added Passthrough layers. These layers connect features in shallow layers to deep layers and associate low-level features with high-level features, which enables the network to obtain fine-grained information in shallow layers. Meanwhile, the enhanced YOLO network improves the ability of shallow feature extraction ability by changing the Passthrough layer. The performance of YOLO networks was evaluated on INRIA dataset [5]. Experimental results show that the improved YOLO-R network outperforms the original YOLO network in IOU, precision, and recall, indicating that this approach can reduce the false detection rate and missed detection rate, and improve the accuracy of pedestrian detection. The missing detection rate of YOLO-R was 10.05% lower than that of the original YOLO model (11.29%).

In general, this paper was well-organized and concise. The authors explained the basic principle of how YOLO model classifies and detects objects, which benefits readers who have deep learning background knowledge but are not familiar with this architecture. In addition, they provided detailed network structures of the proposed YOLO-R model by displaying each layer of the network on a table. Although a variety of evaluation metrics were used to evaluate the performances of their method, only one dataset was applied. Therefore, the generalization performance of the YOLO-R was not evaluated by experiments on different pedestrian datasets. Moreover, the experiments on other deep learning pedestrian detection methods were not conducted as well for comparison. As detecting pedestrians is different from detecting general objects, YOLO neural network should be justified as an effective method particularly for pedestrian detection by comparing with other pedestrian detectors. They did not explain the validity of their proposals in theory, only mentioning the experimental results.

2.6 Special Conditions

Section 2.3-2.5 introduces some widely used deep learning architectures for detecting pedestrians. Using the above mentioned deep learning methods, we can further deal with some challenges that affect the effectiveness of pedestrian detection in automated driving under specific circumstances.

Detecting pedestrians at nighttime is much more challenging because images tend to get blurry and noisy with luminance variation. Moreover, occlusion can also occur due to the low contrast to distinguish pedestrians from background, which makes it more difficult to detect pedestrians in low-illumination environment. Despite the fact that it is more difficult to detect pedestrians at night than during the day, Faster R-CNN has a relatively high performance on night dataset NightOwls with a miss rate of 19% [22].

The infrared thermal camera is also considered as an effective tool for nighttime pedestrian detection using a framework with Faster R-CNN and a Region Decomposition Branch [23]. Unlike the general RPN which proposes the whole body of a person, the region decomposition branch decomposes the region proposal into three most obvious parts in a thermal image for a person: head, foot, and trunk. Therefore, the developed method generates the association between the trunk and other body parts (such as the head and leg), which extracts both global and local features. The experiments show that this thermal-based approach benefits from the combination of multi-region features, and improves the detection accuracy, which is more effective than current night detection strategies and solves the occlusion problem. However, the conclusion from this paper is only partially reliable because the authors did not identify the dataset used in the experiment clearly. Therefore, further demonstration of their work can focus on the evaluation with more common datasets as well as specific dataset designed for night detection like NightOwls.

Similarly, Kumar et al. [24] also applied a thermal based model at night conditions. Additionally, to detect under various illumination conditions, the authors proposed a brightness aware model for multispectral pedestrian detection and used a colour model at daytime. In the above paper, pedestrian detection is divided into in two parts. First, in the brightness aware model, day or night conditions are identified using either deep learning or image processing techniques. The deep learning model is trained by both day and night datasets which totally contains 1722 images and 17 various scenarios using Mobilenet architecture. As for the image processing method, the day/night conditions are detected based on the contrast and brightness of images. The brightness of the image is determined by the average value of all the pixel values, and the contrast is computed by the Root Mean Square method. Both brightness and contrast values are compared with corresponding thresholds to predict whether an image is in daytime or nighttime. Second, the colour model used in daytime learns pedestrian features by PACAL VOC dataset, and the thermal model for low or no light conditions is trained through the thermal dataset FLIR-ADAS. Two evaluation metrics Mean Average Precision (MAP) and Intersection over Union (IoU) are used to evaluate the accuracy of the proposed model. This model has achieved a MAP of 81.27%, which is superior to most current approaches for nighttime pedestrian detection.

Moreover, most traffic accidents happen not only because drivers are inattentive, but also because of adverse weather conditions. Hazy weather is a typical weather condition that leads to more challenging detection and safety problems for autonomous vehicles due to the degradation of visibility, and blurred appearance of pedestrians. Li et al. [25] generated a new pedestrian dataset in hazy weather and developed three deep learning models to address this problem based on YOLO network. The first Simple-Yolo model is similar to the original YOLO framework [16] except the loss function. The other two methods are based on priori boxes which contains multiple boxes of different sizes. The centre of the priori box is set to be at the same position as the centre of the grid cell. Each priori box is associated with only one ground truth box, which requires box translation and transformation. According to the presentation of priori box, two models are divided with different deep learning frameworks. The VggPrioriBoxes-Yolo has similar architecture with the Simple-Yolo but different output information. The MNPrioriBoxes-Yolo is improved by deep separable convolution and bottleneck architectures. In general, these de-haze proposals can automatically extract features from images with faster processing speed and less information loss. The PrioriBoxes-Yolo models are more precise for detection with reduced computational cost.

The difficulty of improving the detection quality in these specific environments lies in the insufficiency of datasets. Current datasets are mostly captured in daytime and normal weather conditions. Common problems occur in these datasets are occlusion, distant objects, and localization, while specific challenges such as low illumination and low vision are not exposed. Considering the importance of safety, functions of autonomous vehicles should be comprehensive enough to respond to various conditions.

3 Summary & Conclusion

The number of deaths in traffic accidents is considerably large, and pedestrians are more vulnerable than drivers or passengers in vehicles. As many accidents happen due to fatigue driving and inattention of drivers, the development of automated driving is expected and required to avoid these problems and reduce traffic crashes by automatically detecting pedestrians and taking actions.

In this literature review, we summarized the fundamental process of pedestrian detection. By different kinds of input devices like cameras, images containing pedestrians in urban environments are captured. A variety of datasets are obtained in this way which can be used for training in deep learning architectures. The common detection process is based on object segmentation and classification of Region of Interest (RoI). We only emphasized the feature extraction in the detection process, more detection details and summarization of traditional detection methods can be found in [4].

Common methods based on HOG features and SVM have gained some achievements in this area, but the detection accuracy still needs to be improved. The pedestrian detection in autonomous driving is challenging because there are many factors that lead to false and miss detection such as occlusion, distant objects, and various environments. These problems are fatal for both individuals and society, hence high detection accuracy is urgently required. Fortunately, deep learning approaches are emerging and make significant enhancement in pedestrian detection. We introduced a most common deep learning strategy in pedestrian detection which was Convolutional Neural Networks (CNN). Though many CNN based approaches have great performances on object detection, their achievements on pedestrian detection are limited because pedestrian detection is more difficult considering different size, postures, appearance, and orientation of people.

We focused on two successful approaches in this area: Faster R-CNN and YOLO networks. The former is based on Region Proposal Networks and Boosted Forest. The latter one also provides potential proposals but with a different regression way. YOLO proposes much fewer detection boxes and address multiple detections for the same instance. These two methods also achieve some improvements in computational speed. Faster R-CNN accelerates R-CNN baseline using fully connected neural networks, and YOLO enables to detect a group of pedestrians simultaneously. These deep learning approaches can be developed to reduce false positive and false negative detection as well as to address specific problems in pedestrian detection. Faster R-CNN was combined with Region Decomposition Branch to process nighttime pedestrian detection [23]. Li et al. extended Yolo networks to improve the detection performances in hazy weather [25]. Research on solutions of pedestrian detection in various environments is currently limited due to the lack of annotated datasets to train networks for these special conditions.

In conclusion, there are a few challenges which remain in pedestrian detection and are vital, especially in automated driving. Even though convolutional neural networks are state-of-the-art approaches in pedestrian detection, the resulting accuracy is impossible to be 100%, which means these algorithms can bring risks to autonomous driving to some extent. Moreover, pedestrian datasets are still small in scale and lack of precise annotations especially in special environments. Future research on this area can focus on solving problems that occur in specific conditions by the extension of corresponding datasets. Finally, the increasing detection accuracy in autonomous driving will expand the markets of this application, and both road users and drivers can benefit from its safety and efficiency.

References

[1] Economic Commission for Europe. *Statistics of road traffic accidents in Europe and North America*. UN, 2015.

[2] C. Urmson and W. ". Whittaker. Self-driving cars and the urban challenge. *IEEE Intelligent Systems*, 23(2):66–68, 2008.

[3] A. Brunetti, D. Buongiorno, G. Trotta, and V. Bevilacqua. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. *Neurocomputing*, 300(jul.26):17–33, 2018.

[4] A. Solichin, A. Harjoko, and A. Putra. A survey of pedestrian detection in video. *neural networks*, 2:8, 2014.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, 2005.

[6] A. Ess, B. Leibe, K. Schindler, and L. Van Gool. A mobile vision system for robust multi-person tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[7] C. Wojek, S. Walk, and B. Schiele. Multi-cue onboard pedestrian detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 794–801, 2009.

[8] M. Enzweiler and D. M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, 2009.

[9] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: A benchmark. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 304–311, 2009.

[10] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012.

[11] V. Neagoe, A. Ciotec, and A. Brar. A concurrent neural network approach to pedestrian detection in thermal imagery. In *2012 9th International Conference on Communications (COMM)*, pages 133–136, 2012.

[12] X. Liu and K. Fujimura. Pedestrian detection using stereo night vision. *IEEE Transactions on Vehicular Technology*, 53(6):1657–1665, 2004.

[13] Y. Xie, M. Pei, G. Yu, X. Song, and Y. Jia. Tracking pedestrians with incremental learned intensity and contour templates for ptz camera visual surveillance. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6, 2011.

[14] M. Karg and C. Scharfenberger. Deep learning-based pedestrian detection for automated driving: Achievements and future challenges. In *Development and Analysis of Deep Learning Architectures*, pages 117–143. Springer, 2020.

[15] M. Szarvas, A. Yoshizawa, M. Yamamoto, and J. Ogata. Pedestrian detection with convolutional neural networks. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, pages 224–229, 2005.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.

[17] R. Girshick. Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015.

[18] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.

[19] L. Zhang, L. Lin, X. Liang, and K. He. Is faster r-cnn doing well for pedestrian detection? In *2016 European conference on computer vision (ECCV)*, pages 443–457. Springer, 2016.

[20] R. Appel, T. Fuchs, P. Dollár, and P. Perona. Quickly boosting decision trees–pruning underachieving features early. In *International Conference on Machine Learning (ICML)*, pages 594–602, 2013.

[21] 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, 2018.

[22] L. Neumann, M. Karg, and S. Zhang. Nightowls: A pedestrians at night dataset. In *2018 Asian Conference on Computer Vision (ACCV)*, pages 691–705, 2018.

[23] Y. Chen, S. Jhong, G. Li, and P. Chen. Thermal-based pedestrian detection using faster r-cnn and region decomposition branch. In *2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–2, 2019.

[24] K. N. Renu Chebrolu and P. N. Kumar. Deep learning based pedestrian detection at all light conditions. In *2019 International Conference on Communication and Signal Processing (ICCSP)*, pages 0838–0842, 2019.

[25] G. Li, Y. Yang, and X. Qu. Deep learning approaches on pedestrian detection in hazy weather. *IEEE Transactions on Industrial Electronics*, 67(10):8889–8899, 2020.