electronics
MDPI

Review

Deep Learning-Based Pedestrian Detection in Autonomous Vehicles: Substantial Issues and Challenges

Sundas Iftikhar [1,†], Zuping Zhang [1,*,†], Muhammad Asim [2,3,*,†], Ammar Muthanna [4,5,†], Andrey Koucheryavy [5,†] and Ahmed A. Abd El-Latif [3,5,6,*,†]

1. School of Computer Science and Engineering, Central South University, Changsha 410083, China
2. School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, China
3. EIAS Data Science Lab, College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia
4. Department of Applied Probability and Informatics, Peoples' Friendship University of Russia (RUDN University), Miklukho-Maklaya, 117198 Moscow, Russia
5. Department of Telecommunication Networks and Data Transmission, The Bonch-Bruevich Saint-Petersburg State University of Telecommunications, 193232 Saint Petersburg, Russia
6. Department of Mathematics and Computer Science, Faculty of Science, Menoufia University, Shebin El-Koom 32511, Egypt

* Correspondence: zpzhang@csu.edu.cn (Z.Z.); asimpk@gdut.edu.cn (M.A.); a.rahiem@gmail.com or aabdellatif@psu.edu.sa (A.A.A.E.-L.)

† These authors contributed equally to this work.

Abstract: In recent years, autonomous vehicles have become more and more popular due to their broad influence over society, as they increase passenger safety and convenience, lower fuel consumption, reduce traffic blockage and accidents, save costs, and enhance reliability. However, autonomous vehicles suffer from some functionality errors which need to be minimized before they are completely deployed onto main roads. Pedestrian detection is one of the most considerable tasks (functionality errors) in autonomous vehicles to prevent accidents. However, accurate pedestrian detection is a very challenging task due to the following issues: (i) occlusion and deformation and (ii) low-quality and multi-spectral images. Recently, deep learning (DL) technologies have exhibited great potential for addressing the aforementioned pedestrian detection issues in autonomous vehicles. This survey paper provides an overview of pedestrian detection issues and the recent advances made in addressing them with the help of DL techniques. Informative discussions and future research works are also presented, with the aim of offering insights to the readers and motivating new research directions.

Keywords: self-driving cars; pedestrian detection; deep learning; CNN; faster R-CNN; MobileNet-SSD; multi-spectral pedestrian detection

Citation: Iftikhar, S.; Zhang, Z.; Asim, M.; Muthanna, A.; Koucheryavy, A.; Abd El-Latif, A.A. Deep Learning-Based Pedestrian Detection in Autonomous Vehicles: Substantial Issues and Challenges. *Electronics* **2022**, *11*, 3551. https://doi.org/10.3390/electronics11213551

Academic Editor: Juan M. Corchado

Received: 2 September 2022
Accepted: 17 October 2022
Published: 31 October 2022

1. Introduction

Pedestrian detection is a computer vision technique and one of the most important functions for autonomous vehicles to be able to detect human motion in their path, which is helpful to ensure the safety of the people, recognizing and pursuing a culprit in a crowd, preventing accidents and avoiding moving vehicles and obstacles. Such detection tasks can be performed with the help of an advanced combination of sensors such as radar, camera, and light detection and ranging (LiDAR). In recent years, a system named Advanced Driving Assistance System (ADS) has been introduced that is helpful in the prevention of unpredictable accidents. This system has many features to substructure multiple tasks such as the protection of the commuter, environment, and drivers. Pedestrian detection is one of its established features. Subsequently, engineers added this feature to autonomous cars. However, with this feature, pedestrian detection still faces a lot of issues that need to be resolved. Through different innovations, many researchers have tried to solve these issues.

These challenging problems are poor obstacle detection under different lighting conditions such as clear visibility problems at night time, occlusion conditions, low resolution, tiny size occurrence, and the tracking and recognition of pedestrians [1–4]. These problems are sorted with the help of different techniques that can be seen in Figure 1. Figure 1 demonstrates the number of papers related to pedestrian detection from 2000 to 2021. In the beginning, traditional techniques such as machine learning techniques were used due to their tremendous results from 2005 to 2015, however, from 2015 to 2017, researchers moved to new "hybrid" approaches because these approaches yielded the best result; however, they also suffered from the same issue as previous traditional techniques, i.e., the features were not manually extracted. Recently, the use of deep learning (DL) has become much more popular compared to previous traditional algorithms because of its great performance, results, and the expertise it has established. Jones and Viola increased the real-time detection capabilities and effectiveness through the famous VJ infrared [5]. Romero and Antonio [6] mostly described DL algorithms, however, some of them were outlined and failed to present the abundant and clear characteristics of the design, for example, the technique and databases it used, problems in its behavior, and the results obtained.

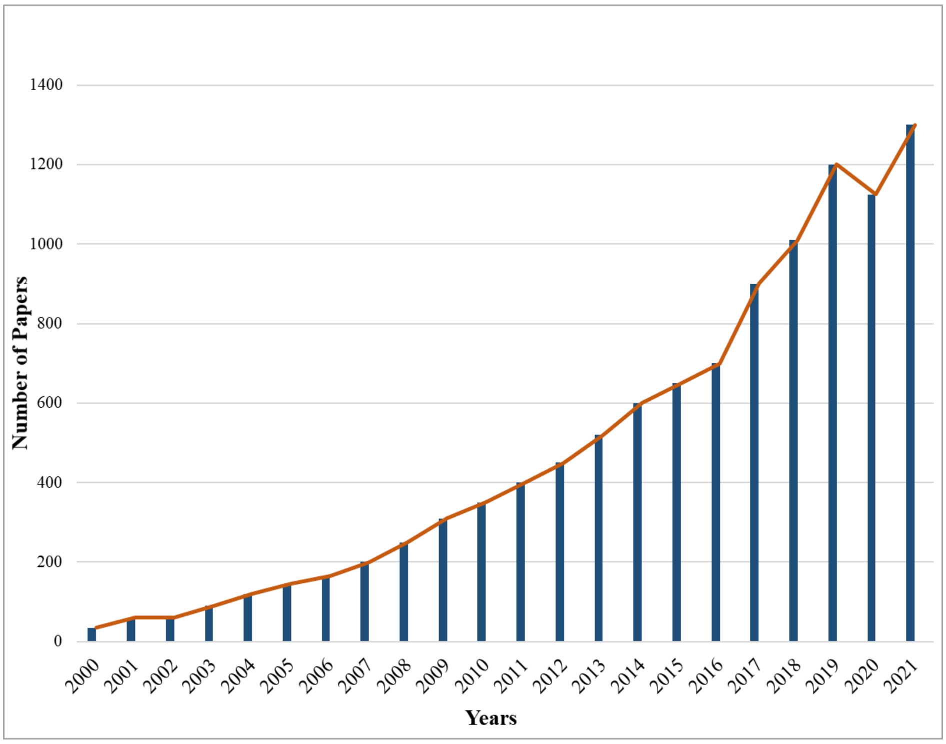


Figure 1. The number of papers affiliated with pedestrian detection increased from 2000 to 2021.

Zhu et al. [7] investigated key issues in long-distance pedestrian detection by the combination of background subtraction and DL techniques. This method has two processing steps. In the first step, this model provides the facts of machine learning detachment frameworks. In the second step, the execution of the identification of small pedestrians by the RefineDet apparatus is enhanced using the attention module. In order to ensure validity when using this technique, it is mandatory to use additional benchmarks that have been gathered from various cartographic locations, with an abundant amplification of pedestrian traffic. In [8], the authors presented YOLOv3, faster R-CNN, and MobileNet-SSD algorithms to determine the true false pedestrian. This approach uses the KITTI and Waymo benchmarks where 110 random samples were marked to verify the actual pedestrian with a learning rate of 0.5–0.9. Moreover, to further enhance this model, data augmentation techniques were utilized and employed the three numerous size dimensions indicator systems 14×14, 27×27, and 53×53 to identify the target. Another study by [9] attempted to determine the difficulty of pedestrian detection using the YOLO configuration.

In particular, the authors refined the real YOLO configuration by initiating a new web design, which they called YOLO-R to more precisely mark pedestrian identification. In the framework proposed by the authors, they enumerated three more transition layers and interchanged the aggregate of the layers that linked in the root layer. The authors investigated their construction on the INRIA benchmark [10] and were able to boost the correctness of pedestrian detection. Some wide-ranging and abundant day–night datasets were proposed, for instance, CityPersons, KITTI, and the color NightOwls datasets by Zhang et al. [11], Neumann et al. [12], and Geiger et al. [13], respectively, to detect pedestrians through a wider size annotation. In [14], a brightness perception model was developed to determine whether it was under day or night conditions. Subsequently, RGB cameras were used to detect pedestrians during the day, while thermal depiction cameras were utilized at night.

Additionally, for a moderate resolution, a deep convolutional generative adversarial networks (DCGANs) approach was suggested in [15] to enhance the attributes of the videos and images because the targets in the interspace largely fade in the videos or images, which causes false detection. This model aimed to detect small-size pedestrians and indeterminate optical features. From deeper surfaces to shallow surfaces, Zhang et al. [16] introduced a saliency loss detection framework that transferred general information about an image. In [17], Navarro et al. used sensor-based automation systems to recognize pedestrians in the applications of autonomous vehicles. Li et al. [18] presented an SAF RCNN model in 2018 based on the knowledge theory. The aim of this approach was to effectively boost the performance of pedestrian detection at various ranges. It can also upgrade the capacity to detect the ordinary target, but since the target-scale variations are more habitual in pedestrian detection areas, the improvements in ordinary target detection are restricted. In addition to this, deep learning technology has also been used to control microbial electrochemical systems such as MFC [19], MEC [20], MDC [21], and MRC [22]. Moreover, a summary of pedestrian detection development in different aspects based on DL approaches such as YOLOv3 and faster R-CNN is outlined in Table 1. The main goal of this survey paper was to review pedestrian detection using the DL approach in autonomous vehicles.

Table 1. Developmental summary of pedestrian detection based on different perspectives.

Authors	Challenges	Area	Models	Results
Kim et al. [23] 2020	Pedestrian detection issues in smart towns	Facing issues due to complex environmental components, parameters, and discord in images	Utilized CNN to build up the advance VGG-16 and vision-based techniques	High accuracy up to 98.8%
Chen et al. [24] 2020 Su Hang et al. [25] 2015	Difficult to identify the pedestrian because the images are captured from one position; no paradigm to stimulate the operations against the movements operated by pedestrians	Pedestrian detection evolution in intelligent transport design	Used the support vector machine (SVM) R-CNN to identify the one- and two-step patterns with the help of Google AVA, Hollywood2, KTH, and UCF sequence	The accuracy rate is 85.5%
Dinakaran et al. [26] 2019 Tian et al. [27] 2015	Reduce long-distance low-resolution problems and control the occlusion handling in pedestrian detection	Detection of vehicles, cyclists, and pedestrians in smart towns due to security issues in transmission generated in IoT systems	Presented a new DCGAN model with cascaded single short detectors (SSD) based on Canadian Institute for Advanced Research (CIFAR) datasets; presented a DeepParts model to handle the occlusion issue based on KITTI and Caltech datasets	Accuracy rates are 80.7% and 70.49%
Wang et al. [28] 2020	An occluded pedestrian resulting in missing information leading to the identification of a false negative pedestrian	The bad reaction of pedestrians to traffic conditions in urban areas of China	Proposed different methods such as FichaDL, THICV-YDM, DH-ARI, and EM-FPS based on KITTI and Caltech datasets	The accuracy rate on the KITTI dataset is 88.27% while that on the Caltech dataset is 81.73%
Hbaieb et al. [29] 2019	To overcome the time response issues in pedestrian detection during the change of weather situations and various road circumstances	Camera quality effects in urban areas	Detection was performed based on support vector machine (SVM), histogram of oriented gradients (HOG), and Haar cascade techniques	The accuracy rate is up to 90% to 93.43%
Navarro et al. [17] 2016	To reduce the pedestrian detection challenges based on a sensors system under real driving circumstances	Perceptions were performed in crowded places	Proposed machine learning (ML) approaches such as SVM, k-nearest neighbors (kNN), Naïve Bayes classifier (NBC)	The accuracy rate is 96.2%
Aledhari et al. [30] 2021	To reduce the poor performance of algorithmic bias in the detection of human skin for instance poor detection due to a darker skin color under a complex situation such as variations in images and illuminations; moreover, a darker skin color also causes occlusion and other issues	Darker skin tones cause serious accidents in some areas of America	Proposed K-Means Cluster, YOLOv3, and CNN for the classification of skin tones based on the Caltech pedestrian detection dataset	The mAP is 43%

In this survey paper, we discuss the three main issues of pedestrian detection with the help of the DL approach, namely occlusion, low-quality images, and multi-spectral images, and present an evaluation of the pedestrian detection performance. Initially, from LiDAR and camera sensors, various benchmark datasets are composed by gathering data from real statistics using DL techniques to solve the pedestrian problem. The second approach consists of applying DL algorithms including CNN, MobileNet-SSD, faster R-CNN, and YOLO versions such as YOLO, YOLOv3, and YOLOv4 to capture the images. Finally, the metrics are illustrated to estimate the execution of the paradigm evaluation. In the evaluation, different image sources including RGB, thermal, and multi-spectral formats are compared for the performance of pedestrian detection.

The rest of this paper is organized as follows. Section 2 presents related works relevant to pedestrian detection. Section 3 provides a short overview of datasets. Section 4 discusses the pedestrian detection structure. Section 5 provides traditional vs. DL approaches. Section 6 describes occluded pedestrian detection. Section 7 provides a comparison of various approaches on different datasets. Finally, Section 8 offers a detailed discussion along with directions for future works, followed by the conclusion in Section 9. For more clarity, the organization of the paper is given in Figure 2.

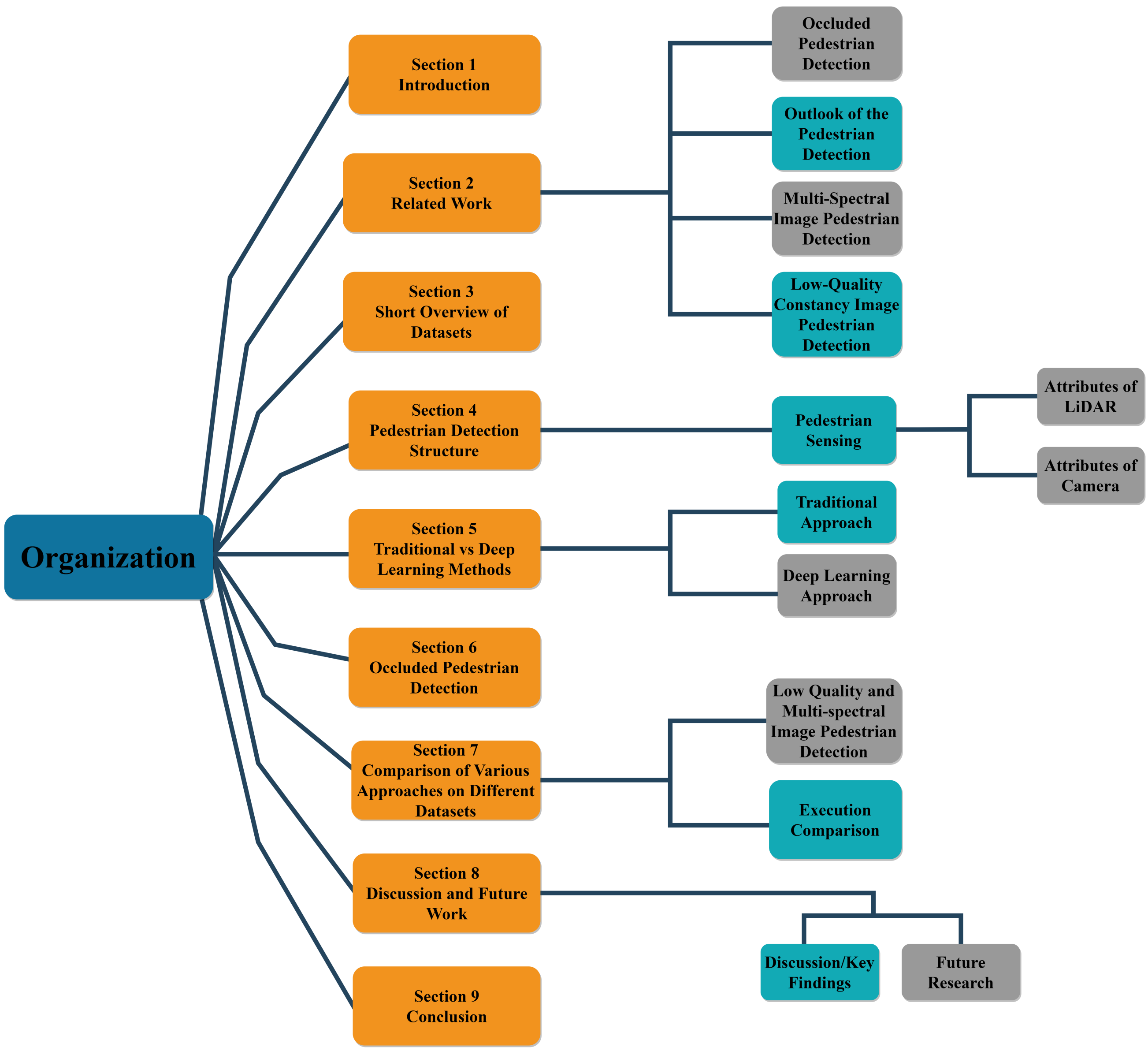


Figure 2. The organization of the paper.

2. Related Work

2.1. Outlook of the Pedestrian Detection

Pedestrian detection in a difficult situation, for instance, occlusion, nighttime conditions, and low resolution is an issue that is still far from being resolved. These shortcomings make it difficult to implement vision-based approaches in applications that require 24/7 operation, such as autonomous driving. To address these issues, different kinds of sensors have been developed in addition to visual optical cameras (VISs), such as depth cameras and infrared (IR) cameras. For the detection of pedestrians, thermal images typically capture the sharp contours of the human body [31,32]; on the other hand, visual optical cameras provide a magnificent visual explanation of human specimens. In fact, thermal and color perception media provide additional facts. Numerous previous studies have solely focused on the detection of pedestrians in color or thermal perception [33–35]. Some current papers use both color and thermal images [36–38]. Nguyen et al. [39] reviewed the progress and problems of the pedestrian detection algorithm. They discussed the latest algorithms during the years 2010 and 2015 and were mainly grounded in traditional approaches. He negotiated that the performance of the pedestrian detection algorithm is heavily dependent on feature extraction which is utilized to create identifiers. The authors solely focused on trained and tested algorithms with the use of the Caltech dataset.

Another study performed by Rajesh and Ragish [40] proposed a comprehensive overview that covered the certain necessity for an advanced driver assistance systems structure. For the detection of pedestrians, they shield the DL and traditional approaches in which various matrices were tested. Furthermore, for pedestrian detection algorithms, they presented trends and tips for future work. However, the indicated DL algorithms, e.g., a recurrent neural network (long short-term memory), were insufficient, and despite the framework of the encoder and decoder, the objects were not declared. On the CityScape and Caltech datasets, the models were trained and tested. However, with the passage of time, convolutional neural networks (CNNs) have gained a significant advantage in finding common objects on a network, such as on MS COCO datasets [41], Pascal datasets, and ImageNet [42]. Li et al., in 2018, presented the situation analysis framework RCNN grounded in the perception hypothesis [18], which successfully enhanced the accomplishment of pedestrian detection at various scales. Now, CNN was further extended to detail the major challenges of pedestrian-like occlusion manipulation by labeling the various body parts, low-quality resolution images, and multi-spectral images including color, RGB images, thermal images, and simultaneous facts in their entirety. Some major previous tasks performed by researchers regarding these issues are given below.

2.1.1. Occluded Pedestrian Detection

Occlusion often occurs in the real world and it is very difficult to find occluded pedestrians, especially in autonomous driving scenarios. Many researchers have used different information for the body parts, for instance, the leg, arm, and head, in addition to other approaches to detect pedestrians with the help of complete body gestures. However, to solve the occlusion problem, researchers have divided pedestrian detection into two kinds of categories: namely traditional methods and the deep neural network approach. In the traditional approach, different methods are used which are the histogram of oriented flow and gradient [10,43], Haar wavelet [44], local binary variant, support vector machine [45], etc., to extract features. However, these methods have limited generalization capabilities to handle handicraft attributes. As such, for further desirable results, researchers have had to move towards adopting DL methods to solve the issue of occluded pedestrian detection. Pedestrian detection based on DL approaches such as MobileNet-SSD, R-CNN, fast-RCNN, and faster R-CNN has created a landmark by enhancing the performance including by manipulating the difference in the radiance level and composite situation with various pedestrians. Zhang et al. [46] presented the occlusion-perceptive R-CNN to enhance the accuracy of the pedestrian in the crowd. On the contrary, various additional works were presented by Zhang et al. [47], Ouyang and Wang [48], to comprehend several occlusion designs in a joint procedure that reforms the abundant testing and training time. Nonetheless, the final choice is even now being carried out by integrating numerous module scores, which when combined, can make the entire operation more difficult and rigid to train. On the other hand, a constant focus vector that is accessible to train and has a low cost is also being studied.

2.1.2. Multi-Spectral Image Pedestrian Detection

In 2015, Huang et al. [36] developed a multi-spectral pedestrian dataset. After their publication, more work on the multi-spectral approach was published later. Choi et al. [49] simultaneously extracted the thermal and RGB images in DNN. In addition, a single shot (SSD) was used for multi-spectral pedestrian tracking [50,51]. Furthermore, Zhang et al. [52] utilized the extended decision trees for the categorization project. Converting thermal and RGB images into a regional proposal network (RPN) may yield a better outcome. For the tracking of a pedestrian on thermal images at night, Chen et al. [53] implemented a carefully controlled encoder–decoder CNN.

2.1.3. Low-Quality Constancy Image Pedestrian Detection

Based on DL [54], the super-resolution began to start using CNN [55] and was named a SRCNN model for the first time for high-resolution reformation. By adopting the end-to-end three-terminal deep convoluted network, resulting in state-of-the-art high-definition performance, Chao et al. [56] presented FSRCNN and suggested that the network be allowed to learn direct denoising filters, thus further increasing speed and correctness. A very deep super-resolution VDSR [57] was the uppermost means of putting spherical residues into super-resolution, which significantly increases the training rate and greatly upgraded the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) diagnostic measures.

3. Short Overview of Datasets

Datasets have played an essential role throughout the history of object identification research. It is not just a common point used to measure and compare algorithms/competitors' performances, but it also being increasingly promoted in the field of research concerned with complex and challenging issues. Especially recently, DL technology has enabled a great success for humankind. Many visual recognition issues and many more interpreted data successfully perform a vital role to access a large number of images on the Internet. A comprehensive dataset can be created to capture the abundance and diversity of objects that have achieved unparalleled efficiency in object identification. Constructing large datasets with minimal deviation is essential for the development of modern computer vision algorithms. In object detection, many well-known datasets and benchmarks have been published in the previous 10 years, and counting public object detection work, pedestrian tracking has its special characteristics. Familiar pedestrian detection datasets currently are the EuroCity [58], TUD known as Brussels Pedestrian Dataset [59], Caltech [60], CityPersons [11], INRIA [10], KITTI [61] and ETH datasets. Some special datasets for pedestrian detection, which are commonly used in experiments, are listed in Table 2.

Table 2. Summary of a few pedestrian detection datasets which are commonly used.

Datasets	Methods	Training Images	Testing Images
KITTI	PCN [62], ECP faster R-CNN [58], faster R-CNN [63], Sub-CNN [64]	7481 images	7518 images
INRIA	PCN [62], SAF R-CNN [34] 2LDCF [65], RF3+LDCF [65]	614 is used as a positive image and 1218 used as a negative image	288 images
Caltech	PCN [62], RPN+FRCNN [63] SAF R-CNN [34], HOG [10]	350,000 images	2300 images
CityPersons	Adam solver ImageNet Model [66]	2975 images	500 images
TUD-Brussels	Part-based model [67]	218 images used as negative	508 images used as positive
ETH	Part-based model [67] Faster RCNN [68,69]	499 used as positive images	1804 negative images

The characteristics of each of the datasets listed in Table 2 are given as:

(1) The KITTI dataset comprises pedestrians of various perspectives, degrees of occlusion, and sizes. For detection, it enhances the DL training; (2) The INRIA dataset is used to verify the model generalization abilities; (3) The Caltech dataset is the most famous dataset which gives the best performance in occlusion handling such as in limited and dense occlusion conditions; (4) The CityPersons dataset manifests high divergence; (5) The TUD-Brussels dataset works efficiently in contrast to deformation; and (6) The ETH dataset works efficiently in contrast to deformation.

KAIST is a dataset that is used as a multi-spectral pedestrian dataset [36]. This dataset contains thermal and RGB images which recorded the data of colleges, rural areas, and roads using three types of labels, for example, human beings, motor-bikers, and people. These data recorded the day and night situations. For training purposes, 14,100 pictures were used in the daytime, and 8058 pictures were used in nighttime situations. It is difficult to find detailed labels for body parts in public datasets, such as the Caltech or CityPersons datasets, which do not have labels for any part of the body to detect pedestrians. In contrast, the Penn–Fudan dataset [70] can easily label different parts of the body. However, there are some shortcomings to the Penn–Fudan dataset. As it is a generated dataset of 1500 images with complete body parts, it must divide the whole body arch into three sections: the legs, the heads, and the arms. This comprises a large part of the pedestrian's body. Furthermore, the data were collected from around the world which caused changes in the dataset (for instance, between the USA, Bangladesh, Malaysia, and India). INRIA is another dataset that was used to train the SAF R-CNN, advanced FCF, and PCN. The miss rate was used as an evaluation metric. The execution of these frameworks was compared to the other 11 models. The error rate ranges between 6.9% and 17.28%. In the models' comparison, the ACF has a bad execution with an error rate of 17.28% and on the other hand, PCN yields the greatest result with an error rate of only 6.9%. Furthermore, benchmarks such as WiderPerson [69], CrowdHuman [71], and Wider Pedestrian [69] are comprised of images on the web to provide more variety and density. This allows the detector to more sharply comprehend the characterization of pedestrians with greater generalization expertise. CityPersons [72] is a more divergent dataset in contrast to CityPersons [65]: whilst the Caltech dataset is registered in 27 various German cities and adjacent countries, in contrast to CityPersons, it is based on approximately 31,000 comments on bounding boxes, and in addition, contains 2975, 500, 1575 images for its training, testing, and validation groups, respectively.

The ETH dataset [68] contains three layouts for testing (1804 images in total). Because the film was shot in the middle of the city, it can accommodate large crowds, making it a suitable testing ground for occluded pedestrian detection. Finally, EuroCity Persons (ECP) [73] is a current dataset for pedestrian detection which outperforms the Caltech dataset and CityPersons dataset in terms of difficulty and heterogeneity. It is noted that it is based on data from 31 towns across 12 states in Europe. Europe has day and night photos (therefore acting as an umpire, as ECP is called daytime and ECP is called nighttime). The defined limit box is higher than 200K. In ECP [73], all examinations and comparisons were performed during the daytime in collaboration with other approaches. A diagnostic server is available. However, test sets and frequency submissions are limited.

4. Pedestrian Detection Structure

Pedestrian detection algorithms mainly follow the primary framework as shown in Figure 3.

In the first step, the sensor system collects the data in the formation of images. In the second step, a regional proposal approach is put into it. ROIs are also known as the region of interest, which has been used as a generally visual technique, for instance, in camera and stereos. However, this is the first and essential step in system tracking. Elements such as borders, lines, and figures are extracted and refined using classifiers to determine the class of a target (for example, whether the target is a human being or not). The ROIs are presented in an image that is suggested to identify pedestrians at the scene. For searching ROIs, different techniques are used such as the sliding window, locally de-correlated channel features (LDCF), and selective search. Then, the ROI features are extracted in the third step. For object detection, the algorithms used for classification and feature extraction are manual or DL-based object detection techniques. Hand-crafted approaches for feature extraction are deployed on models built on lower-level features to manually recommend ROIs [74]. Handcrafted approaches can be bounded and not extremely vigorous as complicated features may be tough to handcraft. DL methods enables the network to specify properties. This can furnish the top level of extraction. At the end of the classifier, features are augmented into the last step. The output element produced by the subtraction step is entered into the classifier to determine whether pedestrians or other obstacles in the form of binary tags are present in the proposed area. However, with the development of DL, more CNN-based sequencing methods are being used for classification compared to previous classifiers, for instance, SVM and AdaBoost.

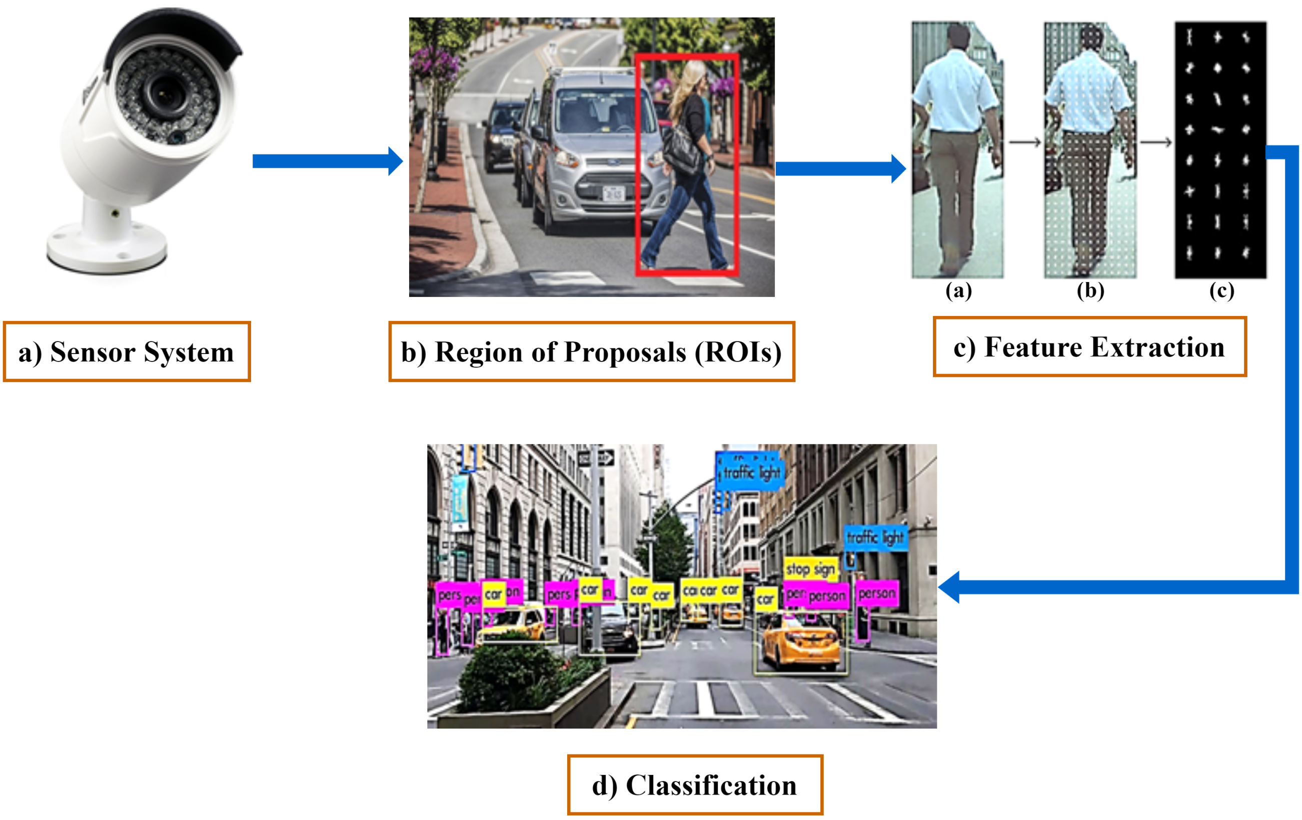


Figure 3. Primary Pedestrian Detection System Framework.

4.1. Pedestrian Sensing

Sensors are an important part of automotive computerized control systems. Automotive triggers must strike a complicated stability between reliability, robustness, manufacturing, compatibility, and low cost. Any pedestrian modeling system should start by gathering sensor information about pedestrians. Detection, tracking, and well-known models can all rely on detailed knowledge at this lower level. Table 3 summarizes the typical autonomous vehicle pedestrian detection sensors and their accuracy and range. Here, we present an overview of LiDAR and camera sensors because the camera is often the most operated sensor as the fundamental element of the pedestrian detection system and LiDAR provided the best accuracy compared to radar, mostly under bad conditions under 200 m.

Table 3. Accuracy and Range of Common AV sensors for the detection of pedestrians.

Sensors	Range	Accuracy
STEREO CAMERAS	Ranges from five hundred centimeters to various tens of centimeters [75] Several tens of meters [75]	Divergence delusion of 1/10 pixels (corresponds to approximately 1 m range delusion if the target is 100 m out of the way) [76]
INFRARED	From a minor centimeter to various centimeters [77,78]	Temperature precision of +/−10 °C, can calculate the temperature up to 3000 °C [77]
ULTRANSONIC	Starting from 20 mm up to 5000 mm [79,80]	Approximately 0.03 cm [79,80]
RFID	Certain meters [81,82]	Certain centimeters [81,82]
LIDAR	Starts range onward 300 m [83,84]	Starts onward from 2 cm [84,85]
RADAR	Short range: 40 m, angle 130° [86–88] Middle range: 70 m–100 m, angle 90° [86,87] Long-range automotive radar: From below 1 m onwards to 300 m (beginning gradient onward +/−30°, a comparative velocity scale of onward +/−260 km/h) [85,86,89]	Short range: below than 15 cm or 1 % [86–88] Middle range: below than 30 cm or 1% [86,87] Long range: 10 cm such as Long-Range-Radar LRR3 Bosch 77 GHz, scale from 250 m [85]

Although humans primarily use their vision and hearing depending on the driving system, there is a method of artificial perception. There are many ways to overcome the shortcomings of a sensor. With a broad variety of detectors used in autonomous vehicles, we divided the reviews into passive and operational detectors. Operational detectors including LiDAR, radar, and sonar actively forward vibration to the surrounding area after which they are identified and reflected; on the other hand, passive detectors, such as monocular and stereo cameras, detect physiological phenomena that already exist in the environment. The idea of AVs is mainly focused on the research of cameras and LiDAR. In the follow up, we narrate the two schemes which were used to gather information from the camera and LiDAR for pedestrian categorization. A more comprehensive current sensor report on AV applications can be found in [89,90].

4.1.1. LiDAR vs. Camera

LiDAR is alike to traditional detectors in that it uses the vibration of infrared light to identify the nearby pedestrian. Traditional detectors use an electromagnetic spectrum. LiDAR uses beams to observe nearby circumstances. LiDAR flashes its laser at an object at the speed of millions of beams per second to generate a 3D graph using an on-board operating system to accommodate the car with knowledge of its neighborhood; this layout—which has a 360-degree vision—assists in operating the car in any kind of situation and measures the change in the car's distance from the object when the laser pulse bounces off and hits the car. The system must be accurate to make quick decisions with a faster response time than humans. A common LiDAR framework, for instance, the HDL-64L [52], uses a series of rotating ray emissions to achieve a 3D point cloud within 360 degrees and a radius of up to 120 m. These detectors can yield 120,000 localities per frame, which is equivalent to 1.2 billion localities per second at 10 Hz frame per second. Velodyne have now released the VLS-128 approach [53] with 128 ray emissions, high angular resolution, and a 300 m orbit limit. Some techniques depend on LiDAR and camera modes. Before combining these approaches, the sensor must be calibrated to obtain a single local reference frame. Park et al. [54] recommended using planner boards in the hope of identifying both methods and produce accurate 3D and 2D connections and obtain the correct sequence.

4.1.2. Benefits of LiDAR

- Among the key benefits of LiDAR are its precision and correctness. The aim is for Waymo to protect its LiDAR structure through its accuracy. Navigate reports that Waymo's LiDAR is very up-to-date, and that it can determine the position of pedestrians and can estimate their movement. Equipped with a Waymo LiDAR, the Chrysler Pacificas can tell which way a cyclist should turn by looking at the gestures used by cyclists.
- Another advantage of LiDAR is that it provides a 3D image for autonomous vehicles. LiDAR is more accurate than cameras because the laser will not be confused by daylight, blazing, shadows, or entering the car front lights.
- In conclusion, LiDAR liberates computing capacity. LiDAR can instantly notify of the distance and direction of an object, while the camera-form program must first take pictures and then examine the images to regulate the distance and velocity of the object, which requires more computing power.

4.1.3. Limitations of LiDAR

- LiDAR also has limitations, as there are still many systems that cannot penetrate well through fog, snow, and rain weather conditions. Ford, which is extraordinarily superior in making self-driving cars, has established a design that can help its LiDAR network distinguish between isolated dewdrops and avalanches. Apart from that, the self-driving car will interpret the avalanche falling in the highway medium as a wall. Ford demonstrated their courage in the Michigan test, but its strategy still has much to comprehend.
- Furthermore, LiDAR does not provide data that the camera can normally see, such as text on signs or the shade of traffic lights. The camera is more suitable for this type of information.
- At last, LiDAR systems are very heavy because they need a laser rotation system to be installed throughout the vehicle, whereas the camera system used in existing Tesla cars is almost invisible.

If one wants to navigate through something such as a crowd of humans, the visible identification of objects is the strategy to proceed with. This is the general reason for using the camera system. The images provided by the camera can be used for high-accuracy analysis using AI software. In Tesla models, the camera is used to yield a 360-degree vision of the surroundings through its autopilot function. It is completely optical and does not depend on range and detection as LiDAR does. Rather than using illumination vibration, the camera uses visible information from the lens optics to return to the onboard software for further inspections. With the evolution of neural networks and computer vision algorithmic programs, the target can be recognized while driving to provide vehicle information. This can assist cars by aiding them in preventing crashes, slowing down in traffic, changing lanes, and using optical character recognition (OCR) to study text on the pavement or highway signs. To date, Tesla has proven that self-driving vehicles can operate in the absence of a LiDAR that uses a camera.

Monocular cameras present comprehensive knowledge about the pixel intensity, display shape, and appearance properties. The appearance and shape data can be applied to determine the roadway geometry, object class, and road signs. The disadvantage of monocular cameras is the absence of depth knowledge needed to assess objects of the correct size and location. One can use the stereo camera to reset the deep channel. This algorithm requires finding the correspondence between two images and calculating the depth of every position at a slower pace than the camera's additional processing intensity. Other modal cameras that provide depth evaluation are time-of-flight cameras, where the depth is derived by calculating the delay in the middle of transmitting and collecting moderate infrared radiation. This technology has been used in vehicle safety applications but still has a lower cost of integration and algorithmic complexity compared to the stereo camera solution.

4.1.4. What Are the Main Reasons for Camera's Popularity?

First, cameras are considerably cheaper compared to LiDAR systems, which lowers the cost of self-driving vehicles, though mostly for end users. They are also easy to integrate (The Hyundai Tesla has eight cameras throughout the car) because on market video cameras are widely available. Tesla could easily buy and improve commercial cameras instead of giving out and innovating some brand-new technologies. Another advantage is that the camera will not turn a blind eye to weather circumstances such as fog, rainfall, and snowflakes. Software improvements should be made to enhance in Ford's LiDAR ability under severe circumstances, however, Tesla's cameras do not have problem resembling as LiDAR's temporal restrictions. Regardless of where the human passenger desires to go, the camera system will follow. The camera can observe the world like a human, and theoretically, unlike LiDAR, it can read road signs and interpret colors. Finally, the camera can be easily integrated into the layout of the car and made invisible among the structures of the car, making it more attractive in consumer vehicles.

In the detection of pedestrians, self-driving cars have a software element that is common to both LiDAR and the camera. Both systems use artificial intelligence technologies such as machine learning and neural networks to investigate information. As the algorithm improves, the result should generate higher accuracy in object recognition and enable self-driving vehicles to manufacture better commitments. It can distinguish between accidents and safe driving.

4.1.5. Limitation of Cameras

- When the lighting conditions change so that the subject becomes blurred, the camera encounters the same problem that humans face, e.g., a situation where intense shadows or glare from the sun or upcoming cars can create chaos. This is a common reason for which Tesla is still adding radar to the forefront of its car to provide further input (which, compared to LiDAR systems radar, is much cheaper).
- Cameras are also relatively "dumb" sensors because they lay out the system with only raw image data, without the exact distance and location of objects as LiDAR does. This means that the camera system must depend on strong machine learning (such as neural network or DL approach) computers that can operate these images to precisely regulate where to place them. As our human brain acts on stereo perception with the eyes to regulate the distance and position.
- To date, neural networks and machine learning systems are not strong enough to transfer massive amounts of data from cameras so that all the information could be prepared in time to make management decisions. Nevertheless, the growth of neural networks has become increasingly complex and can handle real-world inputs better than LiDAR.

5. Traditional vs. DL Approaches

The algorithm for the pedestrian detection structure is divided into three parts: one is the traditional method, the second one is the DL method and the third one is the hybrid method. The hybrid method combines both traditional and DL approaches. Further description of the traditional and DL approach is described in the next part. Moreover, the analysis of this investigation was performed with the help of thermal cameras and HDL-64E Lidar. For target detection, the HDL-64E LiDAR sensor leads to high performance and resolution, while thermal imaging cameras can be used to overwhelm the few limitations of stain cameras such as these cameras not being simulated by the conditions of lighting. Various surveys are using thermal features to detect and monitor pedestrians [91,92].

5.1. Traditional Approach

Different algorithms were created to detect the tasks of pedestrians; for example, in 2000, Haar was suggested by Poggio and Papageorgiou. It can demonstrate the change in the gray level of the image, which includes four groups: border function, line function, central environment function, and special diagnostic line functions. Haar is the basis of pedestrian detection automation, which in addition to Haar and histograms of oriented gradients (HOGs) [93], originated because this approach classified the target by acquiring functional data from the image via edge direction distributions [94]. Moreover, SVMs are used for classification [93]. In addition, Zhang et al. created a new attribute set with the AdaBoost classifier known as Shapelet for pedestrian detection [95]. The traditional detection approach has been used in the design of artificial features and classification. First of all, features must be extracted from the image, comprehending gray-scale, border, complexion, gradient histogram, and further information for the target. Then, the goal of the classifier is to decide which attributes are associated with the pedestrians. In addition, there are two ways in which traditional techniques deal with the main three pedestrian problems—namely (i) occlusion; (ii) multi-spectral images; and (iii) low-quality images problems. First, the objects are divided into various parts, and the visual portion can determine the positions of pedestrians. Second, pedestrians are trained on a specific general classifier to reduce the impact of disruption on daily life and carefully estimate the location of pedestrians. However, from 2015, the work of traditional approaches on different datasets such as the Caltech dataset started to reduce just because of the development of advanced technology such as DL and hybrid technology.

5.2. DL Approaches

In the 1990s, DL was first introduced as the sub-branch of machine learning and artificial intelligence [96]. Compared to the traditional approach, DL can gain a high quantity of abstraction, higher precision, and run time [97]. This is the main advantage of using DL for the detection of any object. However, with the advancement, evolution, and success of DL in pedestrian tracking, detection correctness has improved. The algorithm for the detection of a pedestrian using DL is comprised of three mainframes which are (i) recurrent neural network (RNN); (ii) based on depth belief network (DBN); and (iii) CNN.

The detection of pedestrian DL is divided into two groupings: (i) single-stage detector known as "non-regional proposal method" and "dense prediction"; and (ii) two-stage detector known as "regional proposal method" and "sparse prediction". Single-stage detector joins all the work into a single system structure; on the other hand, a two-stage detector has split the system for choosing regional data, classification, and positioning. A few regional proposal approaches include faster R-CNN [98], regional-fast convolutional network (R-FCN) [99], and region-CNN (R-CNN) [100]. On the other hand, the non-regional approach includes YOLO [101] and SSD [102–104]. These pedestrian detection approaches are the root of CNN, thus becoming the grade for pedestrian detection. However, nowadays, YOLO [105] and faster R-CNN [106,107] are the two main state-of-the-art tools in DL-based pedestrian detection.

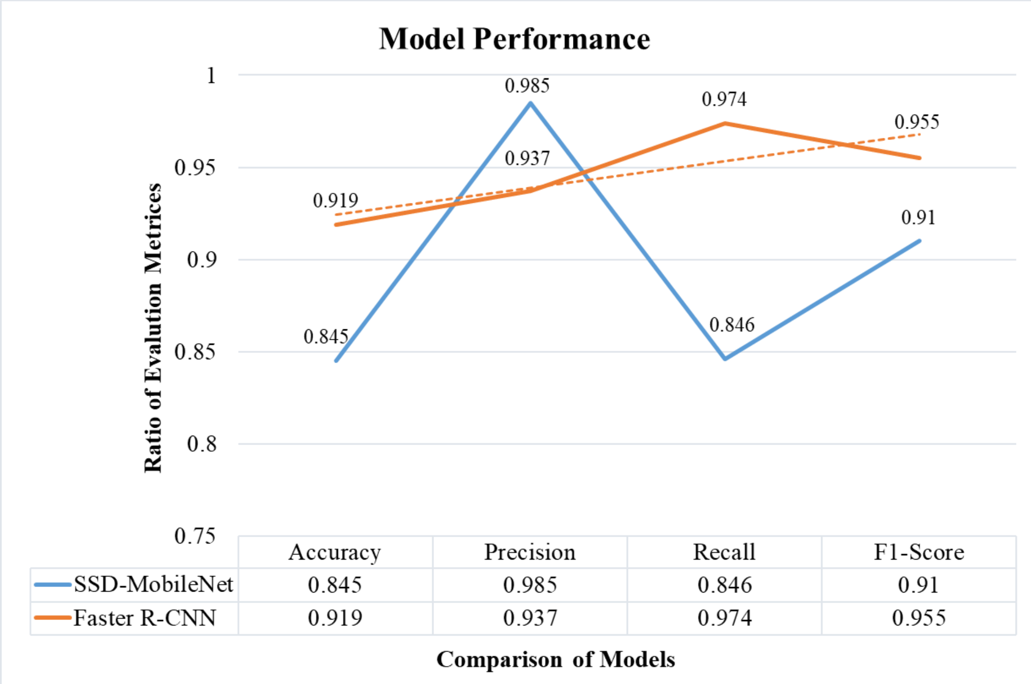
6. Occluded Pedestrian Detection

In pedestrian detection, occlusion has been demonstrated to be one of the crucial drawbacks. Because it is still difficult to find pedestrians who are being stopped by an obstacle or other pedestrians when the number of occluded pedestrians increases, the detection of pedestrians becomes complicated. CNNs are broadly used in pedestrian detection algorithms. In the DL algorithm, there are two schemes to handle the occlusion problem. The first approach is to present the design of the components of the neural network in a particular layer; and the second one is the neural optimization network diagnosis procedure. The framework of DL performs well on the whole body parts of a pedestrian due to their generalization competency. However, as an occluded pedestrian, the performance of the DL is not good enough. For the better performance of the detection of occluded pedestrians, the fusion process combines MobileNet-SSD and faster R-CNN to enhance the performance of occluded pedestrians by taking the whole-body information from public datasets such as Caltech, CityPerson, KITTI, and INRIA. The performance of this network is divided into two groups: one is a classifier and the second one is a localizer. At the structure measure, the detection method can be thought of as a sub-module of the detection system and the camera and LiDAR are used for the detection of nearby occluded pedestrians, so that a warning may be generated by the control system in possible accident situations. The main reason for using faster R-CNN and MobileNet-SSD is that faster R-CNN presents an optimization approach to employ a greater network depth and range in the system which extends the computational cost. The performance can be further enhanced by increasing the number of convolutional layers and reducing the size of the convolutional filters [108]. On the other hand, MobileNet-SSD has the same advantage as it reduces the computational difficulty in contrast to the further conventional CNN. It supplies bounding boxes and records as a result. The position of the whole body part is indicated by the bounding boxes and the probability of the targeted parts is indicated by the record within the boxes [109,110]. However, mainly for occlusion detection, the network is trained with the whole body using CNN subgroups such as faster R-CNN, YOLO and MobileNet-SSD. In the training phase, the dataset is labeled into different parts such as arm, limb, head, and person. The main aim of labeling the datasets is to easily distinguish the occluded pedestrian. For further enhancement, some data were also collected from the crowded environment. The training phase consists of a total of 1500 images using different datasets such as CityPersons, Caltech, Penn–Fudan, and self-created datasets which contain full-body data from various conditions such as the difference in lighting and size, inside and outside conditions, and occlusion obstacles. Nevertheless, the model results in more incorrect detection at a lower threshold, however, with the increase in the threshold up to 0.8, the incorrect detection is minimized.

Moreover, the other reason for the poor performance of occlusion in the detection of pedestrians is the lower ratio of occlusion instances during the training phase. A data augmentation approach was applied which remarkably upgraded the pattern and the amount of the occlusion, diversified the instances in the training phase, and effectively verified the model.

7. Comparison of Various Approaches on Different Datasets

The model prediction is mainly performed with the help of evaluation matrices such as average precision, precision, accuracy, recall, F1-score, and miss rate. Some of the predictions in terms of accuracy, precision, recall, and F1-score are given in Figure 4. From Figure 4, it is clear that faster R-CNN performs better compared to MobileNet-SSD in terms of accuracy, recall, and F1-Score; on the other hand, MobileNet-SSD performs better in terms of precision. Despite this, faster R-CNN predicts better results overall.



	Accuracy	Precision	Recall	F1-Score
SSD-MobileNet	0.845	0.985	0.846	0.91
Faster R-CNN	0.919	0.937	0.974	0.955

Figure 4. Prediction rate between MobileNet-SSD and faster R-CNN.

7.1. Low Quality and Multi-Spectral Image Pedestrian Detection

Pedestrian detection is performed with the help of multi-spectral image information because it contains the RGB, thermal, and color image data. The main aim of multi-spectral images is to reduce the restrictions in pedestrian detection such as insufficient lighting conditions and instances of small-sized pedestrians. However, the performance of multi-spectral images needs to be improved to manage these pedestrian-related issues. Therefore, an effective approach is used in multi-spectral images by combining the information from thermal and color images, and that effective approach is CNN. However, the question that arises here is that of how thermal and color data are meant to deal with the huge and tiny obstacles in pedestrian detection. The solution is based on a simple frame that consists of a duplets sub-network known as "Network-In-Network (NIN)". The implementation of this network is grounded in region-based fully convolutional networks (R-FCN). The main purpose of using two sub-networks was that one tackles the whole image to detect the huge pedestrian obstacles and the other one is used to handle and detect the small obstacles in pedestrians in a midway tiny image portion. Then, the information from both sub-networks' tiny and huge sizes using color and thermal data are then fused by network-in-network. This model has two main benefits: one is that, in the lighting situation, the pedestrian attributes become more dissimilar, and secondly, one can easily handle the small obstacles compared to the traditional region-based fully convolutional networks. This model can easily catch the ordinary attributes of pedestrians of various sizes because this model, as input, extracts the whole and sub-images and presents the detection aggregate as an output. Moreover, the detection of small obstacles is very difficult compared to that of large obstacles, which can be improved using the small-obstacle-detection RPN which is very efficient and successful in detecting small as well as large obstacles. To enable the detection of large-sized obstacles, this approach merges with Conv5; and for the detection of small-sized obstacles, this approach merges with Conv4; after combining with Conv4, the detection of small-sized obstacles is performed with Conv5. Conv4 and Conv5 are the layers of the CNN.

Nonetheless, multi-spectral images also try to solve the detection of pedestrians under bad lighting conditions and weather situations using different DL approaches such as YOLO. Lighting conditions are also an object in pedestrian tracking, as at night, human eyes cannot recognize obstacles because human eyes are oversensitive to illumination sources. Many designs were introduced to increase foresight at night. The first design was based on infrared sensors which comprise far-infrared and near-infrared. Under dim light, Piniarski et al. [111] used these two sensors to detect the obstacles with the help of connected component labeling feature schemes. After the first design, the second design was presented by Kumar and Chebrolu [112], which was known as the " brightness perception model". This model utilized the RGB-deploy model to enhance the obstacle detection in daytime while the thermal-deploy model was used to enhance the pedestrian at nighttime. After the second design, the last design was implemented which is called the "multi-spectral framework" or "Fusion design". This approach starts framing due to the release of multi-spectral pedestrian datasets. Anyhow, for lighting situations, image inventor is the best option for pedestrian detection. With the help of image sources, the performance of pedestrian detection needs to be improved in terms of precision and processing time. The image inventory includes thermal, multi-spectral, and RGB image data. Then, for model better enhancement, the YOLO algorithm needs to be optimized because it can handle the obstacle data into three propositions; in both cases, either the pedestrian stood near the camera or far from the camera because when obstacles stood near to the camera, it appear excessively large; or when it stood far way, then the obstacle size becomes excessively small, which is the main reason for the detection of a pedestrian using YOLO algorithm. This algorithm is one of the greatest one-stage detectors due to its rapid speed. Another system detection YOLO algorithm was replaced by their YOLO v3 version because it can track one or multiple obstacles that are near to one another in a robust way; this version was presented in 2018. The YOLO algorithm mainly uses the KAIST dataset because this dataset contains data in a multi-spectral form, which is why this dataset is also called the "KAIST Multi-spectral Pedestrian Dataset" [112]. Moreover, for small obstacles, further enhancement was performed including four layers of YOLO as an output attribute which are YOLO-3L and YOLO-4L after the YOLO v3 optimization process.

After the detection of huge, tiny obstacles and the detection of obstacles during daytime and nighttime, another drawback that arises in pedestrian detection and surveillance illustration is that of low-attribute images, because in low-quality images, it is difficult to discriminate pedestrians from behind the scenes or discriminate which images which are taken with low-design cameras, and had a blurred view or dense weather. To sort out this issue, a new dataset was presented known as playground (PG). The PG dataset contains images which were taken from two kinds of cameras at various times which comprised daytime and nighttime periods. A super-resolution detection network was also implemented to improve the resolution of low-quality images and can help track the blurry pedestrian behind the scene. After, when the image enhancement was performed by the SRD algorithm, the faster R-CNN model was used to help out the reluctant block pedestrian. The PG dataset is used to validate the effectiveness of the SRD model because this dataset lays out the heavy, occluded, high, and attribute resolution pedestrian data regarding gesture blurring and light intervention under daytime and nighttime conditions as compared to previous datasets, for example, the KITTI [113] and CityPersons dataset [114], because previous datasets lacked some data during the day and nighttime.

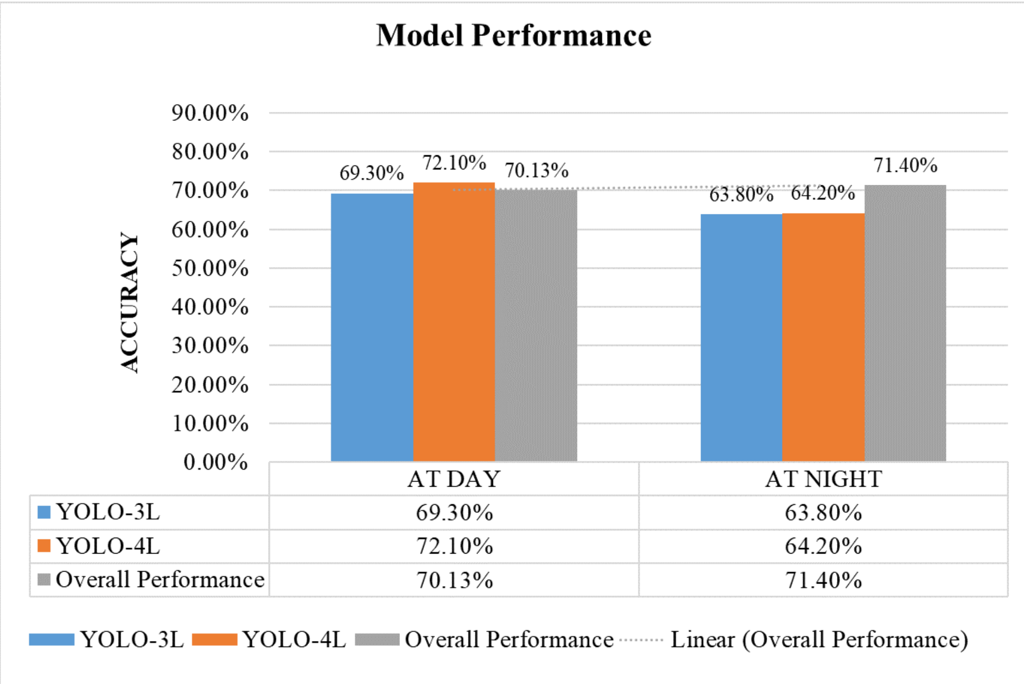
7.2. Execution Comparison

On the KAIST datasets, the network-in-network yields better results during the day and at night compared to previous work on different datasets such as R-FCN and including four faster R-CNN fusion strategies which are available in [115], whilst the range of the previous and network-in-network approach lies between 58% and 84%. NIN successfully achieved higher accuracy through the use of thermal and color data ranges between 40% and 43%. The main reason for the failure of a faster R-CNN on the MS COCO and PASCAL VOC datasets is that it cannot discover the pedestrian under conditions of a small-sized pedestrian and low resolution; moreover, this identical situation happened when work was being performed in multi-spectral pedestrian tracking. Although this was without NIN, the R-FCN has the same problem as it cannot detect the small-sized pedestrian.

Further explanations of the NIN, faster R-CNN, and R-FCN based approaches are available in [106,115,116].

Due to low-light conditions, the KAIST dataset was divided into training and testing phases containing more than 50,000 pedestrians. This dataset contains the records of colleges, highways, and midtown which are further comprised of three labels namely human beings, people, and motor-biker. Multi-spectral is considered to be the best solution for huge and tiny obstacle detection, however, with the addition of YOLO-4L, the detection improved during daytime and nighttime compared to the earliest YOLO algorithm [112]. The performance range of YOLO-3L and YOLO-4L during the daytime and nighttime is shown in Figure 5. In addition, this processing time needs to be improved with the help of compressed design, and the excellent framework achieved 22.76% refinement but still needs to improve the accuracy from 22.76% to 70.7%. Nevertheless, there seem to be some conflicts in which RGB images perform better during daytime while thermal images perform well during nighttime.

As shown in Figure 6, faster R-CNN based on SRD yields the best results on the PG dataset compared to YOLOv3, YOLO v4, SSD, faster R-CNN, and improved faster R-CNN [117], because this model helps to gain more accurate pedestrian tracking under low-attribute images.



	AT DAY	AT NIGHT
YOLO-3L	69.30%	63.80%
YOLO-4L	72.10%	64.20%
Overall Performance	70.13%	71.40%

Figure 5. Performance comparison of YOLO v3, YOLO v4, and over all executions after processing time.

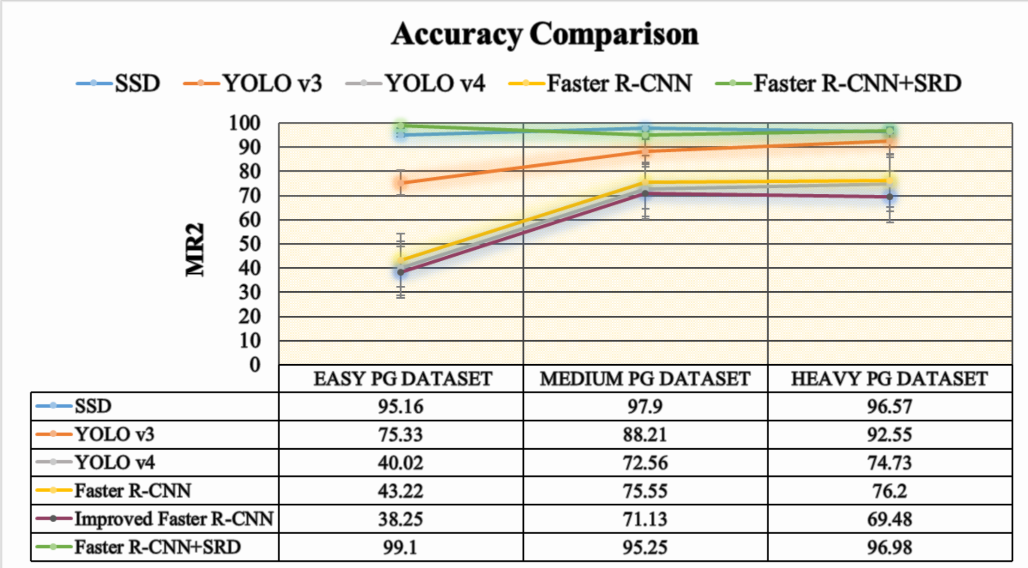


Figure 6. Detection comparison of the pedestrian using faster R-CNN based on the SRD algorithm with other models.

8. Discussion and Future Work

8.1. Discussion/Key Findings

In this study, we discussed the progress made in DL in terms of pedestrian detection. During the study, we found that there are still some main key findings in the generated approach. From the studies, YOLOv3, YOLOv4, and YOLOv5 have undergone peer review, where some writers argued that YOLOv4 is effective while others argued that YOLOv5 is effective, and some writers argued that YOLOv4 and YOLOv5 are similar in terms of detection speed. The reason for various outcomes being declared could be due to numerous factors, for instance, various datasets being used, reform meta parameters, etc. These contradictions stem from specific methods studied by other researchers. To overcome this gap, we will contrast those algorithms by contemplating the effecting circumstances in the future. In practical applications, it is necessary to maintain an equal detection balance between the speed and accuracy because the recent methods have a higher accuracy rate but the speed of detection is lower. Therefore, it is necessary to implement such approaches which can maintain the detection speed and accuracy equally and meet the demands of speed and accuracy according to practical application.

Additionally, there also seem to be some other conflicts as some contend that RGB images perform better in the daytime while thermal images perform well at nighttime during the identification of pedestrians under low illumination factors. Thus, there is a need to sort out this issue based on YOLOv5 on different datasets to verify the enhancement. Datasets play an important role in model performance. In the same model, one dataset performs well and the other datasets maybe perform badly, for instance, PG datasets may boost the enhancement in low weather conditions during the day and night as compared to KITTI [113] and CityPersons [114]. This issue is caused because the divergence of the current datasets is not sufficient. In this case, it is suggested that data augmentation approaches are applied to enrich the divergence of datasets which can enhance the generalization and strength of the frameworks in practical applications. Sensor-based detection is also another important tool to identify the correct pedestrian, as low-quality cameras are affected by urban areas and cause false detection. To overcome this issue, high-quality cameras were implemented which lead to a great computational cost. There is a need to adopt different approaches which may reduce the computational cost.

8.2. Future Research

- For pedestrian detection, better outcomes have been attained with the help of DL approaches. However, to date, the present algorithms are still facing the issue of the detection of small, moderate, and occluded objects. In the future, one can consider/address the aforementioned issues.
- In addition, there is still insufficient work examining how to enhance the detection production under bad lighting and weather conditions. In the future, this problem can be tackled by training both models with daytime and nighttime models as one paradigm and thus increase the generalization abilities.
- Furthermore, there is a need to investigate more techniques that are put together into the detection algorithms to improve the accuracy enhancement. In the future, some powerful techniques can be combined to improve the accuracy of pedestrian detection systems.
- Fuzzy logic-based algorithms can be combined with DL algorithms to improve the pedestrian detection process.
- An interesting future work may be to consider/combine 3D measures with 2D information in order to improve detection and classifications.
- Multi-class approaches should be incorporated, not only to consider different pedestrian models but also to check for other targets (e.g., vehicles) and increase the robustness of the system.
- A DL algorithm-based pedestrian detection has overcome many issues in pedestrian detection, however, these are very slow, and interpretability is very low. As such, the major issue in pedestrian detection is speed and accuracy. Future research may focus on improving the speed of computation and accuracy in detection.

9. Conclusions

This survey paper surveyed DL approaches for pedestrian detection in autonomous vehicles. We first studied pedestrian detection under some critical circumstances including occlusion, low-quality images, detection of light illumination, small- and large-sized obstacle detection by the grip of multi-spectral pedestrian detection. After that, we discussed the framework of pedestrian detection and the importance of pedestrian sensors. Then, we presented an overview of traditional approaches and DL approaches for pedestrian detection. We found that DL has accommodated the more effective techniques for pedestrian tracking as compared to traditional approaches. In addition, we analyzed the pedestrian key issues and challenges using DL approaches which mainly include YOLOv3, YOLOv4, faster R-CNN, and MobileNet-SSD models as well as outlined the best solution based on the best performance. Faster R-CNN estimates better results as compared to MobileNet-SSD regarding evaluation metrics under the low-attribute images. Furthermore, we negotiated that multi-spectral images are the foremost solution for the detection of small- or large-sized pedestrians with the addition of the YOLO version under different lighting conditions. Finally, we presented a useful discussion along with some future research directions.

Author Contributions: All authors have equally contributed. All authors have read and agreed to the published version of the manuscript.

Funding: The studies at St. Petersburg State University of Telecommunications. prof. M.A. Bonch-Bruevich was supported by the Ministry of Science and High Education of the Russian Federation by grant 075-15-2022-1137.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liem, M.C.; Gavrila, D.M. Joint multi-person detection and tracking from overlapping cameras. *Comput. Vis. Image Underst.* **2014**, *128*, 36–50. [CrossRef]
2. Cao, X.; Guo, S.; Lin, J.; Zhang, W.; Liao, M. Online tracking of ants based on deep association metrics: Method, dataset and evaluation. *Pattern Recognit.* **2020**, *103*, 107233. [CrossRef]
3. Zhang, Y.; Jin, Y.; Chen, J.; Kan, S.; Cen, Y.; Cao, Q. PGAN: Part-based nondirect coupling embedded GAN for person reidentification. *IEEE Multimed.* **2020**, *27*, 23–33. [CrossRef]
4. Han, C.; Ye, J.; Zhong, Y.; Tan, X.; Zhang, C.; Gao, C.; Sang, N. Re-id driven localization refinement for person search. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 9814–9823.
5. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I.
6. Antonio, J.A.; Romero, M. Pedestrians' Detection Methods in Video Images: A Literature Review. In Proceedings of the 2018 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 12–14 December 2018; pp. 354–360.
7. Zhu, Y.; Yang, J.; Xieg, X.; Wang, Z.; Deng, X. Long-distanceinfrared video pedestrian detection using deep learning and backgroundsubtraction. *J. Phys. Conf. Ser.* **2020**, *1682*, 012012. [CrossRef]
8. Iftikhar, S.; Asim, M.; Zhang, Z.; El-Latif, A.A.A. Advance generalization technique through 3D CNN to overcome the false positives pedestrian in autonomous vehicles. *Telecommun. Syst.* **2022**, *80*, 545–557. [CrossRef]
9. Lan, W.; Dang, J.; Wang, Y.; Wang, S. Pedestrian detection based on YOLO network model. In Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, 5–8 August 2018; pp. 1547–1551.
10. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
11. Zhang, S.; Benenson, R.; Schiele, B. Citypersons: A diverse dataset for pedestrian detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3213–3221.
12. Stutz, D.; Geiger, A. Learning 3d shape completion under weak supervision. *Int. J. Comput. Vis.* **2020**, *128*, 1162–1181. [CrossRef]
13. Neumann, L.; Karg, M.; Zhang, S.; Scharfenberger, C.; Piegert, E.; Mistr, S.; Prokofyeva, O.; Thiel, R.; Vedaldi, A.; Zisserman, A.; et al. Nightowls: A pedestrians at night dataset. In Proceedings of the Asian Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 691–705.
14. Chebrolu, K.N.R.; Kumar, P. Deep learning based pedestrian detection at all light conditions. In Proceedings of the 2019 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 4–6 April 2019; pp. 838–842.
15. Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2016**, *29*. Available online: https://proceedings.neurips.cc/paper/2016/hash/7c9d0b1f96aebd7b5eca8c3edaa19ebb-Abstract.html (accessed on 13 October 2022).
16. Zhang, X.; Wang, T.; Qi, J.; Lu, H.; Wang, G. Progressive attention guided recurrent network for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 714–722.
17. Navarro, P.J.; Fernandez, C.; Borraz, R.; Alonso, D. A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data. *Sensors* **2016**, *17*, 18. [CrossRef]
18. Divvala, S.K.; Hoiem, D.; Hays, J.H.; Efros, A.A.; Hebert, M. An empirical study of context in object detection. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 1271–1278.
19. Koo, B.; Jung, S.P. Improvement of air cathode performance in microbial fuel cells by using catalysts made by binding metal-organic framework and activated carbon through ultrasonication and solution precipitation. *Chem. Eng. J.* **2021**, *424*, 130388. [CrossRef]
20. Pawar, A.A.; Karthic, A.; Lee, S.; Pandit, S.; Jung, S.P. Microbial electrolysis cells for electromethanogenesis: Materials, configurations and operations. *Environ. Eng. Res.* **2022**, *27*, 200484. [CrossRef]
21. Zahid, M.; Savla, N.; Pandit, S.; Thakur, V.K.; Jung, S.P.; Gupta, P.K.; Prasad, R.; Marsili, E. Microbial desalination cell: Desalination through conserving energy. *Desalination* **2022**, *521*, 115381. [CrossRef]
22. Kang, H.; Kim, E.; Jung, S.P. Influence of flowrates to a reverse electro-dialysis (RED) stack on performance and electrochemistry of a microbial reverse electrodialysis cell (MRC). *Int. J. Hydrogen Energy* **2017**, *42*, 27685–27692. [CrossRef]
23. Kim, B.; Yuvaraj, N.; Sri Preethaa, K.; Santhosh, R.; Sabari, A. Enhanced pedestrian detection using optimized deep convolution neural network for smart building surveillance. *Soft Comput.* **2020**, *24*, 17081–17092. [CrossRef]
24. Chen, L.; Ma, N.; Wang, P.; Li, J.; Wang, P.; Pang, G.; Shi, X. Survey of pedestrian action recognition techniques for autonomous driving. *Tsinghua Sci. Technol.* **2020**, *25*, 458–470. [CrossRef]
25. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view convolutional neural networks for 3d shape recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 945–953.

26. Dinakaran, R.K.; Easom, P.; Bouridane, A.; Zhang, L.; Jiang, R.; Mehboob, F.; Rauf, A. Deep learning based pedestrian detection at distance in smart cities. In Proceedings of the SAI Intelligent Systems Conference, London, UK, 6–9 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 588–593.
27. Tian, Y.; Luo, P.; Wang, X.; Tang, X. Deep learning strong parts for pedestrian detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1904–1912.
28. Wang, K.; Li, G.; Chen, J.; Long, Y.; Chen, T.; Chen, L.; Xia, Q. The adaptability and challenges of autonomous vehicles to pedestrians in urban China. *Accid. Anal. Prev.* **2020**, *145*, 105692. [CrossRef]
29. Hbaieb, A.; Rezgui, J.; Chaari, L. Pedestrian detection for autonomous driving within cooperative communication system. In Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 15–18 April 2019; pp. 1–6.
30. Aledhari, M.; Razzak, R.; Parizi, R.M.; Srivastava, G. Multimodal machine learning for pedestrian detection. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; pp. 1–7.
31. Han, J.; Bhanu, B. Fusion of color and infrared video for moving human detection. *Pattern Recognit.* **2007**, *40*, 1771–1784. [CrossRef]
32. Socarrás, Y.; Ramos, S.; Vázquez, D.; López, A.M.; Gevers, T. Adapting pedestrian detection from synthetic to far infrared images. In Proceedings of the ICCV Workshops, Beijing, China, 21 October 2013; Volume 3.
33. Han, J.; Bhanu, B. Human activity recognition in thermal infrared imagery. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, San Diego, CA, USA, 20–25 June 2005; p. 17.
34. Li, J.; Liang, X.; Shen, S.; Xu, T.; Feng, J.; Yan, S. Scale-aware fast R-CNN for pedestrian detection. *IEEE Trans. Multimed.* **2017**, *20*, 985–996. [CrossRef]
35. Angelova, A.; Krizhevsky, A.; Vanhoucke, V.; Ogale, A.; Ferguson, D. Real-Time Pedestrian Detection with Deep Network Cascades. 2015. Available online: http://www.bmva.org/bmvc/2015/papers/paper032/index.html (accessed on 13 October 2022).
36. Hwang, S.; Park, J.; Kim, N.; Choi, Y.; So Kweon, I. Multispectral pedestrian detection: Benchmark dataset and baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1037–1045.
37. González, A.; Fang, Z.; Socarras, Y.; Serrat, J.; Vázquez, D.; Xu, J.; López, A.M. Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors* **2016**, *16*, 820. [CrossRef]
38. Wagner, J.; Fischer, V.; Herman, M.; Behnke, S. Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks. In Proceedings of the ESANN Conference, Bruges, Belgium, 27–29 April 2016; Volume 587, pp. 509–514.
39. Nguyen, D.T.; Li, W.; Ogunbona, P.O. Human detection from images and videos: A survey. *Pattern Recognit.* **2016**, *51*, 148–175. [CrossRef]
40. Ragesh, N.; Rajesh, R. Pedestrian detection in automotive safety: Understanding state-of-the-art. *IEEE Access* **2019**, *7*, 47864–47890. [CrossRef]
41. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
42. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 84–90. [CrossRef]
43. Dalal, N.; Triggs, B.; Schmid, C. Human detection using oriented histograms of flow and appearance. In Proceedings of the European Conference on Computer Vision, Graz, Austria, 7–13 May 2006; Springer: Berlin/Heidelberg, Germany, 2006; pp. 428–441.
44. Oren, M.; Papageorgiou, C.; Sinha, P.; Osuna, E.; Poggio, T. Pedestrian detection using wavelet templates. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, PR, USA, 17–19 June 1997; pp. 193–199.
45. Mu, Y.; Yan, S.; Liu, Y.; Huang, T.; Zhou, B. Discriminative local binary patterns for human detection in personal album. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
46. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Occlusion-aware R-CNN: Detecting pedestrians in a crowd. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 637–653.
47. Zhou, C.; Yuan, J. Multi-label learning of part detectors for heavily occluded pedestrian detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3486–3495.
48. Ouyang, W.; Wang, X. Joint deep learning for pedestrian detection. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 1–8 December 2013; pp. 2056–2063.
49. Choi, H.; Kim, S.; Park, K.; Sohn, K. Multi-spectral pedestrian detection based on accumulated object proposal with fully convolutional networks. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 621–626.
50. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

51. Hou, Y.L.; Song, Y.; Hao, X.; Shen, Y.; Qian, M.; Chen, H. Multispectral pedestrian detection based on deep convolutional neural networks. *Infrared Phys. Technol.* **2018**, *94*, 69–77. [CrossRef]
52. Zhang, L.; Lin, L.; Liang, X.; He, K. Is faster R-CNN doing well for pedestrian detection? In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 443–457.
53. Chen, Y.; Shin, H. Pedestrian detection at night in infrared images using an attention-guided encoder-decoder convolutional neural network. *Appl. Sci.* **2020**, *10*, 809. [CrossRef]
54. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
55. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
56. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 391–407.
57. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef]
58. Braun, M.; Krebs, S.; Flohr, F.; Gavrila, D.M. The eurocity persons dataset: A novel benchmark for object detection. *arXiv* **2018**, arXiv:1805.07193.
59. Ess, A.; Leibe, B.; Van Gool, L. Depth and appearance for mobile scene analysis. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio De Janeiro, Brazil, 14–21 October 2007; pp. 1–8.
60. Dollar, P.; Wojek, C.; Schiele, B.; Perona, P. Pedestrian detection: An evaluation of the state of the art. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *34*, 743–761. [CrossRef]
61. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? the kitti vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 3354–3361.
62. Wang, S.; Cheng, J.; Liu, H.; Wang, F.; Zhou, H. Pedestrian detection via body part semantic and contextual information with DNN. *IEEE Trans. Multimed.* **2018**, *20*, 3148–3159. [CrossRef]
63. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. Available online: https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html (accessed on 13 October 2022). [CrossRef]
64. Xiang, Y.; Choi, W.; Lin, Y.; Savarese, S. Subcategory-aware convolutional neural networks for object proposals and detection. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 924–933.
65. You, M.; Zhang, Y.; Shen, C.; Zhang, X. An extended filtered channel framework for pedestrian detection. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 1640–1651. [CrossRef]
66. Kingma, D.; Ba, J. Dp kingma and j. ba, adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
67. Najila, A.L.; Shijin Knox, G.U. *A Study on Automatic Pedestrian Detection Using Computer Vision*; IEEE: New York City, NY, USA, 2021; Volume 8, pp. 4553–4557.
68. Tsai, C.Y.; Su, Y.K. MobileNet-JDE: A lightweight multi-object tracking model for embedded systems. *Multimed. Tools Appl.* **2022**, *81*, 9915–9937. [CrossRef]
69. Hasan, I.; Liao, S.; Li, J.; Akram, S.U.; Shao, L. Generalizable pedestrian detection: The elephant in the room. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 11328–11337.
70. Wang, L.; Shi, J.; Song, G.; Shen, I.f. Object detection combining recognition and segmentation. In Proceedings of the Asian Conference on Computer Vision, Tokyo, Japan, 18–22 November 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 189–199.
71. Zheng, A.; Zhang, Y.; Zhang, X.; Qi, X.; Sun, J. Progressive End-to-End Object Detection in Crowded Scenes. *arXiv* **2022**, arXiv:2203.07669.
72. Ding, M.; Zhang, S.; Yang, J. Improving Pedestrian Detection from a Long-tailed Domain Perspective. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 20–24 October 2021; pp. 2918–2926.
73. Gilroy, S.; Glavin, M.; Jones, E.; Mullins, D. Pedestrian Occlusion Level Classification using Keypoint Detection and 2D Body Surface Area Estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3833–3839.
74. Sun, H.; Zhang, W.; Runxiang, Y.; Zhang, Y. Motion planning for mobile Robots–focusing on deep reinforcement learning: A systematic Review. *IEEE Access* **2021**, *9*, 69061–69081. [CrossRef]
75. Bigas, M.; Cabruja, E.; Forest, J.; Salvi, J. Review of CMOS image sensors. *Microelectron. J.* **2006**, *37*, 433–451. [CrossRef]
76. Pinggera, P.; Pfeiffer, D.; Franke, U.; Mester, R. Know your limits: Accuracy of long range stereoscopic object measurements in practice. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 96–111.
77. Fleming, W.J. New automotive sensors—A review. *IEEE Sens. J.* **2008**, *8*, 1900–1921. [CrossRef]
78. Hurney, P.; Waldron, P.; Morgan, F.; Jones, E.; Glavin, M. Review of pedestrian detection techniques in automotive far-infrared video. *IET Intell. Transp. Syst.* **2015**, *9*, 824–832. [CrossRef]

79. Carullo, A.; Parvis, M. An ultrasonic sensor for distance measurement in automotive applications. *IEEE Sens. J.* **2001**, *1*, 143. [CrossRef]
80. Schlegl, T.; Bretterklieber, T.; Neumayer, M.; Zangl, H. Combined capacitive and ultrasonic distance measurement for automotive applications. *IEEE Sens. J.* **2011**, *11*, 2636–2642. [CrossRef]
81. Zhou, J.; Shi, J. RFID localization algorithms and applications—A review. *J. Intell. Manuf.* **2009**, *20*, 695. [CrossRef]
82. Fernandez-Llorca, D.; Minguez, R.Q.; Alonso, I.P.; Lopez, C.F.; Daza, I.G.; Sotelo, M.Á.; Cordero, C.A. Assistive intelligent transportation systems: The need for user localization and anonymous disability identification. *IEEE Intell. Transp. Syst. Mag.* **2017**, *9*, 25–40. [CrossRef]
83. Zhao, F.; Jiang, H.; Liu, Z. Recent development of automotive LiDAR technology, industry and trends. In Proceedings of the Eleventh International Conference on Digital Image Processing (ICDIP 2019), Guangzhou, China, 10–13 May 2019; Volume 11179, p. 111794A.
84. Schalling, F.; Ljungberg, S.; Mohan, N. Benchmarking lidar sensors for development and evaluation of automotive perception. In Proceedings of the 2019 4th International Conference and Workshops on Recent Advances and Innovations in Engineering (ICRAIE), Kedah, Malaysia, 28–29 November 2019; pp. 1–6.
85. de Ponte Müller, F. Survey on ranging sensors and cooperative techniques for relative positioning of vehicles. *Sensors* **2017**, *17*, 271. [CrossRef] [PubMed]
86. Ohguchi, K.; Shono, M.; Kishida, M. 79 GHz band ultra-wideband automotive radar. *Fujitsu Ten Tech. J.* **2013**, *39*, 9–14.
87. Hasch, J.; Topak, E.; Schnabel, R.; Zwick, T.; Weigel, R.; Waldschmidt, C. Millimeter-wave technology for automotive radar sensors in the 77 GHz frequency band. *IEEE Trans. Microw. Theory Tech.* **2012**, *60*, 845–860. [CrossRef]
88. Gresham, I.; Jenkins, A.; Egri, R.; Eswarappa, C.; Kinayman, N.; Jain, N.; Anderson, R.; Kolak, F.; Wohlert, R.; Bawell, S.P.; et al. Ultra-wideband radar sensors for short-range vehicular applications. *IEEE Trans. Microw. Theory Tech.* **2004**, *52*, 2105–2122. [CrossRef]
89. Kuutti, S.; Fallah, S.; Katsaros, K.; Dianati, M.; Mccullough, F.; Mouzakitis, A. A survey of the state-of-the-art localization techniques and their potentials for autonomous vehicle applications. *IEEE Internet Things J.* **2018**, *5*, 829–846. [CrossRef]
90. Van Brummelen, J.; O'Brien, M.; Gruyer, D.; Najjaran, H. Autonomous vehicle perception: The technology of today and tomorrow. *Transp. Res. Part C Emerg. Technol.* **2018**, *89*, 384–406. [CrossRef]
91. Altay, F.; Velipasalar, S. The Use of Thermal Cameras for Pedestrian Detection. *IEEE Sens. J.* **2022**. [CrossRef]
92. Jabłoński, P.; Iwaniec, J.; Zabierowski, W. Comparison of pedestrian detectors for LiDAR sensor trained on custom synthetic, real and mixed datasets. *Sensors* **2022**, *22*, 7014. [CrossRef] [PubMed]
93. Bakheet, S.; Al-Hamadi, A. A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve Bayesian classification. *Brain Sci.* **2021**, *11*, 240. [CrossRef]
94. Buongiorno, D.; Cascarano, G.D.; De Feudis, I.; Brunetti, A.; Carnimeo, L.; Dimauro, G.; Bevilacqua, V. Deep learning for processing electromyographic signals: A taxonomy-based survey. *Neurocomputing* **2021**, *452*, 549–565. [CrossRef]
95. Zhang, L.; Yuan, M.; Zheng, D.; Li, X.Y. M&M: Recognizing Multiple Co-evolving Activities from Multi-Source Videos. In Proceedings of the 2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS), Pafos, Cyprus, 14–16 July 2021; pp. 75–82.
96. Asim, M.; Wang, Y.; Wang, K.; Huang, P.Q. A Review on Computational Intelligence Techniques in Cloud and Edge Computing. *IEEE Trans. Emerg. Top. Comput. Intell.* **2020**, *4*, 742–763. [CrossRef]
97. Sighencea, B.I.; Stanciu, R.I.; Căleanu, C.D. A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction. *Sensors* **2021**, *21*, 7543. [CrossRef] [PubMed]
98. Weina, Z.; Lihua, S.; Zhijing, X. A Real-time Detection Method for Multi-scale Pedestrians in Complex Environment. *J. Electron. Inf. Technol.* **2021**, *43*, 2063–2070.
99. Shivappriya, S.; Priyadarsini, M.J.P.; Stateczny, A.; Puttamadappa, C.; Parameshachari, B. Cascade object detection and remote sensing object detection method based on trainable activation function. *Remote Sens.* **2021**, *13*, 200. [CrossRef]
100. Walambe, R.; Marathe, A.; Kotecha, K. Multiscale object detection from drone imagery using ensemble transfer learning. *Drones* **2021**, *5*, 66. [CrossRef]
101. Indapwar, A.; Choudhary, J.; Singh, D.P. Survey of Real-Time Object Detection for Logo Detection System. In *Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 61–72.
102. Fu, X.B.; Yue, S.L.; Pan, D.Y. Camera-based basketball scoring detection using convolutional neural network. *Int. J. Autom. Comput.* **2021**, *18*, 266–276. [CrossRef]
103. Rundo, F.; Leotta, R.; Battiato, S.; Conoci, S. Intelligent Saliency-based Deep Pedestrian Tracking System for Advanced Driving Assistance. In Proceedings of the 2021 AEIT International Conference on Electrical and Electronic Technologies for Automotive (AEIT AUTOMOTIVE), Online, 17–19 November 2021; pp. 1–6.
104. Xiao, X.; Wang, B.; Miao, L.; Li, L.; Zhou, Z.; Ma, J.; Dong, D. Infrared and visible image object detection via focused feature enhancement and cascaded semantic extension. *Remote Sens.* **2021**, *13*, 2538. [CrossRef]
105. Do, T.N.; Tran-Nguyen, M.T.; Trang, T.T.; Vo, T.T. Deep Networks for Monitoring Waterway Traffic in the Mekong Delta. In *Proceedings of the International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 315–326. Available online: https://link.springer.com/book/10.1007/978-981-16-5685-9 (accessed on 13 October 2022).

106. Chen, L.; Lin, S.; Lu, X.; Cao, D.; Wu, H.; Guo, C.; Liu, C.; Wang, F.Y. Deep neural network based vehicle and pedestrian detection for autonomous driving: A survey. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 3234–3246. [CrossRef]
107. Ozdemir, C.; Gedik, M.A.; Kaya, Y. Age Estimation from Left-Hand Radiographs with Deep Learning Methods. *Trait. Signal* **2021**, *38*, 1565–1574. [CrossRef]
108. Jia, W.; Gao, J.; Xia, W.; Zhao, Y.; Min, H.; Lu, J.T. A performance evaluation of classic convolutional neural networks for 2D and 3D palmprint and palm vein recognition. *Int. J. Autom. Comput.* **2021**, *18*, 18–44. [CrossRef]
109. Wang, I.S.; Chan, H.T.; Hsia, C.H. Finger-Vein Recognition Using a NASNet with a Cutout. In Proceedings of the 2021 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), Hualien, Taiwan, 16–19 November 2021; pp. 1–2.
110. Nagrath, P.; Jain, R.; Madan, A.; Arora, R.; Kataria, P.; Hemanth, J. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. *Sustain. Cities Soc.* **2021**, *66*, 102692. [CrossRef] [PubMed]
111. Pawlowski, P.; Piniarski, K.; Dąbrowski, A. Highly Efficient Lossless Coding for High Dynamic Range Red, Clear, Clear, Clear Image Sensors. *Sensors* **2021**, *21*, 653. [CrossRef] [PubMed]
112. Nataprawira, J.; Gu, Y.; Goncharenko, I.; Kamijo, S. Pedestrian detection using multispectral images and a deep neural network. *Sensors* **2021**, *21*, 2536. [CrossRef] [PubMed]
113. Paigwar, A.; Sierra-Gonzalez, D.; Erkent, Ö.; Laugier, C. Frustum-pointpillars: A multi-stage approach for 3d object detection using rgb camera and lidar. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Event, 11–17 October 2021; pp. 2926–2933.
114. Ding, M.; Zhang, S.; Yang, J. Learning a Dynamic High-Resolution Network for Multi-Scale Pedestrian Detection. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 9076–9082.
115. Zhang, H.; Fromont, E.; Lefèvre, S.; Avignon, B. Guided attentive feature fusion for multispectral pedestrian detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 4–8 January 2021; pp. 72–80.
116. Ding, L.; Wang, Y.; Laganière, R.; Huang, D.; Luo, X.; Zhang, H. A robust and fast multispectral pedestrian detection deep network. *Knowl.-Based Syst.* **2021**, *227*, 106990. [CrossRef]
117. Jin, Y.; Zhang, Y.; Cen, Y.; Li, Y.; Mladenovic, V.; Voronin, V. Pedestrian detection with super-resolution reconstruction for low-quality image. *Pattern Recognit.* **2021**, *115*, 107846. [CrossRef]