

Review Article

Applications of Deep Learning Techniques for Pedestrian Detection in Smart Environments: A Comprehensive Study

Fen He ¹, Paria Karami Olia ², Rozita Jamili Oskouei ³, Morteza Hosseini ⁴,
Zhihao Peng ⁵ and Touraj BaniRostam ²

¹School of Information Engineering, Guangzhou Nanyang Polytechnic College, Guangzhou, Guangdong, China

²Computer Engineering Department, Islamic Azad University, Central Tehran Branch, Tehran, Iran

³Department of Computer Science and Information Technology, Mahdishahr Branch, Islamic Azad University, Mahdishahr, Iran

⁴Department of Engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran

⁵School of Software, Dalian Neusoft University of Information, Dalian, China

Correspondence should be addressed to Rozita Jamili Oskouei; rozita2010r@gmail.com

Received 8 February 2021; Revised 1 April 2021; Accepted 24 August 2021; Published 4 October 2021

Academic Editor: Chunjia Han

Copyright © 2021 Fen He et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Intelligent transportation systems have been very well received by car companies, people, and governments around the world. The main challenge in the world of smart and self-driving cars is to identify obstacles, especially pedestrians, and take action to prevent collisions with them. Many studies in this field have been done by various researchers, but there are still many errors in the accurate detection of pedestrians in self-made cars made by different car companies, so in the research in this study, we focused on the use of deep learning techniques to identify pedestrians for the development of intelligent transportation systems and self-driving cars and pedestrian identification in smart cities, and then some of the most common deep learning techniques used by various researchers were reviewed. Finally, in this research, the challenges in each field are discovered, which can be very useful for students who are looking for an idea to do their dissertations and research in the field of smart transportation and smart cities.

1. Introduction

In recent years, intelligent transportation systems have been developed to help reduce the volume of traffic in metropolitan areas, reduce the rate of accidents and injuries and deaths caused by them, reduce fuel consumption, reduce environmental pollution, and so on. These systems use different technologies (including IoT, machine learning and data mining, neural networks, deep learning, and image processing) for various applications. On the other hand, large automotive and technology companies (such as Google and Tesla) are trying to produce self-driving smart cars that can provide safe travel for people when the driver is drowsy and that can save his life as well as passengers' lives, with automatic control car to prevent any accident. These vehicles must be equipped with sensors to sense the environment and identify objects close to the car and inform the driver and also have actuators to perform real-time operations when the

driver is drowsy or the driver does not pay attention to hazards to prevent accidents.

Some of the most important problems in creating and developing self-driving cars are as follows:

- (i) Lack of accurate patterns to identify pedestrians and roadblocks with very high accuracy in different roads with different light intensities and different image quality
- (ii) Error in identifying pedestrians and obstacles in the paths of self-driving cars
- (iii) People's distrust of self-driving cars
- (iv) Lack of acceptance of people for 24-hour control and monitoring of various urban and interurban roads' paths
- (v) Negative effects of dark weather and snow/ice and fog and rain on the quality and performance of

cameras installed on cars, ultimately reducing the accuracy of pedestrian detection and obstacles in these vehicles

- (vi) Lack of necessary infrastructure to implement intelligent transportation on all urban and interurban roads
- (vii) The need for high investment to implement the necessary infrastructure to implement intelligent transportation and communication systems to connect vehicles to each other (V2V) and vehicles with roadside infrastructure (V2I)

Machine learning (ML) is a field of artificial intelligence that uses statistical techniques to learn hidden patterns from existing data and to make decisions about unseen records. The main task of a machine learner is building a general model on the possible distribution of training examples and then generalizing experience to unseen examples [1]. The learning process depends on the quality of the data displayed. An example is presented in a dataset with different properties. Unfortunately, extracting efficient features can be difficult for some tasks.

Deep learning is an advanced branch of the ML discipline that aims to discover complex representations of simpler representations. Deep learning methods are usually based on artificial neural networks consisting of several hidden layers with nonlinear processing units. The word deep refers to several hidden layers used to change the display of data. Using the concept of feature learning, each hidden layer of neural networks plots its input data in a new display. The layer manages to absorb a higher level of abstraction than the abstract concept in the previous layer. In deep learning architectures, the hierarchy of features learned at multiple levels is finally mapped to the output of the ML work in a single framework. Similar to ML methods, the deep learning architecture is divided into two broad categories: (a) unsupervised learning methods and (b) supervised learning approaches, including deep neural networks.

This research is organized in five sections. In the second part, we provide a complete description of intelligent transportation systems. In the third section, we provide a brief overview of deep learning and some of its applications. In the fourth section, we describe the use of deep learning to identify pedestrians in smart cities and intelligent transportation systems, review some of the research conducted by various researchers in this field, and state the challenges in each area of research. Finally, in the fifth section, we will provide the conclusion.

2. Intelligent Transport Systems (ITS)

Intelligent transportation systems, for automatic road management and real-time operations in the event of a natural accident (such as a mountain fall, avalanche, and icy road floor) or unnatural disasters (such as accidents, road repairs, and car traffic), have been developed. These systems use sensors as a tool to identify and understand the state of the travel environment. The data collected by these sensors using communication technologies (such as WiFi and DSRC) to other vehicles on the route are sent to control centers.

In recent years, organizations responsible for transport management in different countries of the world have shown great attention to the development and use of intelligent transport systems using the creation of intercity networks. The main reason for this is the diverse applications of automotive communication technology in the four areas of safety promotion, mobility improvement, environmental protection, and asset management. Automotive intelligent communication systems provide technical and economic solutions to the transportation challenges of the 21st century. These systems must be able to enable a growing segment of the human population to move freely, without the risk of accidents and with minimal fuel consumption and environmental pollution [2].

Intelligent transportation technology ranges from basic management systems (such as car navigation, traffic signal control systems, variable driving signs, automatic license plate number recognition, or speed camera) to surveillance applications such as advanced CCTV security systems that provide information. From sources such as car park guide information systems, they collect weather information, etc. [2].

Intelligent Transportation System (ITS) means the use of a set of tools, facilities, and expertise such as traffic engineering concepts, software, hardware, and telecommunications technologies in a coordinated and integrated manner to improve the efficiency and security of transportation systems.

Intelligent transportation systems can also be generalized to different modes of transportation, in which, using automated tools and related scheduling, various types of information receiving and processing operations, as well as traffic management and control, are performed. In this system, by limiting the role of human factors in information processing or control and management processes, we improve the quality of decision-making and management processes.

In intelligent transportation systems (ITS), the definition of transportation infrastructure in addition to information and communication technologies leads to the achievement of goals such as improving passenger safety, reducing transportation time, and reducing fuel consumption and wear or tear of car tires. Applications of ITS include accident management, electronic toll collection management, public transportation management, passenger communication management, and traffic flow management [3].

Compared to traditional traffic engineering, the intelligent transportation system (ITS) has created a new transportation system. Due to different national circumstances, the development priorities of these systems are different, and therefore, the content of ITS research is not the same in all countries. In general, ITS uses information, communication, control, computer technology, and other current technologies to create a real-time, accurate, and efficient transportation management system.

2.1. Architecture of Intelligent Transportation Systems. The US Department of Transportation, through RITA Research and Innovation Technology Management, defined a

national architecture for ITS and provided a common structure for designing intelligent transportation systems. The ITS function model (Logic Architecture) provides a functional view of ITS user services. Physical architecture divides the functions defined by logical architecture into classes and subsystems. Figure 1 shows the high-level diagram of the proposed physical architecture (Architecture Development Team 2007a), in which 22 subsystems (white rectangles) are distributed among four classes: passengers, centers, vehicles, and the field or area of operation.

In Figure 1, the communication requirements between these subsystems are supported by four types of communication, which are shown in the form of an oval in the figure: wireless communication over a wide area, fixed point-to-fixed point communication, vehicle-to-vehicle communication, and dedicated short-range communications [4].

The following is a brief description of each of the classes in Figure 1:

- (i) Travelers: different services are provided to passengers (including drivers and occupants of cars), which are generally grouped into two categories, which are as follows [4]:
 - (i) Support to their remote travel: using the installation of surveillance cameras by the Road and Transportation Administration on various routes inside and outside the city, road and transport managers can monitor the movements of passengers on different routes from the headquarters and in case of any problem, whether it is an accident or a fall of a mountain, and etc., and send a team to the place immediately
 - (ii) Access to personal information: it includes monitoring of the crossing of intersections when the traffic lights are red and registering the license plate number of the offending vehicle and then issuing a fine for him
- (ii) Control centers including the centers providing necessary and useful information for drivers, traffic management centers, relief and emergency centers, transport management centers, toll collection center, collected data management centers, transport fleet control centers, and road maintenance management
- (iii) Vehicles including personal vehicles, emergency vehicles, commercial vehicles, freight vehicles, support vehicles and relief vehicles, or vehicles belonging to the police patrol
- (iv) Roadside equipment including equipment installed on the road, equipment related to toll collection, parking management, checking commercial vehicles, and checking the weight of trucks with their load, which, if it is more than a certain weight, should have stopped and in addition to fining them, their burden should also be reduced

2.2. *Some Services Provided in Intelligent Transportation Systems.* Some of the most common services provided in intelligent transportation systems are briefly described in this section [5].

2.2.1. *Accident Management.* We divide the stages of accident management into five stages (depending on the type and severity of the accident) in which one or more stages may occur simultaneously [5]:

- (i) Detection and notification: it is the detection and notification of accidents that are often used in mobile phones at this stage.
- (ii) Verification: the existence of the accident and the exact type and location of the accident (via traffic surveillance cameras) are confirmed.
- (iii) Incident site management: it is a complex process that requires careful coordination, communication, and cooperation between the people present on the scene, all supporting institutions and the general public. Important points for proper incident scene management include the following:
 - (i) Providing accurate information to the dispatch unit in a timely manner, including the exact location of the accident, the severity of the accident, and so on
 - (ii) Establishing evidence in a safe area to minimize oncoming traffic risk given the location of damaged vehicles and rescue personnel
 - (iii) Establishing a command system, especially when major events occur
 - (iv) Asking for help from cleaning companies if there is a possibility of hazardous substances at the scene
 - (v) Public information in case of an accident through mass media, Internet, or SMS to passengers
- (iv) Protection of evidence: it includes the protection of evidence at the scene of the accident and potential evidence that may later be used to prosecute the perpetrator or analyze the data in the future. Any suspicious items at the scene (such as guns, bullets, drugs, and alcohol) should also be protected to be handed over to the police on the scene or the highway warden [5].
- (v) Hazardous Materials: when hazardous materials are spilled at the scene of an accident, they must first be thoroughly inspected by police dispatchers, and then, the necessary measures must be taken to collect those materials and clean the environment.
- (vi) Breakdown and Demobilization: public mobilization to clear the scene and analyze the accident that occurs when all injured people, damaged vehicles, equipment, and debris are removed from the scene. Public mobilization is to create security, expedient, and regular departure of all those present at the

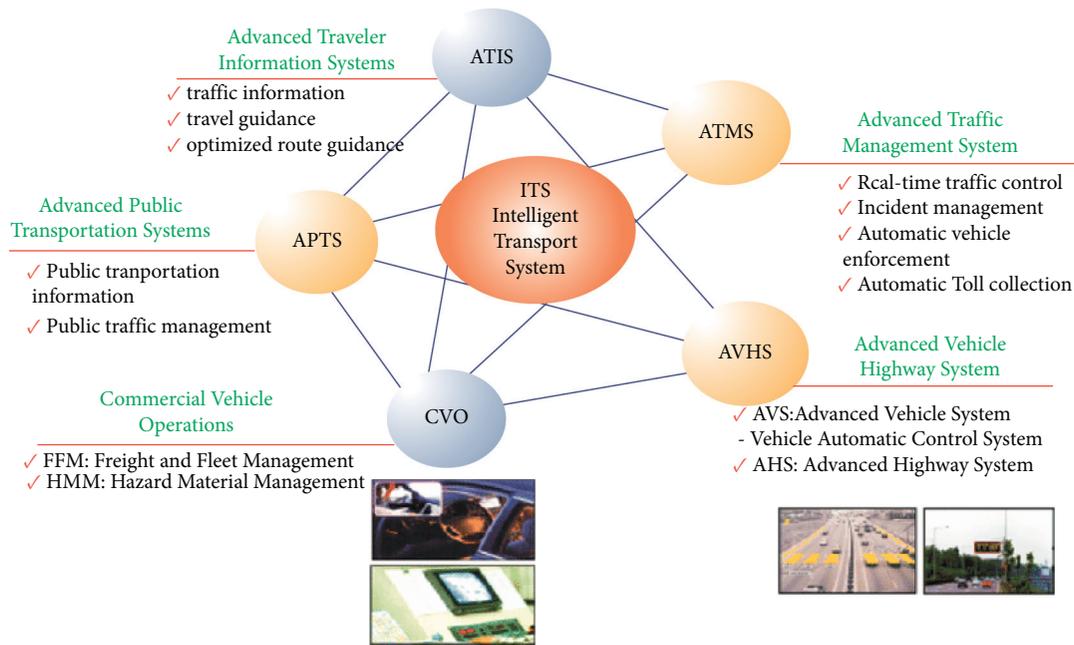


FIGURE 1: High-level architecture diagram for ITS [4].

scene of the accident and equipment and vehicles from the scene and return the affected area to normal with normal traffic flow [5].

2.2.2. APTS Public Transportation Management. These systems use new information management technologies to increase the efficiency and enhance the security of public transportation systems. These systems include instantaneous and real-time passenger information management systems, vehicle location detection systems, bus arrival time notification systems, and bus crossing priority prioritization systems.

2.2.3. Advanced ATIS Passenger Information Systems. These systems provide information on travel routes and weather conditions for transportation system users, so that they can make the right decisions to choose the route, estimate travel time, and avoid getting caught in crowded routes. Several technologies are used for this purpose, which are as follows [5]:

- (i) GPS enabled in car navigation systems
- (ii) Dynamic signs and messages for timely and real-time notification in traffic, turns and passes, and accidents or when the road is closed for various reasons such as repairs
- (iii) Websites are used to indicate congestion on highways, main streets, and urban and interurban road networks

2.2.4. Advanced ATMS Traffic Management Systems. The data and information obtained through different subsystems (such as vehicle type identifiers, in-vehicle messaging

systems, and vehicle connectors) are combined to form a cohesive interface that is capable of parsing and has real-time data analysis and decision-making about the current traffic situation and the assessment of subsequent conditions that may occur, as well as the adoption of appropriate measures to deal with the conditions that have arisen: dynamic traffic control systems, highway operations management systems, accident prevention systems, and making necessary and appropriate decisions when accidents occur, etc. They are considered as advanced traffic management systems [5].

2.2.5. Network Security Management. The main purpose of network safety management, somewhat like the management of sensitive points, means identifying the areas where accidents are most likely to occur. Therefore, there is an urgent need to ensure road safety in those areas. However, there are two important differences between hotspot management and network security management [5]:

- (i) In network safety management, the important goal is to identify roads with different degrees of security and ultimately to identify accident hotspots or sensitive points in the road system (such as intersections).
- (ii) In network safety management, a report on the severity of accidents is prepared, and accident-prone parts of the road are identified. In the management of critical or accident-prone points, the number of accidents at each critical point is usually too high, so this point is given more importance than the severity of the accident.

3. Deep Learning

Learning is the process by which a system improves its performance by using past experiences. Since 2006, deep learning has emerged as a new subfield of machine learning,

affecting a wide range of signal and information processing in both traditional and modern fields. Many traditional machine learning and signal processing techniques use special architectures that contain a single layer of nonlinear features.

Some examples of deep learning in the workplace include a self-propelled vehicle slowing down as it approaches a pedestrian crossing, an ATM rejecting a counterfeit banknote, and a smartphone app instantly translating an installed signboard performing on the street. Deep learning is especially suitable for identification programs such as face recognition, text translation, voice recognition, and advanced driver assistance systems, including and symptom recognition [6].

3.1. The Difference between Deep Learning and Machine Learning. Deep learning is one of the subfields of machine learning. By learning the machine, the features of an image can be extracted manually. With deep learning, raw images can be inserted directly into a deep neural network that learns features automatically. Deep learning usually requires hundreds of thousands or millions of images to get the best results, while machine learning works well with small datasets. Deep learning is also computationally intensive and requires a high-performance CPU [7].

Deep learning is the most effective, supervised, and cost-effective machine learning approach. Deep learning is not a limited learning method, but it follows a variety of methods and topographies that can be used to make broad predictions about complex problems. This technique includes descriptive and distinctive features in a completely categorized way. Deep learning methods with remarkable performance have achieved significant success in a wide range of applications with useful security tools. Deep learning is used in many applications, including business, comparative experiments, biological image classification, computer insight, cancer detection, natural language processing, object recognition, face recognition, handwriting, speech recognition, stock market analysis, and creation and the development of smart cities.

Machine learning is a subset of artificial intelligence (AI) that gives systems the benefits of automatically learning concepts and knowledge without explicit planning. It begins with observations such as direct experiences to prepare features and patterns in the data and to produce better results and decisions in the future. Deep learning relies on a set of machine learning algorithms that model high-level abstractions in data with multiple nonlinear transformations. Deep learning technology works on an artificial neural network (ANN) system. These neural networks continuously use learning algorithms, and by constantly increasing the amount of data, the efficiency of training processes can be improved. The efficiency of deep learning algorithms depends on the volume of large data. The process is called deep training, because the number of neural network levels increases over time.

The operation in the deep learning process generally depends on two stages called the training and inference.

- (i) The training phase involves labeling large amounts of data and determining their adaptive properties.
- (ii) The inference step is to conclude and label new and unseen data, using their prior knowledge. Deep learning is a method that helps the system understand the complex tasks of perception with maximum accuracy. Deep learning is also known as deep structured learning and is a hierarchical learning that consists of several layers that include nonlinear processing units to convert and extract features. Each subsequent layer takes the results from the previous layer as input.

The learning process is performed using the distinct stages of abstraction and multiple levels of representation in a supervised or unsupervised manner. Deep learning or deep neural network uses a basic computing unit, a neuron that receives multiple signals as input. It integrates these signals linearly with the weight and transmits the combined signals to the nonlinear tasks to produce output.

In the “deep learning” method, the term “deep” refers to the multiple layers through which data is converted. These systems are composed of a very special deep credit allocation (CAP) path, which means that the steps were performed to convert the input to output and represent the impact connection between the input layer and the output layer [7]. It should be noted that there is a difference between deep learning and machine learning. Machine learning involves a set of methods that help the machine receive raw data as input and set views for the purpose of detection and classification. Deep learning techniques are simply a type of learning method that has several levels of representation and is at a more abstract level. Figure 2 shows the difference between machine learning and deep learning.

Deep learning techniques in large databases use nonlinear transformations and high-level model abstraction. They also describe how a machine can change the internal features needed to count descriptions in each layer by accepting abstractions and displaying previous layers. This new learning approach is widely used in the areas of adaptive testing, big data, cancer diagnosis, data flow, document analysis and identification, healthcare, object recognition, speech recognition, image classification, pedestrian detection, natural language processing, and voice activity detection.

The deep learning model uses a set of features set for large features using bulk dataset for unique features, then extracts a classification model, and creates an integrated classification to explore a variety of applications.

The key factors on which the deep learning method is based are as follows [8]:

- (i) Nonlinear processing in multiple layers or stages: nonlinear processing in multiple layers refers to a hierarchical method in which the present layer accepts the results of the previous layer and transmits its output as input to the next layer. Hierarchy is created between layers to organize the importance of the data.

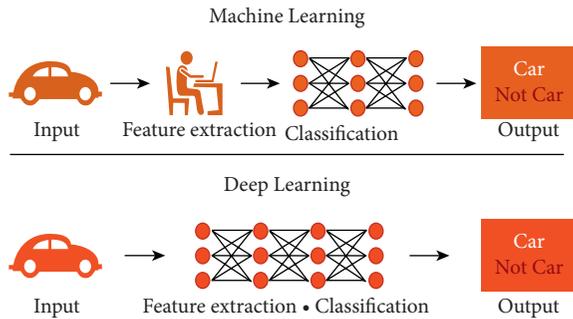


FIGURE 2: The difference between machine learning and deep learning [8].

- (ii) Supervised or unsupervised learning: here, supervised and unsupervised learning are linked to the class goal label. Its availability means a supervised system, and its absence indicates an unattended system.

4. Using Deep Learning to Diagnose Pedestrians

In today's world, where the development of smart cities and smart transportation has received a lot of attention from people, governments, and commercial and manufacturing companies, one of the basic needs is to provide solutions to identify objects around us by sensors and perform appropriate operations according to movements performed by objects. Since in this research we have mainly focused on the development of smart transportation in smart cities, so we will focus only on identifying pedestrians who are influential in the development of smart cars and smart transportation, and studies conducted by various researchers. In the field of pedestrian identification, we have divided these studies into several groups, examining the studies related to each group separately and pointing out the challenges in each.

4.1. Studies Conducted in the Field of Pedestrian Identification in Smart Cities. Belhadi et al. [9] studied the unusual behaviors of pedestrians in smart cities. For this purpose, several algorithms were proposed, which are basically divided into two categories based on performance:

- (i) Algorithms that used different data mining and knowledge discovery techniques to discover the relationship between different behaviors of pedestrians, and finally the knowledge generated to identify abnormal behaviors of pedestrians
- (ii) Algorithms that have been developed based on the history of pedestrian behaviors and based on different characteristics of the user to detect abnormal pedestrian behaviors

To implement these proposed algorithms, the researchers used the HUMBI dataset (<https://humbi-data.net/> (accessed on December 2020)), which contains 164 attributes (including gender, age, and physical condition) that include five basic body parts (including face, hands, body, clothes, and eyes), which were designed using the data in this dataset in this study.

The results of this study showed that the use of deep learning techniques in comparison with the use of data mining techniques both reduces the time of analysis and detection of normal and abnormal behaviors and increases the accuracy of identifying abnormal behaviors of pedestrians. The researchers pointed out that modeling pedestrian behaviors and behavioral analysis in the development of smart cities can help increase the efficiency of smart agents used for various applications in these cities.

Challenge. The limited features used to model pedestrian behaviors and the need to apply metaheuristic algorithms to solve complex intelligent computing problems are among the major challenges in this research to model pedestrian behaviors.

Kim et al. [10] examined pedestrian identification in smart buildings. Because the identification of pedestrians due to noise in images and some environmental factors and parameters faces challenges. The researchers used the Deep Convolution Neural Network (CNN) to create a vision-based model and the optimized version of the VGG-16, called the OVG-16, as the architectural core used to distinguish pedestrians from the multitude of possible images. To evaluate the proposed method, the researchers used the INRIA Dataset (<http://pascal.inrialpes.fr/data/human/> (accessed on December 2020)), which contained 6817 images with 3239 pedestrian images, and the image quality in this dataset was 227×227 pixels. The results of the researchers' studies showed that the proposed method has a high accuracy (approximately 98.8%) for the correct identification of pedestrians compared to other methods of machine learning.

Challenge. The model created on noisy data has not been evaluated, and there is a question: if the set of images and input data has a lot of noise, how accurate will the pedestrian be identified by this proposed model?

Using deep learning, Tomè et al. [11] proposed a system for pedestrian identification. The researchers also proposed a new framework for identifying pedestrians. The researchers also proposed new solutions for different stages of pedestrian detection, which used deep learning to easily implement their proposed algorithm on modern hardware. To implement and evaluate the proposed methods and solutions, they used the NVIDIA Jetson TK1, a GPU-based computing platform (<https://developer.nvidia.com/embedded/jetson-tk1-developer-kit> (accessed on December 2020)), and the Caltech Pedestrian dataset (http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/ (accessed on December 2020)). This dataset contains about 10 hours of video content related to vehicles collected in different weather conditions. This dataset had 250 k frames per 137 minutes of video content with 2300 different pedestrians. Half of the frames had no pedestrians, and 30% of the frames had 2 or 3 pedestrians. The results of the implementation of these researchers showed that their proposed method has high efficiency and accuracy in identifying pedestrians in real time.

Challenge. To implement these methods, we need large amounts of data, and the more data the dataset uses, the more efficient and accurate the proposed method will be. The challenge arises when data collection for various reasons (including privacy) may not be possible in the metropolitan areas of many countries in high-traffic urban areas.

4.2. Studies Conducted in the Field of Pedestrian Identification for the Development of Intelligent Transportation Systems and Self-Driving Cars. Chen et al. [12] examined existing architectures for pedestrian detection when using the automated driving method. These researchers first explained the need to use methods to identify a pedestrian and determine his or her route and then discussed the process of identifying a pedestrian while driving a car. They, then, discussed how to use deep learning techniques (such as R-CNN, SVM) to discover two-step and one-step patterns and test the effectiveness of the patterns discovered to identify pedestrians. Finally, the researchers examined and compared methods proposed by other researchers to identify pedestrians. They also introduced several datasets (such as KTH, the UCF series, Hollywood2, and Google AVA) that are used to examine proposed methods for detecting pedestrian movement.

Challenges. In this research, several important challenges in identifying pedestrians are mentioned, which are as follows:

- (i) The complexity of the environment around the pedestrian can overshadow the operations and methods of recognizing the pedestrian and his movement and, as a result, make it difficult to accurately identify the pedestrian. Therefore, creating methods to identify different perspectives on pedestrian detection and operations performed by him is one of the challenges mentioned in this research.
- (ii) Pedestrian coverage can be extremely effective in the process of identifying him/her. If the images are taken from one perspective, this can affect the accuracy of pedestrian identification and reduce the accuracy of identification. Therefore, it is necessary for researchers to propose new methods for preparing multidimensional images and their simultaneous study and aggregation of the results for early identification of pedestrians, especially in self-driving cars.
- (iii) At present, there is no standard for determining the operations and actions in a vehicle against various movements performed by pedestrians. Better results can be obtained from the effects of identifying pedestrians (such as monitoring the safety of passengers and the driver while traveling and managing environmental pollution in cities) by creating a classification and stating more details about driving practices in self-driving cars.

Said and Barr [13] proposed a new program using deep learning algorithm for fast and accurate pedestrian detection to provide real-time responses in driver assistance systems.

This program uses object classification and pedestrian identification and location tracking. The TensorFlow deep learning framework, Nvidia, cuDNN, and OpenCv acceleration libraries, and the Caltech dataset were used to implement, learn, and test the proposed method. This program is installed for deployment in mobile phones or Embedded Systems connected to self-driving cars in order to develop driver assistance systems.

Challenge. Real-time and accurate detection of objects (such as pedestrians) is one of the major challenges in the automotive industry to create and develop self-driving cars. With all the efforts that have been made, the percentage of pedestrian detection accuracy and the speed of detection of existing methods are not enough, so these methods are not very acceptable for applying real-time responses.

Ahmed et al. [14] first compared the methods and techniques used to diagnose pedestrians and cyclists. They stated that because of, in the detection stage, the possibility of detecting and locating objects (using deep learning techniques such as fast region-convolutional neural network (R-CNN), faster R-CNN, and single shot detector (SSD)) in images and video frames, so the detection stage can be created as a vital part in creating and developing smart applications in a self-driving vehicle. Finally, tracking results can be used to monitor and identify pedestrians or cyclists. The main purpose of this study was to investigate the existing methods for identifying cyclists. The results of their studies showed that the use of appropriate techniques (e.g., sensor fusion and intent estimation) for identifying pedestrians and cyclists can be an important step in maintaining road safety. In this research, first, the challenges in identifying and estimating the purpose and destination are presented, then a history of methods proposed by various researchers for pedestrian detection is presented, and the general steps proposed for object detection are explained. Next, the research conducted by other researchers on the use of deep learning techniques and architectures to identify pedestrians is reviewed, and then, the dataset used by various researchers to implement their proposed methods for pedestrian and cyclist detection is explained.

Challenge. Most of the existing datasets for implementing object detection techniques are focused on pedestrian detection data, and there is no dedicated dataset to implement the proposed techniques for identifying cyclists, so collecting this type of dataset in different areas is currently a challenge.

Zhu et al. [15] studied the challenges of pedestrian detection using infrared and proposed to use deep learning methods for pedestrian detection to overcome these challenges. By combining deep learning and background subtraction methods, the researchers proposed a new method for pedestrian detection. The proposed algorithm had two steps for pedestrian detection, which are as follows:

Step 1: background subtraction methods are performed to provide information between frames for the machine learning module

Step 2: refine Det equipment with a module of attention that is used to improve the accuracy of identifying pedestrians who are small in stature

In this study, a dataset consisting of infrared videos was created that was used to identify pedestrians from a distance and had good performance.

Challenge. The proposed method in this research is based on a dataset created by the researcher, and its performance is well evaluated. This is a set of video data stored by infrared. In order to prove the effectiveness of this method, it seems that it is necessary to use other datasets that have been collected from different geographical locations with different volumes of pedestrian traffic.

Bunel et al. [16] focused on remote pedestrian detection. When the pedestrian is too far from the camera, the size of the pedestrian becomes very small, so it becomes very difficult to detect. These researchers suggested a neural network-based method and convolutional neural network-based method to learn the features with an end-to-end approach to identify pedestrians who are too far away from the camera and too visible. Further, in this research, to implement the proposed method, they used Caltech Pedestrian Datasets. The results showed that the proposed method has a good performance in identifying pedestrians.

Challenge. It seems that the quality of cameras and the amount of pedestrian distance from the camera can be effective in the performance of the proposed method and the accuracy of method detection, which has not been considered in this study. Therefore, it is recommended to accurately define a standard for the best image quality and accuracy of diagnoses with different measurements and different photographs or videos.

Haghighat et al. [17] investigated the application of deep learning models in intelligent transportation systems. In the following, the advantages and disadvantages of embedded systems were discussed, and finally, the use of deep learning techniques to predict the occurrence of traffic on different road routes was examined.

Challenge. All datasets used in research conducted by different researchers have been collected using cameras installed in different areas. It is expected that, with the advances made in the production of self-driving cars and sensors used on cars or on the street floor, the volume of data collected from them will be greatly increased, so we need new techniques in deep learning to be able to analyze this data.

Yu et al. [18] proposed a system for tracking and identifying pedestrians using deep neural networks, which used a UAV and Kalman Filter forecasting method to track objects and pedestrians, and a dataset (YOLOv3) was used to implement the proposed method. To measure the efficiency of the proposed method for tracking and identifying pedestrians, accuracy and execution time and observing and identifying objects were examined. The results of experimental experiments showed that the proposed method had fewer errors in identifying pedestrians.

Challenge. The dataset used in this study had a very small number of records. Therefore, it seems that a larger dataset can be used to get better results. Also, this method should be tested to identify other objects (such as cars and cyclists), and its accuracy should be checked to identify those objects.

Dinakaran et al. [19] proposed generative adversarial networks (GANs) to create a new Cascaded Single Shot Detector (SSD) architecture for remote pedestrian detection. In this architecture, DCGAN is used to improve the image quality for remote pedestrian detection. In this proposed method, several criteria are used to identify the objects in the image. To implement the proposed method, the dataset of the Canadian Institute for Advanced Research (CIFAR) is used. The results obtained from experiments have shown that the proposed method has a high accuracy in identifying vehicles and pedestrians from a distance.

Challenge. Generative adversarial networks (GANs) can be used to remotely detect objects in smart cities. Since security in communications created in IoT-connected networks in smart cities is very important and fundamental, so we conduct research on the use of GANs in improving security in smart cities to identify vehicles and pedestrians. Immediately, it seems very necessary.

To overcome the problem of Occlusion handling, Tian et al. [20] proposed DeepParts, which consists of extensive trackers, instead of using deep learning techniques with an image detector. Some of the features of DeepParts are as follows:

- (i) First, these DeepParts can be trained with poorly labeled data
- (ii) Second, DeepParts is able to handle low IoU positive proposals that shift away from ground truth
- (iii) Third, every part detector in DeepParts is a powerful tracker that can detect a pedestrian by observing only a part of the body

To implement the proposed method in this research, Caltech and KITTI datasets were used, and its performance is compared to other detectors used for pedestrian detection.

Challenge. It is expected that, by using the combination of the results obtained from all the detectors used in the parts, the accuracy of identifying objects, especially pedestrians, will be increased. The use of deep learning techniques and other techniques on data from detectors may improve pedestrian detection accuracy.

In Navarro Lorente et al. [21], an automated sensor-based system was used in applications on self-driving vehicles to identify pedestrians. Different types of sensors are used in self-propelled vehicles, but in this study, researchers focused on the Velodyne HDL-64E LIDAR sensor. The data generated by this sensor was analyzed in three dimensions using machine vision and machine learning algorithms (such as nearest neighbor algorithm, Bayesian classification and support vector machine). A new framework called the Renault Twizy platform was proposed in this study to develop the ability of self-driving vehicles to identify pedestrians. The results of the implementation of their proposed framework showed that their selected features along with the

algorithms used and the quality of the camera can be important factors in better identifying pedestrians and motorcyclists.

Challenge. Implementing the algorithms used is time-consuming, and it is necessary to propose methods for accurate and real-time identification of pedestrians and motorcyclists.

Combs et al. [22] focused on using sensors installed on self-driving cars to reduce the number of deaths due to human-caused traffic accidents. They used the Fatality Analysis Reporting System (FARS) to track the number of human error deaths on US urban and suburban roads. The researchers hypothesized that a car was traveling on a road and had all the necessary sensors to detect a pedestrian and fully effective software to detect and analyze the movements and movements made by a pedestrian to identify that pedestrian. In addition, sensors mounted on the vehicle itself are able to receive signals from the movement of pedestrians. As a result, a model can be developed to be able to easily identify pedestrians and prevent accidents. The proposed model used data from VLC cameras, radar-based detection systems, and light amplitude detection (LiDAR). The results of their practical tests showed that, by using these facilities along with sensors installed on the car body, 90% of accidents caused by human error can be prevented, while using only VLC can reduce the accident statistics by only 30%.

Challenge. There is high cost of using sensors, VLC. LiDAR and car-based radar detection systems prevent automakers from using all of them to prevent rising car prices, or from using low-quality cameras, which could be a reason to reduce its quality and ultimately reduce the accuracy of identifying pedestrians or obstacles on the road.

Song et al. [23] proposed an algorithm for detecting pedestrians on the road. In this study, the pedestrian target area and the results of pedestrian detection on the road by combining the algorithm most similar to the neighbor and the least energy algorithm were accurately divided. In this study, all objects that are around a car and can generate traffic for pedestrian identification (such as cyclists, trees, other cars, and buildings around cars) are divided. And then, an algorithm was proposed for the environmental coverage of the road. The researchers used several experiments to evaluate the performance of the traffic-generating object classification detection system proposed in this study. They selected and examined several sequences of images, including different road scenes, different weather conditions, and different city streets.

Challenge. The time required to identify each of the obstacles (especially pedestrians) is variable, but in changing weather conditions and different road conditions, the time to identify obstacles and pedestrians can be increased or decreased. Therefore, we need to optimize the proposed algorithm in this research to realize response time and identify pedestrians.

Hbaieb et al. [24] proposed a new method for detecting the presence of pedestrians in the path of self-driving cars through an intervehicle communication system. In this method, descriptor (HOG), support vector machine classification (SVM), pedestrian tracker, and feature-based

cascade classification were used to achieve vehicle detection. The performance evaluation results of the proposed method were about 90% for pedestrians and 88% for vehicle detection.

4.3. Proposed Methods for Pedestrian Detection Using Different Techniques. Cai et al. [25], to solve the problems caused by resizing objects at the accuracy and speed of object identification, provide a deep unified neural network, representing the multicast CNN (MS-CNN), for the rapid detection of multifunctional objects. They gave MS-CNN including a proposed subnet and an identification subnet. The proposed subnet has several output layers, in which objects are detected at different scales. The detection subnet uses tracking methods for multipurpose object monitoring. The proposed method was implemented on the KITTI and Caltech datasets, and the results showed that the proposed method has a very good performance in detecting objects with a maximum of 15 frames per second.

Challenge. In this research, the CNN feature approximation has been used as an alternative to input sampling. The challenge is whether other methods can be used to sample the inputs that save more memory and time for calculations. Is it possible to increase the speed of moving and replacing frames (above 15 frames per second)?

Fukui et al. [26] used complex neural network-based (CNN) methods that are highly accurate in a variety of contexts to identify pedestrians. The researchers proposed a new method proposed in this research based on CNN and used Random Dropout and Ensemble Inference Network (EIN) for training and classification, respectively. Random Dropout selects units that have a high and variable flexibility rate for training, while, in a typical dropout, the flexibility rate is fixed. EIN creates multiple networks with different structures in well-connected layers. The researchers used the Caltech and Daimler Mono pedestrian datasets to implement their proposed method.

Challenge. The costs of real-time calculations to identify pedestrians using this proposed method are relatively high, so it is necessary to adopt methods to reduce these costs.

To achieve better performance in applying deep learning theory to pedestrian detection, Cai et al. [27] improved the performance of a poorly supervised hierarchical deep learning algorithm with two-dimensional deep belief networks. In the proposed design of this research, the weaknesses of the structure and training methods used in the various algorithms of the existing classifications are identified, and the following operations are performed to eliminate these weaknesses:

- (i) First, a network of deep one-dimensional beliefs expands to two-dimensional, allowing the image matrix to be loaded directly to preserve more information from the sample space.
- (ii) Second, a lightweight regulation term is added to performance consistent with the training goal

without the use of traditional oversight. With this reform, the main training without supervision becomes weak training under supervision.

- (iii) Third, the ability to distinguish between these extracted features is created.
- (iv) In this research, the INRIA, Daimler, and CVC datasets of Spain have been used to implement and evaluate the accuracy of the proposed method.

Challenge. Working with unstructured data with existing traditional methods faces several challenges (including challenges in the preprocessing, analysis, and grouping stages). It is necessary to adopt methods or algorithms to optimize the performance of existing methods for analyzing that data and identifying pedestrians through the results of those analyzes. It is also necessary to adopt strategies to improve the performance of algorithms and methods used to classify data and semantic information in occlusion conditions.

Saeidi and Ahmadi [28] first examined some of the DCNN-based learning methods and briefly explained the new algorithms proposed by various researchers for these methods. Next, the researchers proposed a deep architectural method and a new training method based on parallel DCNNs for pedestrian detection. The proposed method had two stages of training, which are as follows:

- (i) Learning Candidate Pedestrian Extractor Network (CPEN) Candidate for pedestrian training
- (ii) Parallel training DCNNs (PDCNNs) to teach how to identify a candidate pedestrian by identifying the body parts of that candidate pedestrian

In this study, the Caltech-USA dataset was used to implement the proposed method. The results obtained from evaluating the accuracy of the proposed method in pedestrian detection and comparing it with other methods showed that this method has a higher accuracy compared to other methods.

Challenge. Selecting features for pedestrian detection, especially in multidimensional data, is one of the most important challenges when using deep learning techniques. Deep learning techniques (such as SquaresChnFtrs, InformedHaar, and Katamari) performed poorly in selecting effective features for pedestrian detection, but deep learning techniques have recently been proposed by various researchers (e.g., CompAct-Deep [29], DeepParts [20], and TA-CNN [30]). They performed much better in selecting suitable features for pedestrian detection.

Vasconcelos et al. [31] proposed an automated method for optimizing the efficiency of the training suite by creating deformation and creating a local perspective. Using this method, human figures can be identified in the existing training set by applying monitoring scenarios. Experimental results of applying this method to datasets that included a variety of data and images (selection of 16 features from the imageNet dataset [32]) showed that if these data were entered as input to a convolutional neural network, it will be

able to identify pedestrians with high accuracy. This rich image database can be used in other detectors based on supervised learning architecture.

Challenge. Creating datasets with a number of effective features for pedestrian detection is one of the important challenges that the more datasets we use have a variety of effective features for pedestrian detection and can more accurately identify pedestrians, using the method proposed in this study.

Zeng et al. [33] first focused on in-depth collective public learning about each of the factors used to identify pedestrians using advances in creating a new deep neural network architecture. The proposed architecture in this research has the following parts:

- (i) Filtered information maps are obtained from the first convolution layer.
- (ii) From the second convolution layer, maps are obtained to identify parts of the image.
- (iii) The results obtained by identifying each part of the pedestrian body are used to track maps and work with information obtained from layers. Argument about access to 20 feature parts or parts of the pedestrian body is used to estimate the tag (for example: does a particular window have a pedestrian or not?).
- (iv) The windows are provided in dimensions (height 84 and width 28) that the dimensions of the pedestrian can be identified by 60 by 20.

In other words, the proposed method in this research has four parts for pedestrian detection, which are feature extraction, handling deformation, handling of occasions, and classification.

The proposed method and architecture were implemented using Caltech and ETH datasets, and their efficiency and accuracy in pedestrian identification were compared with the accuracy of other deep learning methods. The results show that the accuracy of the proposed method in this research is higher than that of other methods.

Challenge. To extract the effective features in high-precision pedestrian detection, we need a large dataset with a large number of features that were not available in this study; so, to ensure the accuracy of the proposed method in this study, we need to prepare a very large dataset with more features.

Tarchoun et al. [34] proposed two methods for tracking pedestrians in images taken from moving vehicles:

- (i) In the first method, the block matching algorithm and block matching features are used to identify pedestrians
- (ii) The second method uses a faster R-CNN detector to detect pedestrians

The proposed methods were implemented using the I2V-MVPD database, and the results showed that the first method was able to detect pedestrians in images obtained from moving vehicles in less time but had a higher false

positive rate compared to the second method. The second method had better accuracy and performance in pedestrian detection.

Challenge. Neither of these two methods can be used for real-time pedestrian detection applications, so more research is needed to reduce costs and time on these two methods.

Lee et al. [35] proposed a deep fusion network-based pedestrian detection method that used a single shot multibox detector (DSSD) halfway through. They use correlations between other feature maps to create new properties. In this study, deep fusion network was used to form issues related to the method of recognizing color images at night or pedestrian images in the dark. KAIST dataset was used to implement the proposed method. The results obtained from the implementation and evaluation of the results showed that the proposed method, compared to other methods, had at least 4.28% lower error rate in identifying pedestrians in the dark environments.

Challenge. Correctly identifying and exacting location of pedestrians in the dark using existing methods is still a challenge. Creating ways to connect different features and deep learning techniques can go a long way in increasing the accuracy of identifying pedestrians in the dark.

Ribeiro et al. [36] proposed a deep learning method for pedestrian detection (PD) detection in real time to solve problems related to the human-aware robot navigation problem. To achieve fast and accurate pedestrian detection efficiency, this study developed a combination of Aggregate Channel Features (ACF) detector with a deep convolutional neural network (CNN). In this method, we have tried to use CNN to increase the accuracy of pedestrian detection by trackers. To implement the proposed method and evaluate its accuracy, two sets (called corridor and Mbot) were used, which have real photos taken by the cameras (photos collected from the cameras in the internal and external sensors of the robot), and a typical robot navigation environment was used to evaluate the accuracy of the method in identifying pedestrians, and the results showed that it has sufficient speed and accuracy to be used in these environments and robot navigation applications to identify pedestrians.

Challenge. The performance of the proposed method should be evaluated on datasets collected from cameras located in different places with different light intensities and distances, different types of sensors installed in the environment such as laser sensors.

Hu et al. [37] worked to create a powerful pedestrian detector. For this purpose, the researchers designed the deep convolutional neural network (DCNN) as an image feature to teach a set of enhanced decision models, using redesigned learning algorithms (CFMs) without the use of learning algorithms. To increase the efficiency and accuracy of DCNN-based detectors for image detection of pedestrians, hand-crafted features such as optical flow are used. In this study, they reviewed various datasets that have been used by other researchers to implement their proposed methods for pedestrian detection. They used the

KITTI dataset to implement the proposed method in this research. The results of the evaluation of the proposed method showed that this method reduces the complexity of the detectors and can be more efficient in accurately identifying pedestrians.

Challenge. The proposed method in terms of time required to identify pedestrians may be associated with challenges; i.e., in terms of time, more studies should be done on this method, so that it can be used immediately to detect pedestrians in cars used.

Wagner et al. [38] explored the potential of deep learning techniques in pedestrian identification. They examined two deep fusion architectures and their performance on multi-spectral data. Finally, they used a new deep CNN-based method to detect pedestrians based on multispectral image data to analyze the proposed method. They introduced the first deep CNN application for pedestrian detection based on multispectral image data, and they used three datasets (including ImageNet [32], CALTECH benchmark [2], and KAIST) to implement and evaluate the proposed method. The evaluation results showed that the proposed method had a higher accuracy in pedestrian detection compared to other methods.

Challenge. The most important challenge in the proposed method is that, most of the time, early-fusion architecture is not able to achieve our expected performance. The reason for this may be due to the inability of the early-fusion network to learn the meaningful multistate abstract properties in a given environment.

Kim et al. [39] proposed a system with limited resources for real-world monitoring and identification of moving persons. For this purpose of combination background subtraction and convolutional neural networks (CNNs), they used it to identify and detect moving objects using outdoor CCTV videos. The background subtraction algorithm used to find the desired areas in the video frame and the CNN classifier was used to classify the ROIs obtained in one of the predefined classes. To implement the proposed method in practice, various datasets collected by several real-world CCTV cameras were used. The results showed that the proposed system had a high accuracy in identifying pedestrians and was also less complex than other methods.

Challenge. Occurrence of some problems in the collected data can reduce the performance or accuracy of pedestrian detection. For example, lack of training data may disrupt the training process. On the other hand, using the same images will cause a pedestrian to be repeatedly identified several times, and this will reduce the performance of the proposed system for pedestrian detection.

Lin et al. [40] proposed a framework for pedestrian detection that is based on incorporating pixel-wise information into deep convolutional feature maps. In this context, they used the zooming properties to improve image quality to help easily and accurately identify pedestrians. Therefore, the proposed method in this research helps

TABLE 1: Challenges and proposed solutions.

Challenges	Proposed solutions
Lack of structured data in all studied routes to identify pedestrians at different times of the day with different light intensities during the day and different weather conditions	Installation of devices in the desired routes to monitor the passage of pedestrians and bicycles around the clock and the use of various data mining techniques and deep learning to create patterns to identify cyclists and pedestrians in different weather conditions and different light intensities
Absence of any known standard for selecting appropriate and essential features from the features collected for pedestrian identification or limited features used to model pedestrian behaviors and the need to apply metaheuristic algorithms to solve complex intelligent computing problems	Using more advanced sensors in different directions or applying sensors in the car body, creating a centralized database to consolidate data collected from sources and sensors used in different places
Existence of noise in various images and data collected, poor quality or very low quality of some images collected by cameras installed on the road, especially in cloudy, rainy and icy weather or dark at night	Creating high-precision and high-quality sensors and cameras, installing high-quality cameras on sensitive routes, proposing new techniques for preprocessing data and images, and accurately detecting noise images
Lack of access to clear and uninterrupted data and images for research due to privacy	Establish protocols and standards for collecting data on public places, establish agreements and laws to protect the privacy of the public while collecting 24-hour data from all road routes
The complexity of the environment around the pedestrian can overshadow the operations and methods of identifying the pedestrian and her movement, and as a result, make it difficult to accurately identify the pedestrian	Develop methods to identify different perspectives on pedestrian detection and operations performed by his/her
Pedestrian coverage can be very effective in the process of identifying him. If the images are taken from one perspective, it can affect the accuracy of pedestrian identification and reduce the accuracy of identification	Proposing new methods by researchers to prepare multidimensional images and their simultaneous study and aggregation of the results for early identification of pedestrians, especially in cars
Real-time and accurate detection of objects (such as pedestrians) is an important challenge for car companies to create and develop self-driving cars	Propose methods to increase the accuracy of pedestrian detection and speed up the process of immediate response to avoid accidents when detecting obstacles and pedestrians
Lack of specific data to diagnose cyclists	Collect data on cyclists on different routes with different light intensities and create patterns to identify cyclists
The quality of the cameras, the amount of pedestrian distance from the camera can affect the performance of the proposed method and the accuracy of the method detection, which has not been considered in many studies	Define a standard for the best image quality and detection accuracy with different measurements and different shots or videos
The cost of performing real-time calculations to identify pedestrians using this proposed method is relatively high	It is necessary to adopt methods to reduce these costs It is necessary to adopt methods or algorithms to optimize the performance of existing methods for analyzing that data and identifying pedestrians through the results of those analyzes. It is also necessary to adopt strategies to improve the performance of algorithms and methods used to classify data and semantic information in occlusion conditions
Working with unstructured data with existing traditional methods faces several challenges (including the challenges of preprocessing, analysis, and grouping)	

identify pedestrians who are seen in a very small image by inserting geographical location specifications and pedestrian features. The proposed method uses three datasets: Caltech [41], INRIA [42], and KITTI [43]. The implementation results obtained from the evaluation and comparison with other methods showed that this method is more efficient in terms of reducing the time of pedestrian identification and the number of unidentified cases.

Challenge. Due to the small size of the pedestrian image, it seems that there are complications in recognizing of that pedestrian in an image taken from a low light environment, especially at night, using the method proposed in this research.

Dollár et al. [41] reviewed advances over the past decade in developing methods for pedestrian detection and proposing 40 trackers for pedestrian detection. They analyzed

the performance of these detectors using various datasets including Caltech. In this study, the most widely used datasets were briefly described, and the strengths and weaknesses of each were expressed. Three features (including best features, additional data, and background/conceptual information) were used to conduct practical experiments in this study, which can affect the efficiency of the proposed method for pedestrian detection. Three important and famous trackers (including deformable part models, decision forests, and deep networks) are based on the different learning techniques used.

Challenge. It seems that the most important challenge in the field of pedestrian detection is to develop a deeper understanding in selecting the best features to achieve the highest accuracy and performance in real-time pedestrian detection.

The main challenge ahead seems to develop a deeper understanding of what makes good features, so as to enable the design of even better ones.

5. Conclusion

Nowadays, the amount of data generated during a day from various sensors and other devices is enormous, and technologies such as cloud computing are being used to help store large volumes of collected data. One of the benefits of analyzing this data is the discovery of knowledge and a pattern for use in similar situations by training machines. In this study, we explained deep learning and its difference with machine learning, and then some research is done by various researchers around the use of deep learning techniques in the creation and development of smart cities and pedestrian identification in smart cities and Intelligent Transportation Systems (ITS). Finally, we examined smart transportation and listed the challenges in each of them.

In general, according to the studies conducted in this paper, the most important challenges in identifying pedestrians on the street using the proposed technologies and methods, especially deep learning techniques, can be expressed in Table 1 and some of the solutions that seem useful in solving these challenges are suggested in this table.

Data Availability

This is a survey, and we introduced already all databases in this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] R. E. Schapire and Y. Freund, *Foundations of Machine Learning*, MIT Press, Cambridge, MA, USA, 2012.
- [2] ETSI, "Intelligent transport systems (ITS); security; threat, vulnerability and risk analysis (TVRA)," European Telecommunications Standards, Sophia Antipolis, France, ETSI TR 102 893, 2010.
- [3] T. Rawal and V. Devadas, "Intelligent transportation system in India—a review," *Journal of Management Development*, vol. 2, p. 299, 2015.
- [4] P. Fernandes and U. Nunes, "Vehicle communications: a short survey," in *Proceedings of the IADIS Telecommunications, Networks and Systems*, p. 1556, Lisboa, Portugal, July 2007.
- [5] K. B. Smith, "Typologies, taxonomies, and the benefits of policy classification," *Policy Studies Journal*, vol. 30, no. 3, pp. 379–395, 2002.
- [6] T. Brosch, R. Tam, and Initiative for the Alzheimers Disease Neuroimaging, "Manifold learning of brain MRIs by deep learning," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Nagoya, Japan, September 2013.
- [7] R. Heffernan, K. Paliwal, J. Lyons et al., "Improving prediction of secondary structure, local backbone angles, and solvent accessible surface area of proteins by iterative deep learning," *Scientific Reports*, vol. 5, no. 1, p. 11476, 2015.
- [8] S. Dargan and M. Kumar, "A survey of deep learning and its applications: a new paradigm to machine learning," *Archives of Computational Methods in Engineering*, vol. 27, pp. 1071–1092, 2019.
- [9] A. Belhadi, Y. Djenouri, G. Srivastava, D. Djenouri, J. C.-W. Lin, and G. Fortino, "Deep learning for pedestrian collective behavior analysis in smart cities: a model of group trajectory outlier detection," *Information Fusion*, vol. 65, pp. 13–20, 2021.
- [10] B. Kim, N. Yuvaraj, K. R. Sri Preethaa, R. Santhosh, and A. Sabari, "Enhanced pedestrian detection using optimized deep convolution neural network for smart building surveillance," *Soft Computing*, vol. 24, no. 22, pp. 17081–17092, 2020.
- [11] D. Tomè, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi, and S. Tubaro, "Deep convolutional neural networks for pedestrian detection," *Signal Processing: Image Communication*, vol. 47, pp. 482–489, 2016.
- [12] L. Chen, N. Ma, P. Wang et al., "Survey of pedestrian action recognition techniques for autonomous driving," *Tsinghua Science and Technology*, vol. 25, no. 4, pp. 458–470, 2020.
- [13] Y. F. Said and M. Barr, "Pedestrian detection for advanced driver assistance systems using deep learning algorithms," *IJCSNS (International Journal of Computer Science and Network Security)*, vol. 19, no. 10, 2019.
- [14] S. Ahmed, M. N. Huda, S. Rajbhandari, C. Saha, M. Elshaw, and S. Kanarachos, "Pedestrian and cyclist detection and intent estimation for autonomous vehicles: a survey," *Applied Sciences*, vol. 9, no. 11, p. 2335, 2019.
- [15] Y. Zhu, J. Yang, X. Xieg, Z. Wang, and X. Deng, "Long-distanceinfrared video pedestrian detection using deep learning and backgroundsubtraction," *Journal of Physics: Conference Series*, vol. 1682, 2020.
- [16] R. Bunel, F. Davoine, and P. Xu, "Detection of pedestrians at far distance," in *Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Stockholm, Sweden, May 2016.
- [17] A. K. Haghighat, V. Ravichandra-Mouli, P. Chakraborty, Y. Esfandiari, S. Arabi, and A. Sharma, "Applications of deep learning in intelligent transportation systems," *Journal of Big Data Analytics in Transportation*, vol. 2, no. 2, pp. 115–145, 2020.
- [18] H. Yu, G. Li, W. Zhang et al., "The unmanned aerial vehicle benchmark: object detection, tracking and baseline," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1141–1159, 2020.
- [19] R. K. Dinakaran, P. Easom-McCaldin, A. Bouridane et al., "Deep learning based pedestrian detection at distance in smart cities," in *Intelligent Systems and Applications*, Springer, Cham, Germany, 2019.
- [20] Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep learning strong parts for pedestrian detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, December 2015.
- [21] P. J. Navarro Lorente, C. Fernandez, R. Borraz, and D. Alonso, "A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data," *Sensors*, vol. 17, no. 1, p. 18, 2017.
- [22] T. S. Combs, L. S. Sandt, M. P. Clamann, and N. C. McDonald, "Automated vehicles and pedestrian safety: exploring the promise and limits of pedestrian detection," *American Journal of Preventive Medicine*, vol. 56, no. 1, pp. 1–7, 2019.

- [23] Y. Song, J. Yao, Y. Ju, Y. Jiang, and K. Du, "Automatic detection and classification of road, car, and pedestrian using binocular cameras in traffic scenes with a common framework," *Complexity*, vol. 2020, Article ID 2435793, 17 pages, 2020.
- [24] A. Hbaieb, J. Rezgui, and L. Chaari, "Pedestrian detection for autonomous driving within cooperative communication system," in *Proceedings of the 2019 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, Marrakesh, Morocco, April 2019.
- [25] Z. Cai, Q. Fan, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *European Conference on Computer Vision*, Springer, Cham, Germany, 2016.
- [26] H. Fukui, T. Yamashita, Y. Yamauchi, H. Fujiyoshi, and H. Murase, "Pedestrian detection based on deep convolutional neural network with ensemble inference network," in *Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, Seoul, South Korea, July 2015.
- [27] Y. Cai, Y. He, H. Wang, X. Sun, L. Chen, and H. Jiang, "Pedestrian detection algorithm in traffic scene based on weakly supervised hierarchical deep model," *International Journal of Advanced Robotic Systems*, vol. 14, no. 1, Article ID 1729881417692311, 2016.
- [28] M. Saeidi and A. Ahmadi, "Deep learning based on parallel CNNs for pedestrian detection," *International Journal of Information and Communication Technology Research*, vol. 10, no. 4, pp. 42–52, 2018.
- [29] Z. Cai, M. Saberian, and N. Vasconcelos, "Learning complexity-aware cascades for deep pedestrian detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, December 2015.
- [30] Y. Tian, P. Luo, X. Wang, and X. Tang, "Pedestrian detection aided by deep learning semantic tasks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.
- [31] C. N. Vasconcelos, A. Paes, and A. Montenegro, "Towards deep learning invariant pedestrian detection by data enrichment," in *Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Anaheim, CA, USA, December 2016.
- [32] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "Imagenet: a large-scale hierarchical image database," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, USA, June 2009.
- [33] X. Zeng, W. Ouyang, and X. Wang, "Multi-stage contextual deep learning for pedestrian detection," in *Proceedings of the IEEE International Conference on Computer Vision*, Sydney, Australia, December 2013.
- [34] B. Tarchoun, A. Khalifa, S. Dhifallah, I. Jegham, and M. Mahjou, "Hand-crafted features vs deep learning for pedestrian detection in moving camera," *Traitement du Signal*, vol. 37, no. 2, pp. 209–216, 2020.
- [35] Y. Lee, T. D. Bui, and J. Shin, "Pedestrian detection based on deep fusion network using feature correlation," in *Proceedings of the 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, Honolulu, HI, USA, November 2018.
- [36] D. Ribeiro, R. Mateus, P. Miraldo, and J. C. Nascimento, "A real-time deep learning pedestrian detector for robot navigation," in *Proceedings of the 2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, IEEE, Coimbra, Portugal, April 2017.
- [37] Q. Hu, P. Wang, C. Shen, A. v. d. Hengel, and F. Porikli, "Pushing the limits of deep CNNs for pedestrian detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1358–1368, 2017.
- [38] J. Wagner, V. Fischer, M. Herman, and S. Behnke, "Multi-spectral pedestrian detection using deep fusion convolutional neural networks," in *Proceedings of the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, April 2016.
- [39] C. Kim, J. Lee, T. Kan, and Y.-M. Kim, "A hybrid framework combining background subtraction and deep neural networks for rapid person detection," *Journal of Big Data*, vol. 5, no. 1, pp. 1–24, 2018.
- [40] C. Lin, J. Lu, G. Wang, and J. Zhou, "Graininess-aware deep feature learning for pedestrian detection," in *European Conference on Computer Vision (ECCV)*, Springer, Cham, Germany, 2018.
- [41] P. Dollár, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: a benchmark," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009.
- [42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, IEEE, San Diego, CA, USA, June 2005.
- [43] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Providence, RI, USA, June 2012.