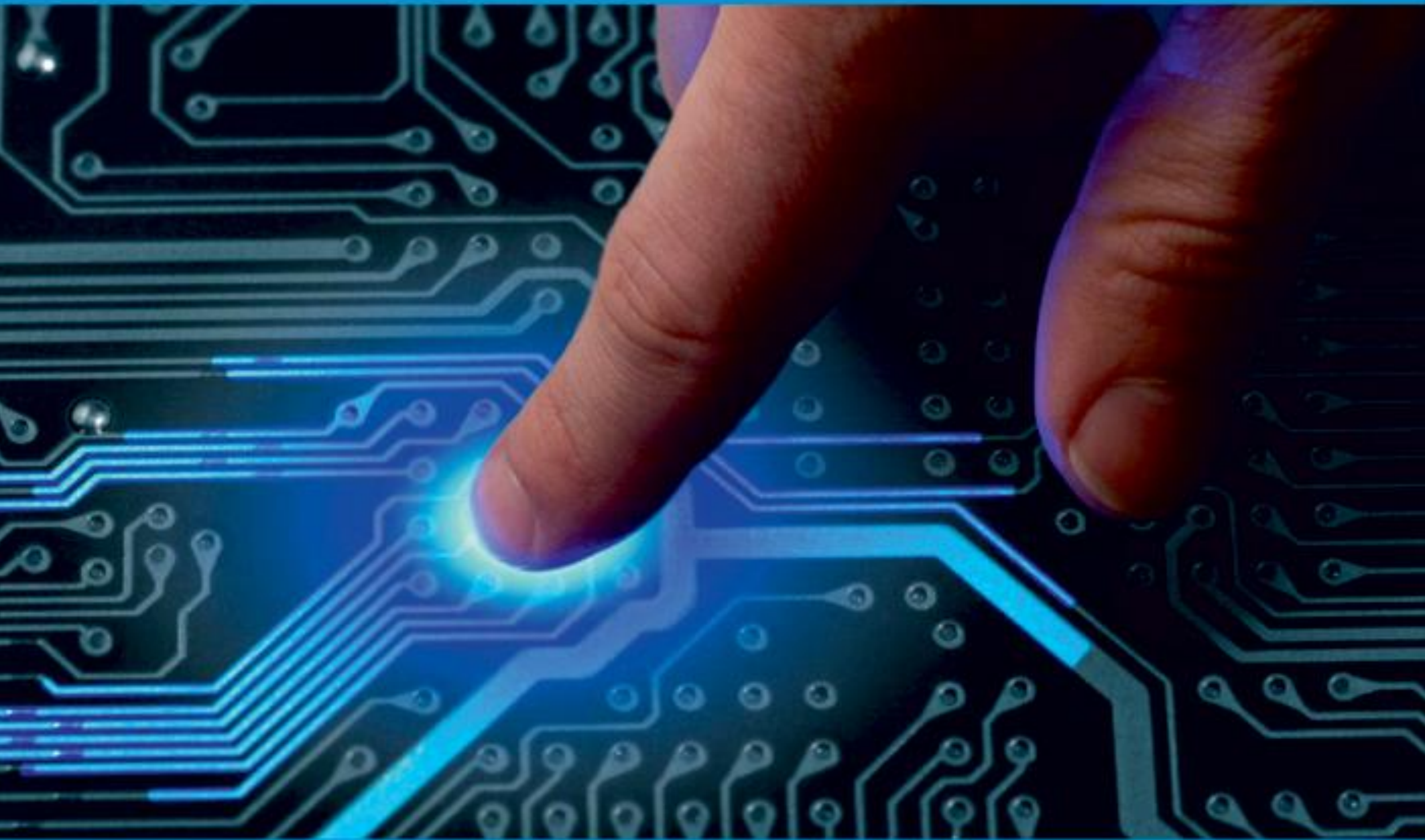
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 3, March 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

A High-Definition Ground Truth Hybrid Road Dataset

P. Shalini, Dr. V. Annapoorani

II MCA, Department of MCA, Paavai Engineering College, Paavai Nagar, Pachal, Namakkal Dist, India

Professor, Department of MCA, Paavai Engineering College, Paavai Nagar, Pachal, Namakkal Dist, India

ABSTRACT: A robust visual understanding of complex urban environments using passive optical sensors is an onerous and essential task for autonomous navigation. The problem is heavily characterized by the quality of the available dataset and the number of instances it includes. Regardless of the benchmark results of perception algorithms, a model would only be reliable and capable of enhanced decision making if the dataset covers the exact domain of the end-use case. For this purpose, in order to improve the level of instances in datasets used for the training and validation of Autonomous Vehicles (AV), Advanced Driver Assistance Systems (ADAS), and autonomous driving, and to reduce the void due to the no-existence of any datasets in the context of railway smart mobility, we introduce our multimodal hybrid dataset called ESRORAD. ESRORAD is comprised of 34 videos, 2.7 k virtual images, and 100 k real images for both road and railway scenes collected in two Normandy towns, Rouen and Le Havre. All the images are annotated with 3D bounding boxes showing at least three different classes of persons, cars, and bicycles. Crucially, our dataset is the first of its kind with uncompromised efforts on being the best in terms of large volume, abundance in annotation, and diversity in scenes. Our escorting study provides an in-depth analysis of the dataset's characteristics as well as a performance evaluation with various state-of-the-art models trained under other popular datasets, namely, KITTI and NUScenes. Some examples of image annotations and the prediction results of our 3D object detection lightweight algorithms are available in ESRORAD dataset. Finally, the dataset is available online. This repository consists of 52 datasets with their respective annotations performed.

KEYWORDS: Autonomous Vehicles, Advanced Driver Assistance Systems, Road Dataset.

I.INTRODUCTION

The reasons are from four aspects: 1) The existing methods of environment perception, e.g., detection [3], tracking and segmentation [6] of participants in traffic scenes, still produce inevitable errors in real environment; 2) The driving environment is rather complex, unpredictable, dynamic, and uncertain; 3) Deep traffic scene understanding, such as understanding the geometry/topology structure of scene, and spatio-temporal evolution of participants (pedestrian, vehicle, etc.), is studied far from sufficient, whose ultimate goal is to semantically reasoning the scene evolvement so as to provide clues for behavior decision and autonomous vehicle control. Actually, it is difficult to study because these elements are implicitly contained in the driving environment and cannot be directly observed; 4) The deployment of autonomous vehicle faces social dilemma and involves moral issue [7]. Complementary to our survey, Janai et al. [8] exhaustively reviewed the traffic participant recognition, detection and tracking, scene reconstruction, motion estimation, semantic segmentation, and many other vision-based tasks. Xue et al. [9] made an overview on autonomous vehicle systems from the perspectives of self-localization and multi-sensor fusion for obstacle detection and tracking, and emphasized vision-centered fusion of multiple sensors. studied the latest progresses on lane detection, traffic sign/light recognition in the perception of intelligent vehicles. These surveys, to a great extent, give a comprehensive and detailed investigation concerning with the first reason mentioned above.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

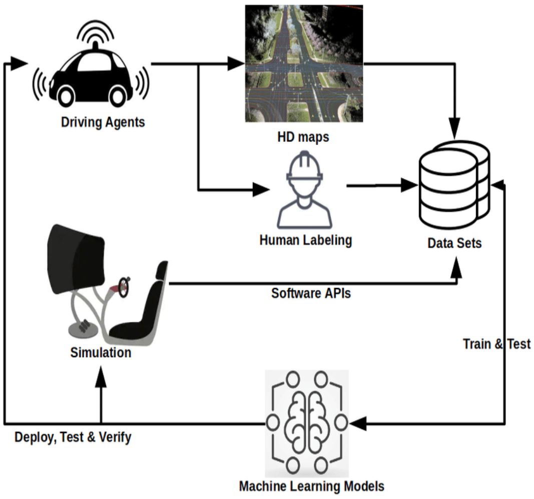


Fig 1: Road Driving Machine Learning Models

In this paper, we focus on the third aspect: survey on the deep understanding of traffic scene for autonomous vehicles. This paper aims to explore the evolution of traffic scene from an event reasoning view. That is because event can reflect the dynamic evolution process of scene with tractable reasoning strategy [11]. In order to provide a clear and logical investigation, this paper reasons the event from its representation, detection, as well as prediction stages. In the representation stage, the main goal is to obtain high-level clues for the following stages. In this stage, we expound the saliency, the contextual layout, and the topology rules for autonomous driving. As for the detection stage, we review the event detection with respect to different participants, such as pedestrian and vehicles. For the prediction stage, this paper elaborates the intention of autonomous vehicles with regard to the expected time span for future prediction. We classify the prediction of intention as long-term intention prediction and short-term prediction.

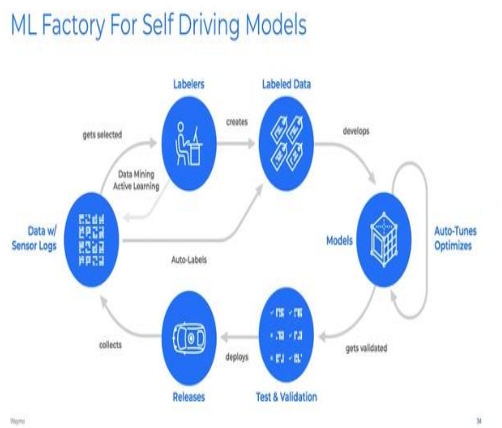


Fig 2: Self Driving model

Developing autonomous systems aim to assist humans in handing everyday tasks. Autonomous driving system, a system for closely related to humans' everyday trips, has become people's one of the most typical pursuits. It can free hands from the steering wheel, and spare time for tackling many other things. Meanwhile, the equipped sensors of autonomous vehicle can also recognize the surrounding condition immediately and ensure safe driving, thus decreasing traffic accidents. Encouraged by those merits, researchers are diligently pursuing autonomous driving all the time. There are two kinds of driving force in the development of autonomous driving. One is the projects launched and challenges posed by different governments, research institutes and vehicle manufacturers. The other we want to emphasize is the publicly available benchmarks.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

II. LITRATURE SURVEY

Defining event is a difficult problem in cognition science. What kind of scene variation should be taken as an event? Why does the event occur? We attempt to answer these questions from scene representation. Specifically, this paper focuses on the aspects of traffic saliency, content layout and topology rules for self-driving. Reasons are that: 1) Traffic saliency formulates where the scene should or may be looked when driving in different traffic situations. An event always influences and changes the attention of human drivers. In other words, traffic saliency can provide locally instantaneous clue for event reasoning. 2) Context layout specifies the relationship of traffic elements of scenes, such as geometrical layout of road scene, providing prior knowledge for the event definition. That is to say that context layout supplies globally spatial. 3) Topology rules intuitively denote the operational logic of traffic flow and the reasonable running rules with a relatively long time accumulation. Bluntly speaking, topology rules generate the spatial-temporally logical clue for event reasoning.

Replacing human drivers with autonomous control systems, however, comes at the price of creating a social interaction void. Besides being a dynamic control task, driving is a social phenomenon and requires interactions between all road users involved to ensure the flow of traffic and to guarantee the safety of others

Social interaction can play an important role in resolving various potential ambiguities in traffic. For example, if a car wants to turn at a non-signalized intersection on a heavily travelled street, it might wait for another driver's signal indicating the right of way. In the case of pedestrians, interaction can help them to understand when it is safe for them to cross the road, e.g. by receiving a signal from the driver [4]. Recent field studies of autonomous vehicles show how the lack of social understanding can result in traffic accidents [5] or erratic behaviors towards pedestrians
Given that autonomous vehicles may commute without any passengers on board, they are subject to malicious behavior, similar to those observed against a number of autonomous robots used in malls [7]. For example, some people might step in front of the autonomous vehicles to force them to change their route or interrupt their operation. Understanding the true intention of these people can help the vehicles act accordingly. A large body of studies in the field of behavioral psychology have addressed the social aspects of driving and identified numerous factors that can potentially influence the way road users behave [9]–[11]. Factors such as pedestrians' demographics [12], road conditions [11], social factors [10], and traffic characteristics [13] are shown to significantly influence pedestrian crossing decisions. However, there is a missing component in the literature, namely a holistic view of pedestrian crossing behavior to identify the extent of these factors and to explain in what ways they are interrelated.

III.METHODS

The methods of studying human behavior (in traffic scenes) have transformed during past decades as new technological advancements have emerged. Traditionally, written questionnaires or direct interviews were widely used to collect information from traffic participants or authorities monitoring the traffic. Some modern studies still rely on questionnaires especially in cases where there is a need to measure the general attitudes of people towards various aspects of driving, e.g. crossing in front of autonomous vehicles. These forms of studies, however, have been criticized for the bias people have in answering questions, the honesty of participants in responding or even how well the interviewees are able to recall a particular traffic situation.

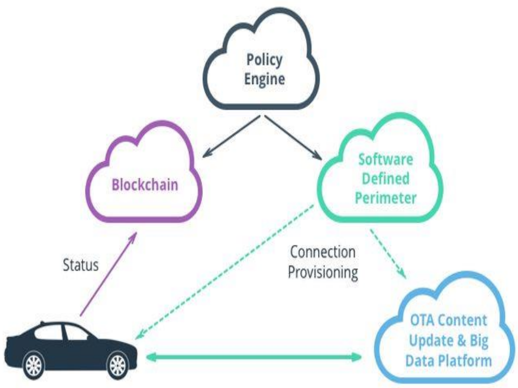


Fig 3: Autonomous Driving

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

Naturalistic recording of traffic scenes (both videos, is, perhaps, one of the most effective methods for studying traffic behavior. Although the first instances of such studies date back to almost half a century ago, they have gained tremendous popularity in recent years. In this method of study, a camera (or a network of cameras) are placed either inside the vehicles or outside on roadsides. Since the objective is to record the natural behavior of the road users, the cameras are located in inconspicuous places not visible to the observees. In the context of recording driving habits, although the presence of the camera might be known to the driver, it does not alter the driver's behavior in the long run. In fact, studies show that the presence of cameras may only influence the first 10-15 minutes of the driving, hence the beginning of each recording is usually discarded at the time of analysis [26]. An added advantage of recording compared to on-site observation is the possibility of revising the observation and using multiple observers to minimize bias.

Naturalistic recording, similar to on-site observation, may also be affected by observer bias. Moreover, in some cases, it is hard to recognize certain behaviors or underlying motives, e.g. whether a pedestrian notices the presence of the car or looks at the traffic signal in the scene and why. To remedy this issue, it is common to employ a hybrid approach where recordings or observations are combined with on-site interviews. Using this method, after recording a behavior, the researcher approaches the corresponding road user and asks questions regarding their experience, for example, whether they looked at the signal prior to crossing. Overall, the hybrid approach can help resolve the ambiguities observed in certain behaviors.

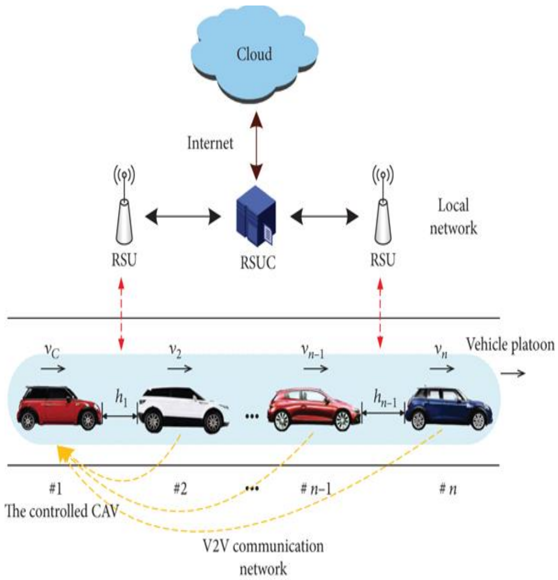


Fig 4: Road Event Analysis

Social norms even influence the way people interpret the law. For example, the concept of "psychological right of way" or "natural right of way" has been studied [21]. This concept describes the situation in which drivers want to cross a non-signalized intersection. The law requires the drivers to yield to the traffic from the right. However, in practice drivers may do quite the opposite depending on the social status (or configuration) of the street. It is found that factors such as street width, lighting conditions or the presence of shops may determine how the drivers would behave.
Result Analysis

IV.EXPERIMENTAL RESULTS AND ANALYSIS

The annotation of the images consisted in documenting them with the ground truth. The objective was to identify the different objects in the environment (vehicle, pedestrian, and cyclist) as well as their distance from the vehicle. We used the BAT 3D tool that we modified for the annotation of the images. The annotation consisted in displaying the LiDAR point clouds and manually placing 3D boxes corresponding to the class of the identified object, then displaying the result on the color/RGB image. illustrates not only some examples of image annotations but also the prediction results of our lightweight 3D object detection algorithm trained for 300 epochs on the Nuscenes.

A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams called "3D Bounding Box Annotation Tool" (BAT 3D) [58] that is a novel open access annotation system devoted to a 2D and 3D

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

dataset annotation. It is efficient, accurate, and includes tools for 3D object localization and their dynamic movement using full-surround multi-modal data streams [58]. It is open-source and cross-platform compatible for the 3D annotation. It also consists of many features that the other 3D annotation tool fails to offer. BAT 3D satisfied all of our requirements for the annotation of our dataset. It also consists of many features that the other 3D annotation tool fails to offer.

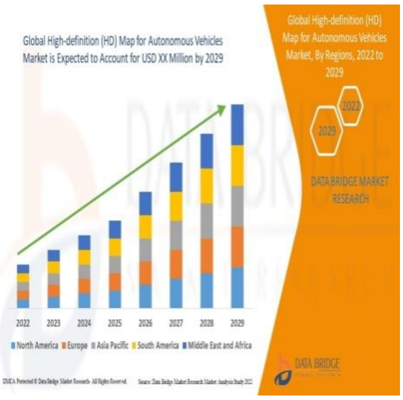


Fig 5: Road Analysis

Once the projection is validated, the post-processing part begins, i.e., the documentation of the images with the ground truth. The objective was for each image, to identify in the environment our three main classes (vehicle, pedestrian, cyclist) and also the distance of each object to the vehicle. All this information will constitute the ground truth of the real dataset that will be dedicated to the training and testing of both 2D and 3D object detection algorithms. We have therefore carried out a Benchmarking of all the annotation tools that exist in the literature. The idea is to display the LiDAR point clouds, to manually place 3D boxes corresponding to the class of the identified object, and then to display the result on the color image. This labeling step will be the last step of the dataset development.

V.CONCLUSIONS

In this paper, we presented our new Esigelec engineering high school and Segula technologies ROad and RAilway Dataset (ESRORAD). Our hybrid multimodal dataset is a large scale, first of its kind dataset, that can be institutively utilised for training and validation of models in both road and railway smart mobility applications. The hybrid nature of the dataset is achieved by including synthetic sequences developed with GTAV as a simulator and fully synchronised real images with 3D point clouds acquired using our Instrumented IRSEEM vehicle dedicated for data acquisition. In terms of quantity, our dataset includes 34 videos, 2.7 k virtual images, and 100 k real images collected in two Normandy cities, Rouen and Le Havre. The complete dataset is acquired in several fields of view in numerous road and railway traffic conditions. The images are annotated with 3D bounding boxes showing at least three different classes of objects, namely, persons, cars, and bicycles. We also provide open access to our dataset to allow researchers and companies to conduct their research works. Additionally, ESRORAD was validated with promising results, while benchmarking was conducted using our different CNN models that were trained under KITTI and NUScenes datasets. Hence, our dataset is an excellent choice for training and validation of different perception algorithms performing tasks such as 3D object detection, object tracking, semantic segmentation, scene parsing, and understanding. However, ESRORAD is extensible to many other challenging tasks. Thus, we are currently preparing other quantitative data that could enrich our dataset.

REFERENCES

1. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [**Google Scholar**] [**CrossRef**][**Green Version**]
2. Mauri, A.; Khemmar, R.; Decoux, B.; Ragot, N.; Rossi, R.; Trabelsi, R.; Boutteau, R.; Ertaud, J.Y.; Savatier, X. Deep Learning for Real-Time 3D Multi-Object Detection, Localisation, and Tracking: Application to Smart Mobility. *Sensors* **2020**, *20*, 532. [**Google Scholar**] [**CrossRef**] [**PubMed**][**Green Version**]
3. Mauri, A.; Khemmar, R.; Decoux, B.; Haddad, M.; Boutteau, R. Real-time 3D multi-object detection and localization based on deep learning for road and railway smart mobility. *J. Imaging* **2021**, *7*, 145. [**Google Scholar**] [**CrossRef**] [**PubMed**]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 4, April 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1104236 |

4. Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Duerig, T.; et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *arXiv* **2018**, arXiv:1811.00982. [**Google Scholar**] [**CrossRef**][**Green Version**]
5. 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. [**Google Scholar**]
6. Cai, Z.; Fan, Q.; Feris, R.S.; Vasconcelos, N. A Unified Multi-scale Deep Convolutional Neural Network for Fast Object Detection. *arXiv* **2016**, arXiv:cs.CV/1607.07155. [**Google Scholar**]
7. Wang, X.; Yin, W.; Kong, T.; Jiang, Y.; Li, L.; Shen, C. Task-Aware Monocular Depth Estimation for 3D Object Detection. *arXiv* **2019**, arXiv:cs.CV/1909.07701. [**Google Scholar**] [**CrossRef**]
8. Godard, C.; Mac Aodha, O.; Brostow, G.J. Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 270–279. [**Google Scholar**]
9. Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G. Digging into self-supervised monocular depth estimation. *arXiv* **2018**, arXiv:1806.01260. [**Google Scholar**]
10. Tosi, F.; Aleotti, F.; Poggi, M.; Mattoccia, S. Learning monocular depth estimation infusing traditional stereo knowledge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9799–9809. [**Google Scholar**]
11. Zhou, T.; Brown, M.; Snavely, N.; Lowe, D.G. Unsupervised learning of depth and ego-motion from video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1851–1858. [**Google Scholar**]
12. Aladem, M.; Chennupati, S.; El-Shair, Z.; Rawashdeh, S.A. A Comparative Study of Different CNN Encoders for Monocular Depth Prediction. In Proceedings of the 2019 IEEE National Aerospace and Electronics Conference (NAECON), Dayton, OH, USA, 15–19 July 2019; pp. 328–331. [**Google Scholar**]
13. Praveen, S. Efficient Depth Estimation Using Sparse Stereo-Vision with Other Perception Techniques. In *Advanced Image and Video Coding*; IntechOpen: London, UK, 2019. [**Google Scholar**]
14. Laga, H.; Jospin, L.V.; Boussaid, F.; Bennamoun, M. A Survey on Deep Learning Techniques for Stereo-based Depth Estimation. *arXiv* **2020**, arXiv:2006.02535. [**Google Scholar**] [**CrossRef**]
15. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [**Google Scholar**] [**CrossRef**]
16. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [**Google Scholar**] [**CrossRef**][**Green Version**]
17. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully Connected CRFs. *arXiv* **2016**, arXiv:1606.00915. [**Google Scholar**] [**CrossRef**][**Green Version**]
18. Mukojima, H.; Deguchi, D.; Kawanishi, Y.; Ide, I.; Murase, H.; Ukai, M.; Nagamine, N.; Nakasone, R. Moving camera background-subtraction for obstacle detection on railway tracks. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3967–3971. [**Google Scholar**]
19. Rodriguez, L.F.; Uribe, J.A.; Bonilla, J.V. Obstacle detection over rails using hough transform. In Proceedings of the 2012 XVII Symposium of Image, Signal Processing, and Artificial Vision (STSIVA), Medellin, Colombia, 12–14 September 2012; pp. 317–322. [**Google Scholar**]
20. Yanan, S.; Hui, Z.; Li, L.; Hang, Z. Rail Surface Defect Detection Method Based on YOLOv3 Deep Learning Networks. In Proceedings of the 2018 Chinese Automation Congress (CAC), Xi'an, China, 30 November–2 December 2018; pp. 1563–1568. [**Google Scholar**]

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING