Urban Digital Twins for Intelligent Road Inspection

Rui Fan[1], Yikang Zhang[1], Sicen Guo[1], Jiahang Li[1], Yi Feng[1], Shuai Su[1], Yanting Zhang[2], Wenshuo Wang[3], Yu Jiang[4], Mohammud Junaid Bocus[5], Xingyi Zhu[6], Qijun Chen[1]

***Abstract*— Urban digital twin (UDT) technologies offer new opportunities for intelligent road inspection (IRI). This paper first reviews the state-of-the-art algorithms used in the two key components of UDT-based IRI systems: (1) multi-temporal, multi-dimension, multi-score, and heterogeneous road data acquisition, and (2) road distress detection. This paper then summarizes the UDTIRI competition, organized in conjunction with IEEE Bigdata 2022. More details on our competition are available at `sites.google.com/view/udtiri-workshop/bigdata-2022`.**

I. INTRODUCTION

The concept and model of *digital twin* (DT) were first introduced in 2002 by Grieves at a Society of Manufacturing Engineers conference [1]. Grieves's DT architecture has three dimensions: physical entity, virtual model, and their communication connection [2]. A five dimensional (5D) DT architecture introduced by Tao *et al.* [3] has been universally adopted since 2018. As shown in Fig. 1, besides the aforementioned three dimensions, Tao's DT architecture adds services (*e.g.*, detection and prediction) and DT data (physical entity data, virtual model data, service data, and domain knowledge) [4].

Urban digital twins (UDTs) – defined as the application of DT technology to cities, are 3D virtual replicas of the real-world environment, connected to massive amounts of data and software tools [5]. UDT technologies are expected to offer new opportunities for current transportation infrastructures and systems in terms of their evaluation and maintenance and will make such systems intelligent and self-sustaining in the future [6].

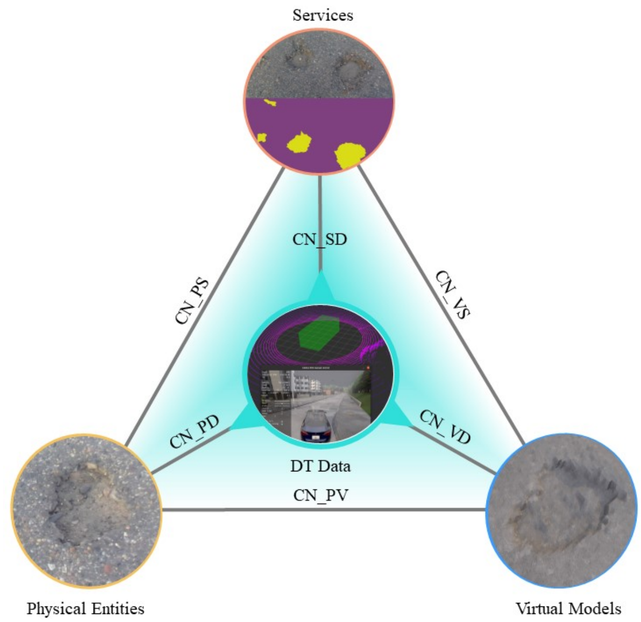

Fig. 1. An illustration of 5D urban digital twin architecture for intelligent road inspection. "CN_PS", "CN_SD", "CN_VS", "CN_PD", "CN_PV", and "CN_VD" represent communication connections.

Road inspection is becoming increasingly crucial due to a drastic increase in the number of vehicles and consequently road usage [7]. The success of a road transport system is inherently dependent on the riding quality and comfort level of the users, for which timely detection of road distresses and ensuing maintenance are of utmost importance [8]. Traditional manual road inspection is cumbersome, time-consuming, and expensive [9]. To warrant long-standing structural integrity and safety levels, *intelligent road inspection* (IRI) systems are becoming increasingly demanding [10]. IRI systems need to integrate innovative technologies that will employ next-generation distributed sensors as well as artificial intelligence and big data algorithms to help in the evaluation, classification, and localization of road distresses in a timely and cost-effective manner [11].

A UDT-based IRI system should be able to offer two main functionalities [4], [10]:

1) Multi-temporal, multi-dimension, multi-score, and het-

This work was supported by the National Key R&D Program of China under Grant 2020AAA0108100 (awarded to Prof. Qijun Chen), the National Natural Science Foundation of China under Grant 62233013 (awarded to Prof. Qijun Chen), the Science & Technology Commission of Shanghai Municipal under Grant 22511104500 (awarded to Prof. Rui Fan), the Fundamental Research Funds for the Central Universities under Grants 22120220184, 22120220214, and 2022-5-YB-08 (awarded to Prof. Rui Fan), the National Natural Science Foundation of China under Grants 51922079 and 61911530160, the Shanghai Municipal Science & Technology Major Project (2021SHZDZX0100), and the Fund of Science & Technology Commission of Shanghai Municipality (20DZ2251900).

[1]R. Fan, Y. Zhang, S. Guo, J. Li, Y. Feng, S. Su, and Q. Chen are with the Robotics & Artificial Intelligence Laboratory (RAIL), the Department of Control Science & Engineering, Frontiers Science Center for Intelligent Autonomous Systems, and State Key Laboratory of Intelligent Autonomous Systems, Tongji University, Shanghai 201804, P. R. China. Emails: {rfan, 22666078, guosicen, 2230745, 2230744, sushuai, qjchen}@tongji.edu.cn

[2]Y. Zhang is with the Department of Computer Science & Technology, Donghua University, Shanghai 201620, P. R. China. Email: ytzhang@dhu.edu.cn

[3]W. Wang is with the Department of Civil Engineering, McGill University, Montreal, QC, H3A 0C3, Canada. Email: wenshuo.wang@mcgill.ca

[4]Y. Jiang is with the CTO office at ClearMotion, Billerica MA, 01821, USA. Email: yjiang@clearmotion.com

[5]M. J. Bocus is with the Department of Electrical & Electronic Engineering, the University of Bristol, BS8 1UB, Bristol, U.K. Email: junaid.bocus@bristol.ac.uk

[6]X. Zhu is with the Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 200092, P. R. China. Email: zhuxingyi66@tongji.edu.cn

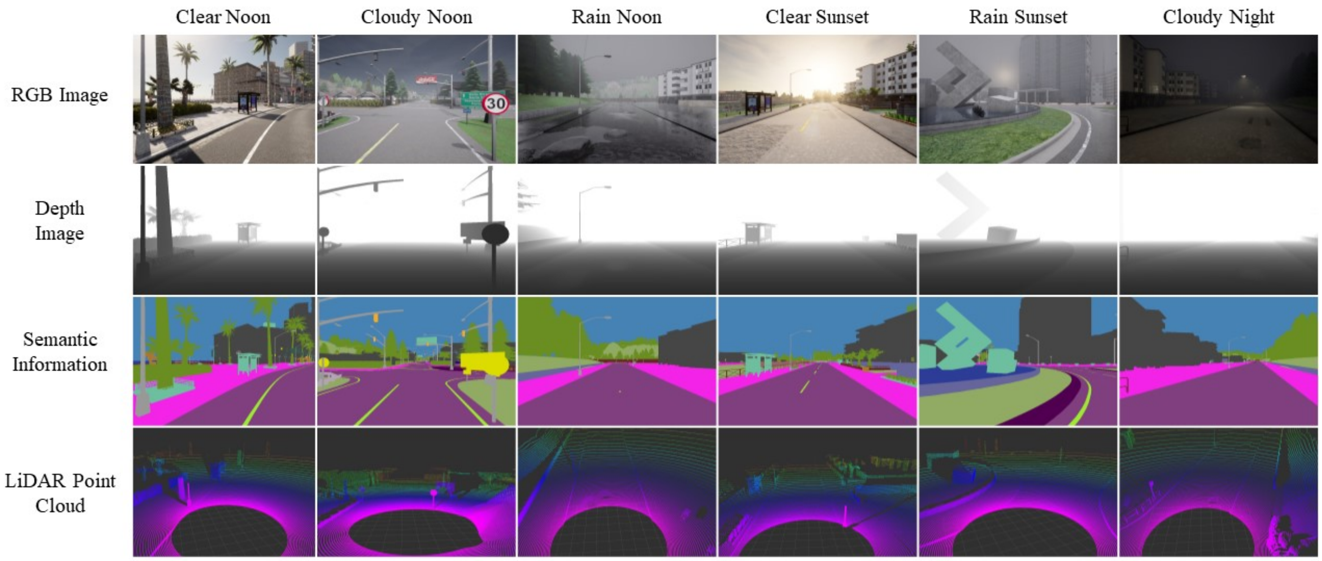

Fig. 2. Examples of our generated road data using the CARLA simulator. The 3rd, 4th, and last columns show the road data with road distresses.

erogeneous road data acquisition (3MH-RDA);

2) Road distress detection (RDD).

3MH-RDA is generally achieved using laser scanners [12], [13], Microsoft Kinect sensors [14], [15], digital cameras [7], [16], LiDARs [17], [18], vibration sensors [19], [20]. RDD is typically realized using neural networks that perform classification or regression [21]–[23]. This paper briefly reviews the state-of-the-art (SoTA) algorithms employed in current UDT-based IRI systems and introduces our Urban Digital Twins for Intelligent Road Inspection (**UDTIRI**) Competition[1], organized in conjunction with the 2022 IEEE International Conference[2] on Big Data (**IEEE BigData 2022**).

II. Multi-Temporal, Multi-Dimension, Multi-Score, and Heterogeneous Road Data Acquisition

Over the past two decades, various sensing technologies, *e.g.*, vibration sensing, active sensing, and passive sensing, have been prevalently employed for road data acquisition.

[24] presents a crowd-sourcing system for road distress detection and localization via the analysis of the accelerometer data obtained from multiple vehicles. Vibration sensors are cost-effective and only require a small storage space [16]. Nonetheless, the road distress shape and volume cannot be explicitly inferred from the vibration sensor data [25]. Furthermore, road construction joints are often mistaken for road distresses. Hence, researchers have focused on developing IRI systems based on active and passive sensing. [13] introduces the Georgia Institute of Technology sensing vehicle equipped with two laser scanners for 3D road data collection. However, such vehicles are not widely used due to their hefty price tag and long-term maintenance costs [16].

The most commonly used passive sensors for 3D road data acquisition include Microsoft Kinect and other types of digital cameras [7]. In [14], a Kinect sensor was used to acquire road depth information. However, the Kinect sensor was not designed for outdoor use and often fails to perform well when exposed to direct sunlight [26]. Therefore, it is more effective to acquire road data using digital cameras, as they are cost-effective and work well in outdoor environments [16]. To extrapolate the 3D information from a given driving scene, images from multiple views are required [27]. These images can be captured using either a single moving camera or a collection of synchronized cameras [25]. The former is typically known as structure from motion [28] or optical flow [29], while the latter is generally referred to as stereo vision (in case two cameras are used) [25].

Acquiring the aforementioned road data is, nevertheless, challenging, because accurate multi-sensor time synchronization has to be conducted [30]. Moreover, the collected road data must be human-annotated before being used to develop IRI algorithms. This process is always exhausting and costly. Therefore, developing IRI algorithms on synthetic datasets that reflect real-world environments has become popular in recent years. CARLA[3] [31] is a commonly used open-source client-server simulator for synthetic autonomous driving data generation. Based on Unreal Engine 4[4], CARLA can provide various game services, including dynamic weather, collisions, traffic rules, agent and pedestrian simulations. Furthermore, we can import user-defined high-definition maps and easily obtain their corresponding semantic information, as OpenDRIVE[5] standard is supported.

However, for UDT-based IRI applications, we have to translate real-world data into models that the simulator can

[1] https://sites.google.com/view/udtiri-workshop/bigdata-2022

[2] https://bigdataieee.org/BigData2022.

[3] https://carla.org/

[4] https://www.unrealengine.com

[5] https://www.opendrive.com/

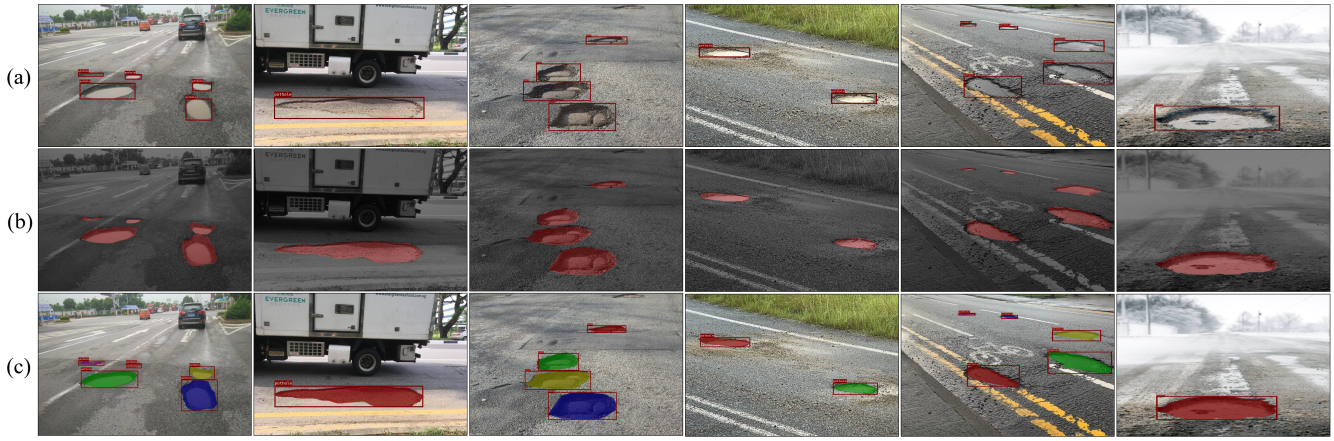

Fig. 3. Examples of the UDTIRI dataset: (a) object detection annotations; (b) semantic segmentation annotations; (c) instance segmentation annotations.

interact with. In our experiments, we first collect 3D road point clouds from the real world and build their corresponding meshes. The collision models can then be generated using Blender[6] and simulated in CARLA. Some examples of our generated road data (some of which include road distresses) are shown in Fig. 3.

III. Road Distress Detection

Road distress detection is generally achieved with deep learning algorithms. Our recent publication [8] provides a taxonomy of such algorithms. Deep learning-based algorithms address the road distress detection problem using image classification, object recognition, or semantic segmentation algorithms, solvable with SoTA convolutional neural networks (CNNs) [23].

With the advances in computational resources and the increase in training data sample size, deep CNNs have been extensively used for road distress detection [8]. Compared to the traditional approaches based on support vector machine (SVM) [32], deep CNNs are capable of learning more abstract representations, which have significantly improved the road distress detection performance [21]. [21], [33]–[36] are the most typical image classification-based road distress detection approaches. [37] compared 30 SoTA image classification deep CNNs in terms of detecting road cracks. The authors of [37] conclude from the experimental results that binary road distress detection is a relatively easy task compared to the image classification tasks in other application domains. Furthermore, we typically require instance-level or pixel-level road distress detection results. Therefore, researchers have focused on object detection and semantic segmentation algorithms in recent years.

Instance-level road distress detection approaches are generally developed based on three representative object detection algorithms: (1) single shot multi-box detector (SSD) [38], (2) region-based CNN (R-CNN) series [39], [40], and (3) you only look once (YOLO) [41], [42] series. Amongst them, R-CNN and YOLO series are the most widely used ones for road distress detection. However, the existing instance-level road distress detection approaches generally re-train the original object detection networks without additional modifications. They are infeasible when pixel-level road pothole detection results are desired.

The existing semantic segmentation networks are classified as: (1) single-modal and (2) data-fusion. Single-modal networks generally segment RGB images with encoder-decoder architectures [23], [43]–[45]. On the other hand, data-fusion networks typically learn deep representations from two different types of vision sensor data (color images and depth maps were used in FuseNet [46], color images and surface normal maps were used in SNE-RoadSeg series [47], [48], and color images and transformed disparity images were used in AA-RTFNet [11]) and fuse the learned deep representations to provide a better semantic understanding of the environment. Unfortunately, the SoTA semantic segmentation networks are strong data-driven algorithms that require a considerable amount of data [8]. Therefore, road distress detection with unsupervised or self-supervised learning is a popular area of research that requires more attention.

IV. UDTIRI Competition

The 1st UDTIRI workshop aims to bring together industry professionals and academics to brainstorm and exchange ideas on the advancement of UDT technologies for IRI. In this half-day workshop, we organized a UDTIRI competition in addition to the BigData Cup Challenges. As discussed in Sec. III, object detection and semantic segmentation networks have been widely employed for road distress detection. However, following a thorough literature review, we have found that there are currently no prior arts on instance segmentation-based road distress detection. Therefore, our competition focuses entirely on instance segmentation.

The images in our UDTIRI dataset are retrieved from online search engines. We also captured a collection of real images and included them in the UDTIRI dataset. This results in a total number of 1,000 images. We split our dataset into a training set (600 images containing 885 road potholes), a validation set (100 images containing 151 road potholes),

[6] https://www.blender.org/

and a test set (300 images containing 426 road potholes). A few examples of the images present in our UDTIRI dataset are shown in Fig. 3. The dataset is publicly available at https://www.kaggle.com/datasets/jiahangli617/udtiri. The competitors' submitted results are also published on our workshop webpage. In the future, we will release more UDTIRI datasets on our online benchmark https://www.udtiri.com/.

V. Conclusion

This paper briefly reviewed the state-of-the-art techniques leveraged in the two key components of urban digital twin-based intelligent road inspection systems: (1) multi-temporal, multi-dimension, multi-score, and heterogeneous road data acquisition and (2) road distress detection. This paper also introduced the UDTIRI competition, organized in conjunction with IEEE Bigdata 2022, and a large-scale instance segmentation-based road distress dataset.

References

[1] M. W. Grieves, "Virtually intelligent product systems: Digital and physical twins," *Complex Systems Engineering: Theory and Practice*, vol. 7, p. 175, 2018.

[2] M. Grieves, "Digital twin: manufacturing excellence through virtual factory replication," *White paper*, vol. 1, no. 2014, pp. 1–7, 2014.

[3] F. Tao, M. Zhang, Y. Liu, and A. Y. Nee, "Digital twin driven prognostics and health management for complex equipment," *Cirp Annals*, vol. 67, no. 1, pp. 169–172, 2018.

[4] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, and A. Nee, "Enabling technologies and tools for digital twin," *Journal of Manufacturing Systems*, vol. 58, pp. 3–21, 2021.

[5] F. Dembski, U. Wössner, M. Letzgus, M. Ruddat, and C. Yamu, "Urban digital twins for smart cities and citizens: The case study of herrenberg, germany," *Sustainability*, vol. 12, no. 6, p. 2307, 2020.

[6] S. Guo, Y. Bai, M. J. Bocus, and R. Fan, "Digital transformation for intelligent road condition assessment," in *Intelligent Systems in Digital Transformation*. Springer, 2023, pp. 511–533.

[7] R. Fan, U. Ozgunalp, Y. Wang, M. Liu, and I. Pitas, "Rethinking road surface 3-d reconstruction and pothole detection: From perspective transformation to disparity map segmentation," *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 5799–5808, 2022.

[8] N. Ma, J. Fan, W. Wang, J. Wu, Y. Jiang, L. Xie, and R. Fan, "Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms," *Transportation Safety and Environment*, vol. 4, no. 4, 11 2022.

[9] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Transactions on Image Processing*, vol. 29, pp. 897–908, 2019.

[10] R. Fan, J. Jiao, J. Pan, H. Huang, S. Shen, and M. Liu, "Real-time dense stereo embedded in a uav for road inspection," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2019, pp. 535–543.

[11] R. Fan, H. Wang, M. J. Bocus, and M. Liu, "We learn better road pothole detection: from attention aggregation to adversarial domain adaptation," in *European Conference on Computer Vision Workshops (ECCVW)*. Springer, 2020, pp. 285–300.

[12] J. Laurent, J. F. Hébert, D. Lefebvre, and Y. Savard, "Using 3d laser profiling sensors for the automated measurement of road surface conditions," in *7th RILEM international conference on cracking in pavements*. Springer, 2012, pp. 157–167.

[13] Y.-C. Tsai and A. Chatterjee, "Pothole detection and classification using 3d technology and watershed method," *Journal of Computing in Civil Engineering*, vol. 32, no. 2, p. 04017078, 2018.

[14] M. R. Jahanshahi, F. Jazizadeh, S. F. Masri, and B. Becerik-Gerber, "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor," *Journal of Computing in Civil Engineering*, vol. 27, no. 6, pp. 743–754, 2013.

[15] T. Banerjee, M. Yefimova, J. M. Keller, M. Skubic, D. L. Woods, and M. Rantz, "Exploratory analysis of older adults' sedentary behavior in the primary living area using kinect depth data," *Journal of Ambient Intelligence and Smart Environments*, vol. 9, no. 2, pp. 163–179, 2017.

[16] R. Fan, X. Ai, and N. Dahnoun, "Road surface 3d reconstruction based on dense subpixel disparity map estimation," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3025–3035, 2018.

[17] R. Wang, J. Peethambaran, and D. Chen, "Lidar point clouds to 3-d urban models : a review," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, pp. 606–627, 2018.

[18] K. Ji, H. Chen, H. Di, J. Gong, G. Xiong, J. Qi, and T. Yi, "Cpfg-slam: A robust simultaneous localization and mapping based on lidar in off-road environment," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 650–655.

[19] Z. Ye, H. Xiong, and L. Wang, "Collecting comprehensive traffic information using pavement vibration monitoring data," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 2, pp. 134–149, 2020.

[20] J. Lekshmipathy, N. M. Samuel, and S. Velayudhan, "Vibration vs. vision: Best approach for automated pavement distress detection," *International Journal of Pavement Research and Technology*, vol. 13, no. 4, pp. 402–410, 2020.

[21] R. Fan, M. J. Bocus, Y. Zhu, J. Jiao, L. Wang, F. Ma, S. Cheng, and M. Liu, "Road crack detection using deep convolutional neural network and adaptive thresholding," in *2019 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2019, pp. 474–479.

[22] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.

[23] R. Fan, H. Wang, Y. Wang, M. Liu, and I. Pitas, "Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms," *IEEE transactions on image processing*, vol. 30, pp. 8144–8154, 2021.

[24] A. Fox, B. V. Kumar, J. Chen, and F. Bai, "Multi-lane pothole detection from crowdsourced undersampled vehicle sensor data," *IEEE Transactions on Mobile Computing*, vol. 16, no. 12, pp. 3417–3430, 2017.

[25] R. Fan, L. Wang, M. J. Bocus, and I. Pitas, "Computer stereo vision for autonomous driving: Theory and algorithms," in *Recent Advances in Computer Vision Applications Using Parallel Processing*. Springer, 2023.

[26] R. Fan, S. Guo, L. Wang, and M. J. Bocus, "Computer-aided road inspection: Systems and algorithms," in *Recent Advances in Computer Vision Applications Using Parallel Processing*. Springer, 2023.

[27] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[28] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4104–4113.

[29] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.

[30] J. Janai, F. Güney, A. Behl, A. Geiger *et al.*, "Computer vision for autonomous vehicles: Problems, datasets and state of the art," *Foundations and Trends® in Computer Graphics and Vision*, vol. 12, no. 1–3, pp. 1–308, 2020.

[31] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[32] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[33] V. Pereira, S. Tamura, S. Hayamizu, and H. Fukai, "A deep learning-based approach for road pothole detection in timor leste," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*. IEEE, 2018, pp. 279–284.

[34] W. Ye, W. Jiang, Z. Tong, D. Yuan, and J. Xiao, "Convolutional neural network for pothole detection in asphalt pavement," *Road materials and pavement design*, vol. 22, no. 1, pp. 42–58, 2021.

[35] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016, pp. 3708–3712.

[36] K. E. An, S. W. Lee, S.-K. Ryu, and D. Seo, "Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2018, pp. 1–2.

[37] J. Fan, M. J. Bocus, L. Wang, and R. Fan, "Deep convolutional neural networks for road crack detection: Qualitative and quantitative comparisons," in *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*. IEEE, 2021, pp. 1–6.

[38] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[39] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[40] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[41] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[42] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *CoRR*, 2018.

[43] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[45] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.

[46] C. Hazirbas, L. Ma, C. Domokos, and D. Cremers, "Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture," in *Asian conference on computer vision*. Springer, 2016, pp. 213–228.

[47] R. Fan, H. Wang, P. Cai, and M. Liu, "Sne-roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection," in *European Conference on Computer Vision*. Springer, 2020, pp. 340–356.

[48] H. Wang, R. Fan, P. Cai, and M. Liu, "Sne-roadseg+: Rethinking depth-normal translation and deep supervision for freespace detection," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1140–1145.