Pavement Surface Type Classification Based on Deep Learning to the Automatic Pavement Evaluation System

Aline Calheiros Espíndola[1,2], Ernesto Ferreira Nobre Júnior[2], Elias Teodoro da Silva Júnior[3]

[1]*Technology Center, Federal University of Alagoas*
Av. Lourival Melo Mota, S/N, Tabuleiro do Martins, Zip Code: 57072-970, Maceió, Alagoas, Brazil
[2]*Dept. of Transport Engineering, Federal University of Ceará*
Av. Mister Hull, s/n – Pici, Zip Code: 60455-760, Fortaleza, Ceará, Brazil
[3]*Dept. of Computer Science, Federal Institute of Ceará*
Av. Treze de Maio, 2081 - Benfica, Zip Code: 60040-531, Fortaleza, Ceará, Brazil

aline.espindola@ctec.ufal.br, aline@det.ufc.br, nobre@det.ufc.br, elias@ifce.edu.br

Abstract. Computer vision techniques, image processing, and machine learning became incorporated into an automatic pavement evaluation system with technological advances. However, in most research, the models developed to identify defects in the pavement assume that all the segments evaluated are paved and with one specific pavement surface type. Nevertheless, there is a wide variety of road surface types, especially in urban areas. The present work developed models based on a deep convolutional neural network to identify the pavement surface types considering five classes: asphalt, concrete, interlocking, cobblestone, and unpaved. Models based on ResNet50 architectures were developed; also, the Learning Rate (LR) optimization "one-cycle" training technique was applied. The models were trained using almost 50 thousand images from Brazil's states highway dataset. model results are excellent, highlighting the model based on ResNet50, in which it obtained accuracy, precision, and recall values of almost 100%.

Keywords: Pavement Surface, Image Processing, Convolutional Neural Network.

1 Introduction

Transportation infrastructures are by far the most critical asset of any country. Mature and well-maintained highway networks are crucial to the development of a nation's economy, mainly for countries with a transport matrix based on the highway modal, such as Brazil. Brazilian road transport concentrates approximately 61% of the movement of goods and 95% of passengers and a density of the paved road network of only 25.1 km/1000km^2. As much as Brazil has 1,720,700 km of roads, only 12.4% of the roads are paved, and of those, 59.0% were considered inadequate on at least one of criterion (geometry, signage, or pavement), with 52.4% of the roads assessed showing some pavement defect [1]. In Brazil, only 213,453 km are paved roads; most of them are dirt roads, and mostly unpaved roads are under municipal administration [1]. For example, the Management Report [2] of highways in Ceará State showed that it has 55,827.7 km of roads, with 69.7% under municipal management, and they have only 1.3% of paved highways and small towns no minimum mesh information.

The problem of small cities is systematic, [3] conducted surveys with managers of small and medium-sized Brazilian city halls and concluded that cities are interested in implementing a pavement management system (PMS). Still, human and financial resources are restricted to road assessment activities. Furthermore, over the years, the reality has not changed. The absence of primary data on the network makes the implementation of infrastructure management unfeasible, preventing the definition of the maintenance and restoration plan, the investment plan for implementing new highways, and guidelines for choosing the type of coating on roads with low traffic volume.

The manual approach falls out of favor due to its subjectivity, high financial costs, time, and human resources. In addition, manual inspection is becoming impractical due to the required scale and frequency. With technological advances, there is an increasing tendency to use automated surveys to assess the pavement with equipment ranging from cameras to high technology, such as sensors and computer vision. To deal with the vast amount of data, analysis techniques have also progressed steadily. Researchers have developed a simple regression model for a sophisticated analytical tool such as a neural network to bring fast and accurate results [4].

Researchers have used the neural network for the past twenty years to classify pavement deterioration and quantify severity. In the previous eight years, a convolutional neural network (CNN), a particular type of neural network widely known as deep learning (DL), has gained notable popularity in many sectors, especially in computer vision [5]. However, in pavement engineering, DL techniques' application is constantly evolving, with advances in hardware and software technologies. After a literature review about the application of deep learning in the pavement, the main focuses are on road defects and reveals that attention has been predominantly on the classification, detection, and segmentation of cracks [4].

In this context, the work aims to present a method that incorporates a deep convolutional neural network-based model in automating the registration and diagnosis of roads to enable a pavement management system for small and medium-sized cities.

2 Literature Review

Deep learning methods (DNN – Deep Neural Network) focus on learning hierarchies of features through higher levels formed by a combination of lower-level features. Learning features at multiple levels of abstraction allows learning complex functions mapping the input to the output directly from data [6]. Deep learning technologies have achieved success in various computer vision tasks such as image classification, object detection, and image segmentation.

When analyzing floating-point operations per second (FLOPS) in structures and architectures for the different computer vision tasks, it appears that classification structures require significantly less computational power, DNN typically has billions of FLOPs (multiplication-adds) in the functions of classifications with VGG16 15.5GFLOPs, ResNet 34, and ResNet50 4GFLOPs [7], detection with YOLOv3-Darknet53 65GFLOPs [8] and segmentation U-Net 221GFLOPs or SegNet 341GFLOPs [9].

Therefore, in the development of this research, the ResNet-based multi-class classification was used, as it presents significantly smaller FLOPS and, as will be shown below, satisfactory performance.

2.1 ResNet Architecture

CNN typically requires many labeled images to achieve high prediction accuracy. However, sometimes it is hard to collect thousands of images and manually label them. Transfer learning allows building accurate models with fewer input resources. Rather than starting the learning process from scratch, the transfer learning apporach chooses a model already trained in a larger dataset, such as ImageNet, to solve a similar problem [10]. Pre-trained CNN's weights are taken as initial weights before the training process. Transfer learning helps solve image classification problems of various domains, and it is not different for pavement images [11].

Microsoft Research released a deep Residual Learning Network (ResNet) in 2015. They investigated the performance of the deep neural network by stacking more layers. It was observed that the problem with increasing the number of layers is vanishing or exploding gradients. So residual blocks were used to mitigate those issues and proved that a DNN of thousands of layers is trainable [7]. The resulting network architecture, called ResNet, won in several categories in the 2015 ImageNet competition. The same network is also used by the DAWNBench competition winners, prioritizing reducing training cost while obtaining a 94% test accuracy on the CIFAR-10 dataset and 93% top-5 accuracy scores on ImageNet, respectively. These results in competitions show the computational efficiency of ResNet while obtaining good accuracy in image recognition tasks [12]. The use of ResNet was defined as a pre-trained model due to the good results widely known in literature.

2.2 Systematic Learning Rate Finder

The learning rate (LR) is a tuning parameter in an optimization algorithm that determines the step size at each iteration while moving toward a minimum of a loss function [13]. Smith first proposed [14] and updated [15] a method to find the optimum LR, based on cyclical learning rates (CLR) method, called a "one-cycle" training schedule. This study shows that increasing the LR might have a short-term negative effect and still reach a longer-term beneficial effect. It reinforces the idea that the LR should vary within a range of values instead of adopting a stepwise fixed or exponentially decreasing value. One sets the minimum and maximum boundaries, and the LR cyclically varies between these bounds.

The study proposes one training run for a few epochs to estimate the minimum and maximum boundary values. It is an "LR range test", run your model for a few epochs while letting the LR increase linearly between low and high LR values. The Smith [15] method allowed for higher LR and was still able to achieve great accuracy results on CIFAR-10 and CIFAR-100 in fewer epochs.

Distress pavement examples of studies that used the CLR method to define the LR range were Lau et al. on automated pavement crack segmentation [16] and Aparna et al. on a binary classifier (pothole/non-pothole) for thermal imaging [10] researches. With the CLR method, they were able to enlarge accuracy a smaller number of epochs.

2.3 Related Work

Before detecting the defect, it is important to determine the type of road pavement. Because the pavement types influence on defects type, so it is interesting to have an distress automatic detection model for each type of pavement. For example, there is no bleeding on concrete pavements, only on asphalt. This way, Table 1 shows other researches developed to identify the road pavements, primarily focused on autonomous vehicles and not exactly performing road registration and diagnostics.

Table 1. Road pavement researches

Research	Data	Labels	Description
[17]	accelerometer information	concrete panels, cobblestones, asphalt, and dirt road	Testing CNN, Long short-term memory network (LSTM), and Reservoir computing (RC) models, the best model was based-CNN, accuracy, precision, recall, and F-measure of 85%.
[18]	accelerometer information	dirt roads, cobblestone, and asphalt	The proposed models were based on LSTM, GRU, and deep CNN, with CNN's best performance. The CNN-based obtained an average training accuracy of 93.04%, classifying asphalt segments with an F1-score of 98.60%, cobblestone with 86.09%, and dirt with 90.78%.
[19]	Image from Smartphone	paved and unpaved road	Random forest and support vector machine models were compared with the proposed CNN. The model achieved the best performance to test the dataset, with precision 98.0%, recall 98.4%, and F1-Score 98.2%.
[20]	Image from Google Street View	divide roads (C1), express roads (C2), paved street roads (C3), unpaved street roads (C4), parquet streets (C5), and distorted roads (C6)	The proposed approach shows that the road types were determined with an accuracy of 91.41%. The model had the best performance to identify express roads (C2) with 96.0% and the worst accuracy to paved street roads (C3) with 79.2%
[21]	Image from a camera into survey vehicle	asphalt, paved and unpaved road	A classifier-based deep CNN model was proposed, using three datasets during the training KITTI [22], CaRINA [23], and RTK [21]. The lowest accuracy result was 86.05% for RTK, 76.23% to KITTI and 99.9% to CaRINA.

3 Methods

3.1 Dataset

The dataset is composed of images from highways and local roads located in Ceará state, Brazil. The right of way (ROW) video images with 1920x1080 were captured by a Garmin VIRB Ultra30 Action camera mounted over a survey vehicle. The input images are cropped to 16% of the original size, as shown in Figure 1. This part of the image is ROI (Region of Interest). This procedure is adopted because the large part of the image is not pavement elements, and the low clearance of the road sections is far from the survey vehicle. After defining the ROI, the images are resized to 300x300 pixels.

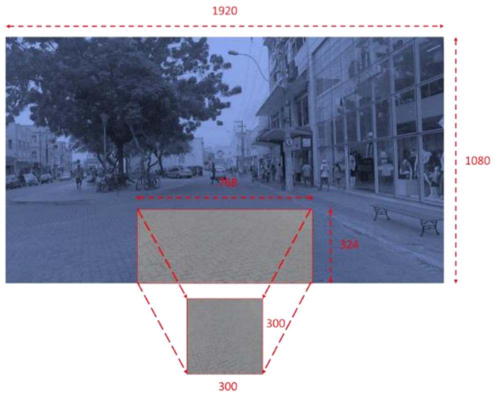

Figure 1. Pre-processing action to ROI (Region of Interest).

In total, 49173 images of 300x300 pixels were obtained. The classes of pavement types are asphalt, concrete, cobblestone/stone, interlocking, and unpaved, as shown in Figure 2. The amount of images in each class can be seen in Table 2.

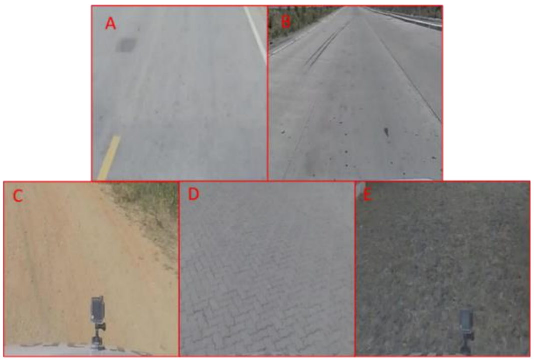

Figure 2. Pavement Classes (A) Asphalt, (B) Concrete, (C) Unpaved, (D) Interlocking and (E) Cobblestone/Stone

Table 2. An overview of images for each pavement class.

	Asphalt	Cobblestone	Concrete	Interlocking	Unpaved	Total	Train Set	Valid Set
Images Amount	19992	8973	3701	7674	8833	49173	39339	9834
Proportion	41%	18%	8%	16%	18%			

3.2 Training the network

The training of a CNN is a procedure comprising mathematical optimization, using a cost function, called loss function, so the neural network's weights and biases can be automatically tuned. In this process, the training dataset with expected labels is presented to the optimization procedure to find a global minimum loss function.

CNN training has massive computational time consumption. Therefore, it is necessary to use GPUs that can process data in a parallel fashion; the other alternatives are field-programmable gate arrays or application-specific integrated circuits. For this purpose, the computing resources of Google Colab were used to conduct CNN training. Due to the nature of Google Colab, each training is performed on a randomly allocated GPU, and Tesla T4 GPU was allocated for the training. The DL framework used is PyTorch version1.5.0+cu101 with the fastaiv2 library.

The input parameters used for model training are ResNet50, as a pre-train model; image size of 224x224 pixels; the batch size of 64; 15 epochs and cross-entropy (flattened loss) the loss function. And the dataset was divide in 80% of the images are used for the training set and 20% for the validation set, each chosen randomly.

Accuracy, precision, recall, and F1-score were the metrics to evaluate the model performances. In addition, the confusion matrix will be used to to identify which classes are being mistaked by the classifier and the intensity of this confusion.

Building CNN with a pre-training model has been shown a suitable pavement distress identification, as stated by Nie and Wang [24]. Depending on the resemblance of the data that the CNN will deal with in its lifecycle, the weights produced by training need to be altered. Thus, the last layers of the network often need enormous changes in parameters, while deeper levels that are already well trained in detecting basic features (as edges and outlines) need minor modifications. Since transfer-learning is used in network architecture, CNN is fine-tuned in two steps. The first step freezes the first layer group so that all the parameters in it are not updated—the last layer-groups train for 15 epochs. In the second step, the first layer group is unfrozen. The training is continued from the previous training state for an additional 15 epochs. In other words, the first layer group is trained after the other layer groups are well optimized. By training the latter parts of the pre-trained ResNet-50, this procedure takes maximum advantage of transfer learning.

During the training of the first step, in which the firsts layers were frozen, it is used the CLR method to define LR, so the graphs loss vs. LR (Figure 3) was plotted, and used to choose the maximum LR. Furthermore, the minimum LR value is set to 0,25% of the maximum LR. The range value of LR is between 10^{-4} to 4.10^{-2}. For the second step, more tests were needed to reach the LR with better performance because after training the second step of CNN with the interval of 5.10^{-5} and 10^{-4}, there was no accuracy gain. It was observed that the scale of losses in Figure 3 is in the centesimal, varying from 0.03 to 0.02, it was decided to expand the LR range, after tests the range of 10^{-6} and 10^{-4} presented the most significant gain of performance, where the minimum LR value was 1% of the maximum LR.

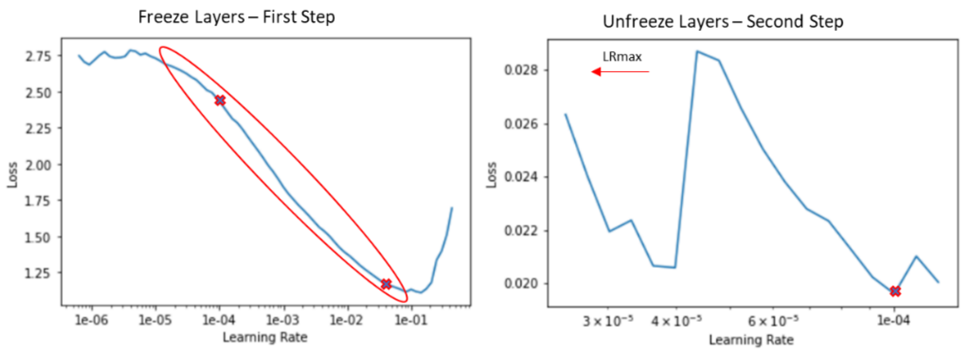

Figure 3. The graph of validation loss vs LR.

4 Results and Discussion

The final model showed extremely low losses, with a training loss of 0.002055, validation loss of 0.034294, and an error rate of 0.001424. The CNN-based model's general results are presented in Table 3; the results of the four metrics for the road pavement type model presents an almost perfect performance, close to 100%.

Table 3. Performance of Pavement surface Type Model

	Accuracy	Precision	Recall	F1-score
Asphalt	99,99%	99,98%	100,00%	99,99%
Concrete	100,00%	100,00%	100,00%	100,00%
Interlocking	100,00%	100,00%	100,00%	100,00%
Unpaved	99,86%	99,77%	99,42%	99,59%
Cobblestone	99,87%	99,49%	99,77%	99,63%
Average	99,94%	99,85%	99,84%	99,84%

Concrete and Interlocking images did not present any false-positive or false-negative, with 100% performance in the four metrics (Accuracy, Precision, Recall, and F1-score). The asphalt class presented a single false-positive and no longer achieved 100% in the metrics. The Unpaved and Cobblestone classes have false-positive or false-negative, but still, low amounts, getting values on all metrics above 99.4%.

9834 images are separated for in the validation set, images not used in training, so the model did not know its characteristics and labels; there were only 14 mistakes in identifying the pavement type. For example, one unpaved image was classified as asphalt, four cobblestone/rough stone images were classified as unpaved, and nine unpaved images were classified as cobblestone/rough stone.

Unpaved and cobblestone were classes with lowest accuracy, according to Table 3. Figure 4 shows the confusion matrix for the proposed classifier, which shows that the CNN mostly confused images from those classes unpaved and cobblestone/rough stone.

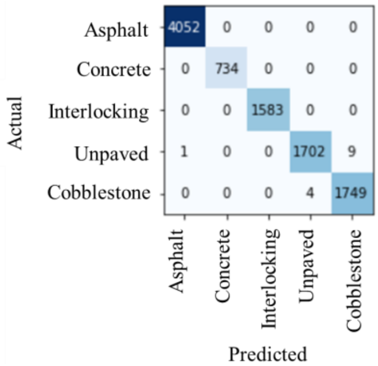

Figure 4. Confusion Matrix

Figure 5 presents an example in which the neural network mistakes the classes. The original label was the cobblestone/stone road, and the model predicts unpaved road. However, it is an understandable error since the surface has deteriorated, covered with dirt, and rough stones are not visible. And Figure 6 presents several images that have been correctly classified.

Figure 5. An example image that the model has misclassified.

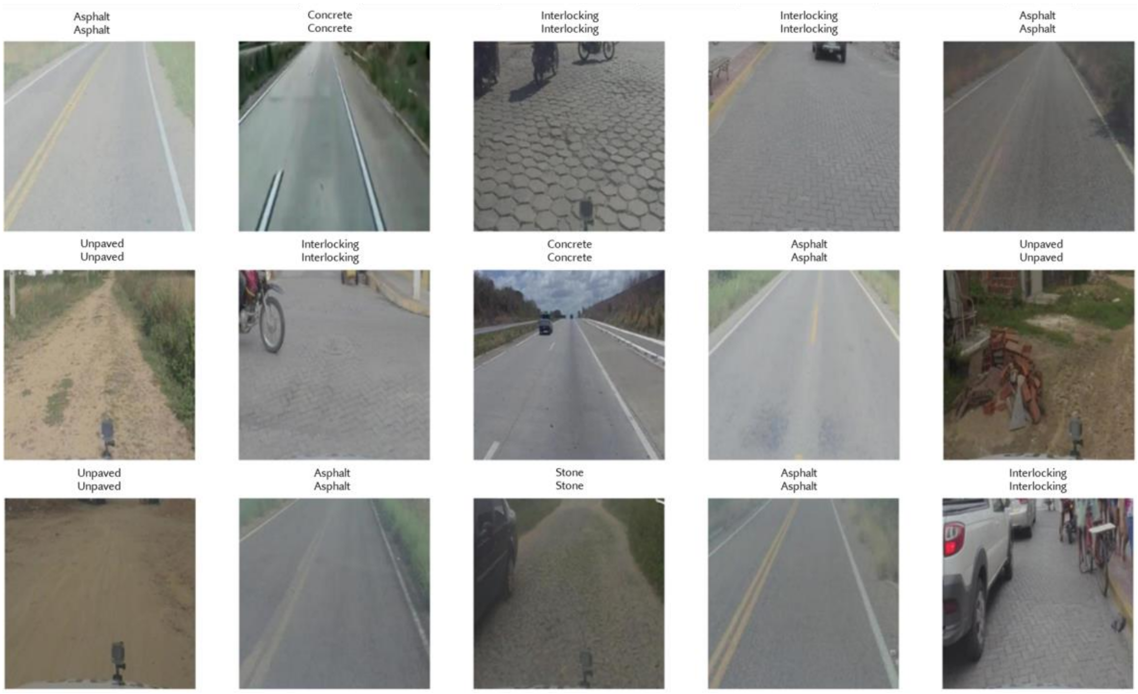

Figure 6. Examples of images and the results of the classification of the pavement surface

5 Conclusions

It is possible to create a high accuracy multi-class classifier based on a convolutional neural network to identify road pavement type. Besides, transfer learning and “systematic learning rate finder” techniques facilitate outstanding performance achievement.

Using a multi-class classifier allows the improvement of techniques for automating pavement evaluations. It would be the first step of the automatic network level PMS because it identifies the pavement surface type. After, the defect detection model must be applied for each pavement type.

The continuation of the research are training and testing the model on other datasets with characteristics total different from the original dataset. A down face and back camera, different position and angulation on vehicle, image size (pixels), and lighting clarity. To develop the model's generalization to identify the road pavement type regardless of the image characteristics inputted.

Acknowledgements. This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-Brazil (CAPES) under Grant [Finance Code001]. The authors would like to express their appreciation for CERTARE Engenharia e Consultoria Ltda for providing the images dataset.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors or has the permission of the owners to be included here.

References

[1] CNT, "Pesquisa CNT de Rodovias 2019," Brasília, 2019.
[2] S. D. O. P. SOP, "Informativo Gerencial," 2018.
[3] P. Causim, "Estudo de um Sistema de Gerencia de Pavimentos para Cidades de Pequeno e Medio Porte," UNIVERSIDADE ESTADUAL DE CAMPINAS, 2001.
[4] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Constr Build Mater*, vol. 157, pp. 322–330, 2017.
[5] K. Gopalakrishnan, "Deep learning in data-driven pavement image analysis and automated distress detection: A review," *Data*, vol. 3, no. 3, 2018.
[6] C. R. Ikram, "A benchmark for evaluating Deep Learning based Image Analytics," UNIVERSITY OF OSLO, 2019.
[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
[8] X. Long *et al.*, "PP-YOLO: An effective and efficient implementation of object detector," *arXiv*, 2020.
[9] L. Ding *et al.*, "MP-ResNet: Multi-path residual network for the semantic segmentation of high-resolution PolSAR images," *arXiv*, pp. 1–5, 2020.
[10] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, "Convolutional neural networks based potholes detection using thermal imaging," *J King Saud Univ - Comput Inf Sci*, 2019.
[11] N. Tajbakhsh *et al.*, "Convolutional Neural Networks for Medical Image Analysis : Full Training or Fine Tuning ?," vol. 35, no. 5, pp. 1299–1312, 2016.
[12] C. Coleman *et al.*, "DAWNBench : An End-to-End Deep Learning Benchmark and Competition," in *31st Conference on Neural Information Processing Systems*, 2017.
[13] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. 2012.
[14] L. N. Smith, "No More Pesky Learning Rate Guessing Games," 2015.
[15] L. N. Smith, "Cyclical learning rates for training neural networks," *Proc - 2017 IEEE Winter Conf Appl Comput Vision, WACV 2017*, no. April, pp. 464–472, 2017.
[16] S. L. H. Lau, X. Wang, Y. Xu, and E. K. P. Chong, "Automated Pavement Crack Segmentation Using Fully Convolutional U-Net with a Pretrained ResNet-34 Encoder," *IEEE*, 2020.
[17] B. Varona, A. Monteserin, and A. Teyseyre, "A deep learning approach to automatic road surface monitoring and pothole detection," *Pers Ubiquitous Comput*, vol. 24, no. 4, pp. 519–534, 2020.
[18] J. Menegazzo and A. Von Wangenheim, "Multi-Contextual and Multi-Aspect Analysis for Road Surface Type Classification Through Inertial Sensors and Deep Learning," *IEEE X Brazilian Symp Comput Syst Eng*, 2020.
[19] V. Pereira, S. Tamura, S. Hayamizu, and H. Fukai, "Classification of Paved and Unpaved Road Image Using Convolutional Neural Network for Road Condition Inspection System," *ICAICTA 2018 - 5th Int Conf Adv Informatics Concepts Theory Appl*, pp. 165–169, 2018.
[20] V. Tumen, O. Yildirim, and B. Ergen, "Recognition of road type and quality for advanced driver assistance systems with deep learning," *Elektron ir Elektrotechnika*, vol. 24, no. 6, pp. 67–74, 2018.
[21] T. Rateke, K. A. Justen, and A. von Wangenheim, "Road surface classification with images captured from low-cost camera-road traversing knowledge (RTK) dataset," *Rev Inform Teor e Apl*, vol. 26, no. 3, pp. 50–64, 2019.
[22] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research," *Int J Rob Res*, no. October, pp. 1–6, 2013.
[23] P. Y. Shinzato *et al.*, "CaRINA dataset: An emerging-country urban scenario benchmark for road detection systems," *IEEE Conf Intell Transp Syst Proceedings, ITSC*, pp. 41–46, 2016.
[24] M. Nie and K. Wang, "Pavement Distress Detection Based on Transfer Learning," *2018 5th Int Conf Syst Informatics, ICSAI 2018*, no. Icsai, pp. 435–439, 2018.