



Article

A Low-Cost Deep Learning System to Characterize Asphalt Surface Deterioration

Diogo Inácio ¹, Henrique Oliveira ^{2,3,*} , Pedro Oliveira ⁴ and Paulo Correia ^{1,2}

¹ Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal; diogobpalma@tecnico.ulisboa.pt (D.I.); plc@lx.it.pt (P.C.)

² Instituto de Telecomunicações (IT-Lisboa), 1049-001 Lisboa, Portugal

³ Instituto Politécnico de Beja, 7800-111 Beja, Portugal

⁴ Tecnofisil, Av. Luís Bivar 85A, 1050-143 Lisboa, Portugal; pedro.oliveira@tecnofisil.pt

* Correspondence: hjmo@lx.it.pt

Abstract: Every day millions of people travel on highways for work- or leisure-related purposes. Ensuring road safety is thus of paramount importance, and maintaining good-quality road pavements is essential, requiring an effective maintenance policy. The automation of some road pavement maintenance tasks can reduce the time and effort required from experts. This paper proposes a simple system to help speed up road pavement surface inspection and its analysis towards making maintenance decisions. A low-cost video camera mounted on a vehicle was used to capture pavement imagery, which was fed to an automatic crack detection and classification system based on deep neural networks. The system provided two types of output: (i) a cracking percentage per road segment, providing an alert to areas that require attention from the experts; (ii) a segmentation map highlighting which areas of the road pavement surface are affected by cracking. With this data, it became possible to select which maintenance or rehabilitation processes the road pavement required. The system achieved promising results in the analysis of highway pavements, and being automated and having a low processing time, the system is expected to be an effective aid for experts dealing with road pavement maintenance.

Keywords: crack segmentation; crack classification; deep neural networks; characterization of highways condition



Citation: Inácio, D.; Oliveira, H.; Oliveira, P.; Correia, P. A Low-Cost Deep Learning System to Characterize Asphalt Surface Deterioration. *Remote Sens.* **2023**, *15*, 1701. <https://doi.org/10.3390/rs15061701>

Academic Editors: Valerio Baiocchi, Alessandro Mei and Xianfeng Zhang

Received: 30 December 2022

Revised: 13 March 2023

Accepted: 20 March 2023

Published: 22 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to their constant use, road pavement surfaces are subject to continuous degradation, with cracks being the first sign of pavement surface deterioration [1]. Therefore, crack detection is crucial for monitoring and maintaining pavement surfaces, ensuring the security of the drivers. However, if traditional human inspection procedures are used, crack detection can be extremely tedious, time-consuming, and subjective [2]. Implementing automatic pavement condition monitoring systems can overcome these limitations, allowing a more precise, faster, and safer analysis than traditional methods, minimizing experts' effort and human subjectivity.

The automatic detection and classification of cracks in road pavements are challenging due to the different characteristics of pavement materials and the wide variability of crack shapes and their inhomogeneous textures, among other factors. Moreover, depending on the sensors used to gather road pavement imagery, the resulting images may be affected by different types of noise, shadows, debris, or oil and water spots. These challenges make developing a robust and reliable automatic crack detection and classification system a difficult task.

1.1. Contributions

This paper proposes an automatic crack detection and classification system based on deep neural networks to help speed up highway pavement surface monitoring. This

work was conducted in collaboration with the Portuguese company Tecnofisil [3]. The contributions of this paper are:

1. A segmentation deep neural network, based on U-Net architecture, for identifying crack regions in images that was trained using five publicly available datasets.
2. A classification deep neural network used to classify detected cracks into one of four classes: (i) alligator; (ii) longitudinal; (iii) transverse; or (iv) non-crack.
3. A system used to automatically estimate the percentage of cracking present in a road pavement segment to serve as an aid for experts in their maintenance planning.
4. An anomaly detection method based on isolated cracking classifications that relates uncertainty to the proposed system results.

The paper is organized as follows. After this introduction, Section 1.2 briefly reviews the related work. Section 2 presents the proposed methodology and provides details about the development process, while Section 3 presents the experimental results, and Section 4 discusses the experimental results. This paper ends with Section 5, in which conclusions and proposals for future work are presented.

1.2. Related Work

Crack detection in road pavements has been a hot topic of research over the years due to its importance, high level of complexity, and challenging characteristics. It has seen a remarkable evolution and a huge variety of proposals based on different methodologies. What most differentiates the available methodologies using a camera sensor is the strategy used to extract features from images that allow for identifying pavement surface distresses [4].

In the scientific literature, two main categories of methods can be identified: (i) those based on traditional image processing; (ii) those using machine learning techniques. Traditional image processing techniques can be further divided into three subcategories: (i) threshold segmentation; (ii) edge detection; and (iii) region growing [5].

Crack detection using threshold segmentation assumes that pixels belonging to a crack are darker compared to its surroundings [6]. These methods start with a pixel-based analysis to determine whether they belong to a crack or to the background, where selecting a suitable threshold value is the key to good method performance. For instance, Oliveira and Correia [7] identified dark pixels belonging to potential cracks using the dynamic thresholding technique that comprised two steps. The first threshold was able to label image pixels into “image background” and “potentially belonging to cracks”. Then, after dividing the resulting segmented image into non-overlapping blocks of a given dimension, they were finally classified into “crack” and “non-crack” after applying a second threshold determined based on the set of entropy values obtained from the analysis of the pixel intensities of each non-overlapping image block (as many entropy values as the number of non-overlapping blocks in the image).

Edge detection methods are based on identifying the edges of regions of the image presenting pavement surface distress, enabling the outlining of their contours [6]. Various edge detection operators include Sobel, Roberts, Prewitt, and Canny operators. However, using a single operator can hardly reach the desired results. Hence, many researchers used edge detection operators with other techniques to improve crack detection performance [5]. Ayenu-Prah and Attoh-Okine [8] proposed a road crack detection method, combining a bi-dimensional empirical mode decomposition (BEMD) with Sobel edge detection with BEMD, extending the original empirical mode decomposition [9] to remove noise from the input signal efficiently.

Region-growing strategies aim to gather similar pixels together to form regions. Then, the characteristics of pavement surface defect regions can be estimated. Li et al. [10] proposed an automatic crack detection method, the F* Seed growth Algorithm (FoSA), to improve the detection of discontinuous and blurred cracks. The FoSA is an extension of the F* method [11], exploiting a seed-growing strategy (as illustrated in Figure 1) to eliminate the requirement that start and endpoints should be known or pre-defined in advance. The

global search area is reduced to local to improve search efficiency. However, the results of FoSA may be impaired by the lighting conditions of the images to be analyzed.

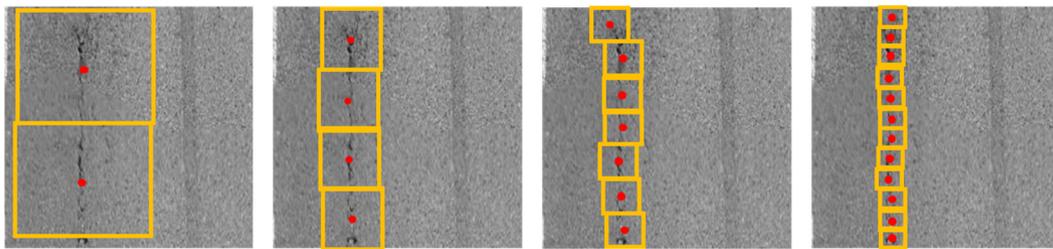


Figure 1. Illustration of the seed-growing strategy described in [10] for different seed-growing radius.

More recently, through the enormous development of machine learning techniques, deep learning has dominated proposals for the detection of cracks in pavement surface images taken during road surveys, as well as their classification, by allowing a computer to automatically learn new characteristics and classification rules based on a sufficiently large set of representative training examples. Adopting a supervised learning approach requires that the training set contains images labelled with information about the presence and location of cracks, a task that typically requires a substantial previous effort by human experts to produce this labelling information and that can bring out the subjectivity of human analysis.

Methods based on deep learning have achieved great success in many related areas, such as image classification, object detection, and image segmentation. Therefore, deep learning methods, mainly using convolutional neural networks (CNN), have also been proposed for automatic crack detection and/or crack-type classification.

Crack detection and crack-type classification can be seen as image classification problems, and two different approaches can be followed: (i) assigning a label to the whole input image; or (ii) dividing the image into blocks and classifying each block as belonging to a particular class. The labels to be considered can be binary (“crack” or “non-crack”) or else indicate the type of crack that was detected (longitudinal, transverse, and crocodile skin, among others).

For example, Lei et al. [12] proposed a deep learning-based algorithm that divides pavement surface images of size 3264×2248 pixels into smaller blocks of size 99×99 pixels (keeping the 3 RGB channels). They used a CNN to perform a binary classification of each image block according to the probability of whether or not it contains cracks. The CNN used by the authors was ConvNet, and its architecture is illustrated in Figure 2.

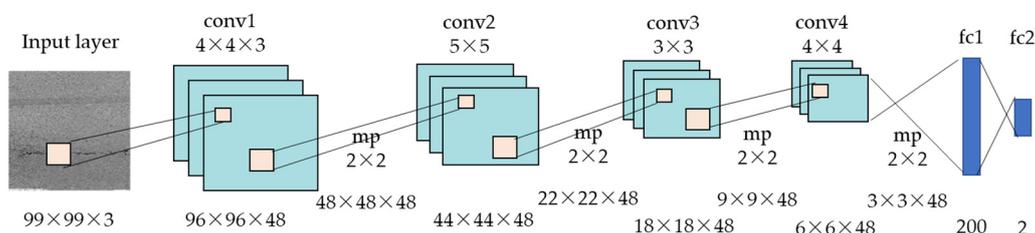


Figure 2. Illustration of the ConvNet architecture presented in [12].

Aslan et al. [13] implemented a CNN that classifies several types of surface distress present in images of road pavements. The proposed CNN architecture (Figure 3) consists of two convolutional units with 32 filters, each applied before a 16-neuron fully-connected layer and a softmax classification layer with four output neurons. The developed model was trained through a balanced dataset that includes four types of pavement surface defects: (i) longitudinal crack; (ii) transversal crack; (iii) alligator crack; and (iv) pothole.

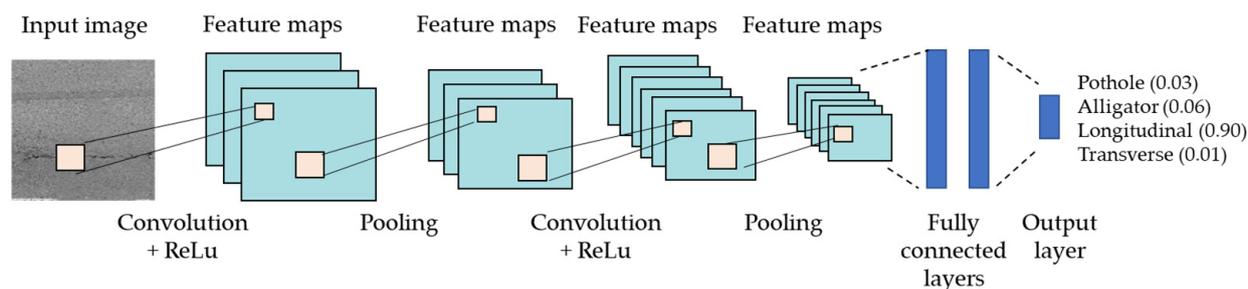


Figure 3. Illustration of the CNN architecture proposed in [13].

Anand et al. [14] proposed an autonomous real-time system with an associated camera to detect cracks and potholes in images of road pavement surfaces. The method, denoted Crack-pot, uses traditional image processing techniques, such as edge detection, to locate and generate the bounding boxes of the detected objects, combined with a CNN employed as a classification model.

Jenkins et al. [15] proposed a fully convolutional neural network [16] to classify all pixels presented in pavement surface images. The network architecture proposed by the authors is an encoder–decoder architecture based on the U-Net model [17]. In this type of architecture, the layers belonging to the encoder reduce the size of the input image through down-sampling operations and learn a high density of lower-level feature maps. The layers belonging to the decoder then map the encoded features back to their original resolution, using pooling layers to perform up-sampling operations efficiently. The network ends with a classification layer to assign an individual label to each pixel. During the evaluation of this method, the authors concluded that the training data were scarce and contained a low diversity of cracks, which meant that the performance of the proposed classification system could still improve, given the network’s capabilities.

Zhang et al. [18] proposed a fully convolutional network for per-pixel semantic segmentation/detection. The authors stated that their system, CrackGAN, aims to solve issues found in published research works dealing with the crack detection problem in images of pavement surfaces: the one known as “All black”, which occurs in FCN-based pixel-level crack detection when using partially accurate ground truths, where the network treats all the pixels as belonging to the image background, along with the data imbalance issue concerning the network’s training step. Their system processed images from three publicly available image datasets, but most of them were captured by very sophisticated imaging systems, namely laser imaging systems. Although the proposed approach achieved state-of-the-art performance, the calculation of the pavement surface distress level was not performed, nor the classification of distress types.

Liu et al. [19] proposed a feature fusion encoder–decoder network (FFEDN) with two novel modules to improve the crack detection accuracy by enhancing the representation capability of tiny cracks and reducing the influence of background interference. The FFEDN system processed images from three publicly available image datasets mainly captured by cell phones, and published results showed that the proposed system outperformed all the eight methods used for comparison. Nevertheless, all the images presented in the published work are free of artifacts in the pavement surface, such as white lane markings (continuous or interrupted), oil spots, and shadows cast by objects near the road edges, which does not allow us to assess on the system performance in processing images with such artifacts.

A recently published research paper by Ma et al. [20] addressed problems regarding data acquisition (image and video) and the defect counting of road pavement surfaces, aiming at increasing the efficiency and accuracy of traditional detection systems. The authors developed a generative adversarial network (PCGAN) to generate realistic crack images, aiming to solve the problem of small amounts of training data, an important issue when using deep neural networks. Regarding crack detection, an improved version of a

regression-based object detection and median flow model was developed (YOLO-MF) that allows for obtaining the number of cracks in an image/video captured by an imaging sensor carried by an unmanned aerial vehicle. Although the published performance metrics are auspicious, the images shown in the paper exhibit very prominent cracks that appear to be easily detectable. On the other hand, the images presented in the paper did not show some often-visible artifacts on road pavement surfaces, such as oil spots or even white lane markings. Additionally, the authors concluded that their method presented difficulties in dealing with small and complex pavement distresses and detecting extensive alligator cracks.

CrackW-Net was the methodology developed by Han et al. [21], namely a skip-level round-trip sampling block structure with the implementation of convolutional neural networks aimed at segmenting cracks in images of the pavement surface at the pixel level. The authors tested their proposed approach with only two image datasets, namely the Crack500 [22] and a self-built dataset. Processed images were free from well-known artifacts (oil spots and white lane markings, among others), with most of them exhibiting simple and fairly noticeable cracks. CrackW-Net training costs were compared to FCN [23], U-Net [24], and ResU-Net [25], which were significantly higher concerning the number of neural network parameters and the training speed. The authors concluded that extensive validations are still needed before the use of CrackW-Net becomes feasible.

As highlighted, numerous studies have been devoted to deep learning in crack detection and classification due to convolution neural networks' excellent feature representation capabilities. The approach presented in this paper follows this trend, proposing, however, a segmentation deep neural network to identify crack regions and a classification deep neural network to classify detected cracks into types to handle more complex images of road pavement surfaces that frequently present noticeable artifacts' visible and more complex crack shapes and textures.

2. Materials and Methods

This section describes the proposed methodology for the developed crack detection and classification system. It consists of five main steps, as represented by the generic architecture presented in Figure 4.

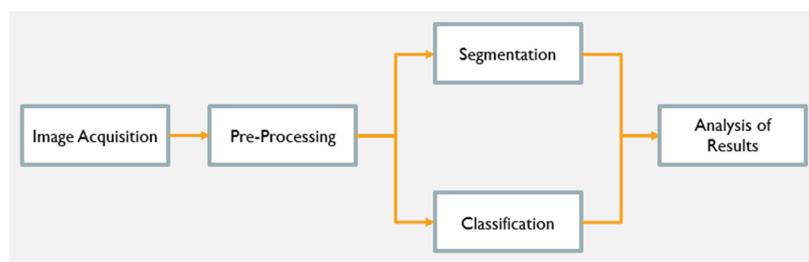


Figure 4. Generic architecture of the proposed system.

The five main steps of the proposed system consist of the following:

1. Image Acquisition—This step deals with the acquisition of the road pavement images to be analyzed. A set of representative images was considered to create a dataset for usage along the development of the crack detection and classification system.
2. Preprocessing—This step performs a set of image manipulations and transformations, notably for normalization purposes.
3. Segmentation—This step creates an image map identifying which image pixels are affected by cracking.
4. Classification—This step assigns a crack type classification for each detected crack.
5. Analysis of Results—This step is responsible for combining the results produced by the segmentation and classification steps.

The proposed system captures pavement surface images using a conventional 2D camera. Each road is analyzed considering segments of a pre-specified length (e.g., 100 m). The output consists of road pavement crack segmentation maps and a list containing the type of cracking that the system assigned to each road segment. The analysis of results module then outputs a set of tables summarizing the calculated information, including a distressing percentage per road segment, which provides valuable help for the quick identification of road segments that require careful analysis by experts when planning road maintenance operations.

2.1. Image Acquisition

The proposed system includes a simple image acquisition system consisting of a conventional 2D camera that can be easily mounted on a regular vehicle. The system was provided by Tecnofisil [3], and an illustration of its setup is included in Figure 5; for the purpose of this work, the camera was set up on the back of a small van and mounted to guarantee the desired slant angle towards the ground.



Figure 5. Image acquisition system consisting of a conventional 2D camera mounted on the back of a vehicle.

The image sensor used is a Teledyne Lumenera Lt16059H high-performance camera [26], which has a 16 Megapixel Charge-Coupled Device (CCD) sensor and an electronic shutter, that can capture high-quality and high-speed images. The maximum resolution of this camera is 4864×3232 pixels. A sample acquired image is shown in Figure 6, where it is possible to see that the road pavement surface region of interest to be analyzed, corresponding to the lane where the van is travelling, appears with a trapezoidal shape marked in red in the figure.

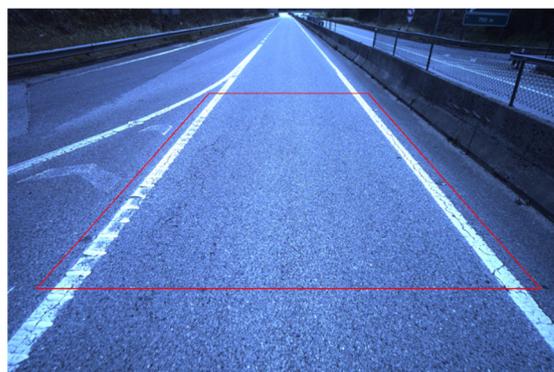


Figure 6. Sample image acquired by the proposed system. The area of interest for analysis is delimited in red.

Image acquisition needs to ensure that the entire width of the road lane is captured. Additionally, the acquisition system should not skip any portion of the road. For this reason, it was set up so that consecutive images have some overlap in the captured regions of interest. The setup of the image acquisition system requires setting the following parameters:

- Camera slant angle;
- Camera focal length;
- Camera frame rate;
- Camera resolution;
- Vehicle speed.

Image acquisition tests were conducted on Portuguese highways, considering an initial slant angle of 30° towards the ground and camera focal length values of 16 and 28 mm, corresponding to the minimum and maximum values supported by the imaging sensor, respectively. Figure 7 shows sample images captured with these parameters. Please note that the present work is mostly focused on the analysis of highways and other roads for circulating at relatively high speeds and not so much for very twisty roads.



Figure 7. Images captured with a slant angle of 30° and a camera focal length of: (a) 16 mm; (b) 28 mm.

As expected, when considering a short focal length, as illustrated in Figure 7a, a more extensive road area is captured due to a wider angle of view, although the details of the pavement surface are smaller due to the lower magnification. On the other side, with a longer focal length, as illustrated in Figure 7b, a smaller road area is captured, but the pavement surface details are more visible. This second setting would require the driver to keep the vehicle in the middle of the road lane, which can be challenging depending on the road trajectory.

After the initial tests, it was decided to adopt a camera focal length of 22 mm, which ensures an easy capture of the entire road lane width, with good focus and detail in the region of interest of the pavement surface. A sample image acquired in these conditions is presented in Figure 6.

The camera slant angle was kept at 30° towards the pavement surface because, for higher angles, the images could suffer from shadow effects cast by the vehicle itself, and the van's roof could also appear. On the other hand, for angles less than 30° , the pavement surface closest to the van was no longer captured, but the region of interest would be captured at a greater distance and, consequently, at a lower resolution.

The selection of the remaining three parameters, namely camera frame rate, camera resolution, and vehicle speed, should be performed in such a way as to ensure that an overlap area exists between the regions of interest of consecutive images to make sure that there is no loss of pavement surface area in the analysis. The camera offers a trade-off between frame rate and resolution; with the maximum resolution of 4864×3232 pixels,

the camera can operate at a maximum of 3 frames per second. For higher frame rates, the resolution should be lower. For instance, reducing to a quarter of the maximum resolution (2432×1616 pixels) increases the frame rate to 8 frames per second. With this in mind, several tests were conducted with images captured on a 3 km stretch of a major Portuguese road to decide the values of the above-mentioned parameters. A summary of the tests conducted is presented in Table 1, with the overlap area of the pavement surface between consecutive images included in the last column. The length of the region of interest was set to 6.5 m.

Table 1. Tests conducted to help set the values of the camera frame rate, camera resolution, and vehicle speed. The last column shows the overlap between the regions of interest in consecutive images.

Trial	Speed (km/h)	Frame Rate	Resolution	Overlapping Region (m)
1	50	3	4864×3232	2/6.5
2, 3, 4	70	3	4864×3232	not guaranteed
5, 6	70	8	2432×1616	4.5/6.5
7, 8	90	8	2432×1616	3.5/6.5

Table 1 shows the results for four combinations of speed, frame rate, and resolution parameters. During these tests, it was noticed that the camera frame rate underwent minor variations during the acquisition, meaning that even if the frame rate is set at 3 or 8 frames per second, it will vary, typically in the range of ± 0.1 – 0.3 frames per second. These oscillations influence the overlapping region achieved between consecutive images. The technical team setting up the project decided that an overlap of 3.5 m, with a region of interest of length of 6.5 m, would be adequate.

A camera resolution of 2432×1616 pixels was considered sufficient for crack detection when using the previously defined values of a 30° slant angle and a focal distance of 22 mm. The value of the camera frame rate (and consequently of the time lapse between successive shots) was adjusted according to the vehicle speed, as summarized in Table 2, to achieve a 3.5 m overlap between the regions of interest of consecutively captured images.

Table 2. Tradeoff between vehicle speed and camera frame rate to achieve a 3.5 m overlap between consecutively captured images.

Speed (km/h)	Frame Rate (fps)	Time Lapse (ms)
50	4.6	217
70	6.5	154
90	8	120

The proposed system image acquisition can be used to acquire images for training, validation, and testing sets; however, other images extracted from public datasets were used to train the neural networks that compose the proposed system to increase image diversity in training. As automatic crack detection in pavement surface images is an emergent topic that has been investigated over the years, several image datasets of pavement surfaces are publicly available, allowing for the validation and testing of new models and systems. They differ by various parameters, such as the sensor with which the images are captured and the scale used, among others. The public datasets that were used in this work are as follows:

- CrackForest: Contains 118 crack images of a road pavement surface in Beijing that are 480×320 pixels in size. The sensor used was the camera of an iPhone5 [27].

- AigleRN: Contains 38 pre-processed grayscale images of a road pavement surface in France. Half of these images are of size 991×462 pixels, and the other half are of size 311×462 pixels [28].
- CrackTree260: Contains 260 images of size 800×600 pixels, captured by an area-array camera under visible light illumination conditions [29].
- CRKWH100: Contains 100 images of a road pavement surface of size 512×512 pixels, captured by a line array camera under visible light illumination conditions [30].
- CrackLS315: Contains 315 images of size 512×512 pixels, captured by a line array camera under laser illumination [31].

These five public datasets have differences in lighting conditions, texture, pavement type, and other characteristics, ensuring image diversity in the neural network training process. Along with image diversity, it is also appropriate to include images of the training set, such as those that the neural networks analyze in their applications to improve neural network performance. Accordingly, it was considered advantageous to include images of Portuguese road pavements in the training step of the classification model, namely some images also obtained during the survey of the “IP3”.

It should be noted that, in order to be able to use images of different sizes, coming from several public datasets, for the purposes of the training, validation, and testing of the neural networks used by the proposed system, those images were first resized to match the input size expected by each of the neural networks used, notably considering an input size of 480×320 for the segmentation network and an input size of 512×512 for the classification network.

2.2. Preprocessing

Image pre-processing includes two main tasks: (i) identifying the region of interest of the captured pavement surface image and converting its trapezoidal shape to a rectangular shape; (ii) creating a more extensive image of the pavement surface area for distress analysis by compositing several consecutive reshaped regions of interest. This second task is performed because, from the point of view of the road analysis experts, observing surface distress in images of road lane sections longer than 6.5 m in length provides more relevant results.

In the first task, all images were converted to grayscale and then subjected to a perspective transformation, as illustrated in Figure 8. This task was accomplished using the functions `getPerspectiveTransform()` and `warpPerspective()` available in the OpenCV library [32].

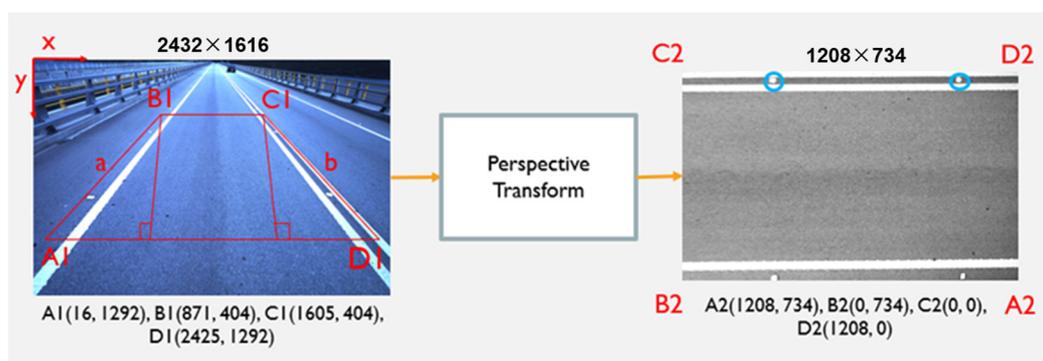


Figure 8. Selection of the region of interest (left) and the output of perspective transformation (right).

In order to define the region of interest, a 20% lower cut and a 25% uppercut were made relative to the image’s vertical axis, fixing the vertical (y) value of points A1 and D1 and points B1 and C1, respectively. The 20% lower cut ensures that the white lane markings are visible at the bottom of the region of interest. On the other hand, the 25% uppercut is necessary since only a portion of the entire pavement surface that is visible in the images

can be used for analysis because its upper part presents a less detailed pavement surface (less prominent distresses may not even be visible) compared to the lower part.

Once the height of the trapezium relative to the image's vertical axis is fixed, its width is set to cover the width of the road lane. However, since it can be challenging to keep the vehicle carrying the image acquisition system in the road lane's center, the width of the trapezium was slightly increased. Thus, margins were created for the white lane markings to remain in the region of interest, even when vehicles deviate from the road lane's center. With this in mind, lines were drawn parallel to the white lane markings along the image. By intersecting them with the horizontal lines that define the trapezium's height already specified by the lower and uppercuts, it was possible to determine the image region of interest for the acquired images (Figure 8).

As illustrated in Figure 8, the region of interest is given by a rectangle with height equal to the length of the segment defined by points B1-C1, namely 734 pixels, while the rectangle's width was set equal to the shorter length of the segments defined by points A1-B1 and D1-C1, namely 1208 pixels $\left(\sqrt{(x_{D1} - x_{C1})^2 + (y_{D1} - y_{C1})^2}\right)$. After defining the dimensions of the region of interest in pixels, it was essential to match them to the true dimensions in meters relative to the pavement surface. To perform this, two blue circles were marked in the image as shown on the right side of Figure 8, whose real distance was known to be 3.5 m. Taking this measurement as a reference, the actual dimensions considered for the region of interest were 6.5 m long by 3.5 m wide.

Since two neural networks were used in the proposed system, independently and articulated in parallel, the second pre-processing step comprised two independent paths to conveniently prepare the images for the segmentation and classification steps.

The images resulting from the first pre-processing step (a sample is shown on the right side of Figure 8) only required a resize operation to be processed by the classification network, which expects a smaller image as input to keep reasonable values of required memory and processing time.

For the segmentation step, images of road lane sections approximately 18.5 m long by 3.5 m wide were created from those resulting from the first pre-processing step. An overlapping area between two consecutive images was essential in forming such sections. For this purpose, two options are presented: (i) construction of a panoramic image; (ii) image concatenation. In the first alternative, a panorama was created using the functions `createStitcher()` and `stitch()` from the OpenCV library. Figure 9 presents a panoramic image sample.

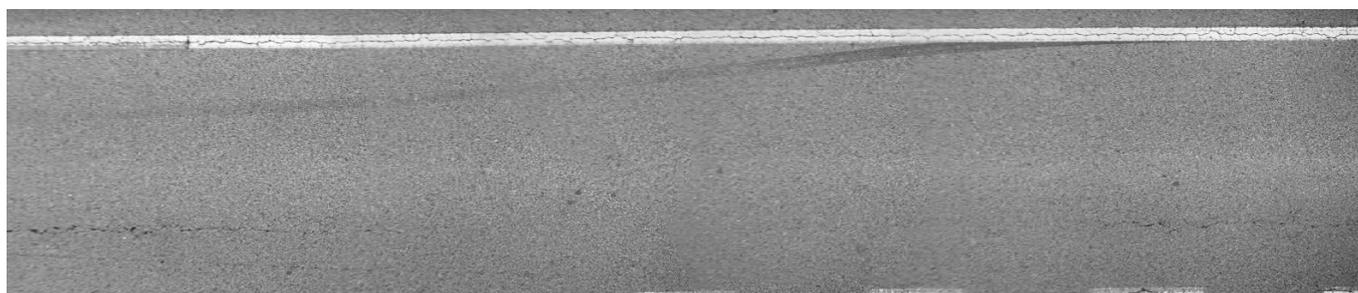


Figure 9. Sample of a panoramic image.

To build a panorama, the images must share an overlapping area featuring a reasonable number of salient points, also known as key points, which are used as reference points to successfully join two or more images into one. However, pavement surface images often exhibit a uniform structure and texture, a most noticeable feature when the road pavement is free of distress, making it very difficult to identify the key points. Therefore, the panoramic construction of a road lane section may fail, and if it does, it is crucial to have an alternative, such as image concatenation, which is a feasible solution because the

overlap area between consecutive images is known. An image concatenation sample for the same road lane section shown in Figure 9 is presented in Figure 10.

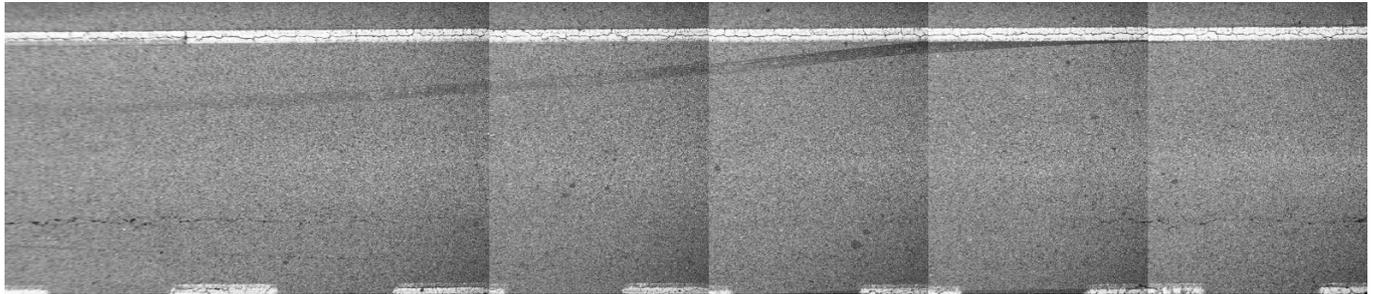


Figure 10. Road section generated through the image concatenation method.

Analyzing the two alternatives, it was decided that panoramic construction would be the first option, and in case of failure, the image concatenation would ensure the generation of the respective road lane section as a fallback solution.

2.3. Segmentation

This step segments the road lane sections generated in the preprocessing step, whose proposed architecture is shown in Figure 11. The road lane sections are submitted to a segmentation neural network, whose architecture is identical to the U-Net [17], illustrated in Figure 12. The difference between U-Net and the proposed neural network for the segmentation task is the dimension of the input image, corresponding to 480×320 pixels. A dataset was built with images from five public datasets, as described in Section 2, to train the proposed segmentation neural network. Furthermore, the parameters used for training were Categorical Crossentropy Loss and Adam optimizer with a learning rate of 0.0001.

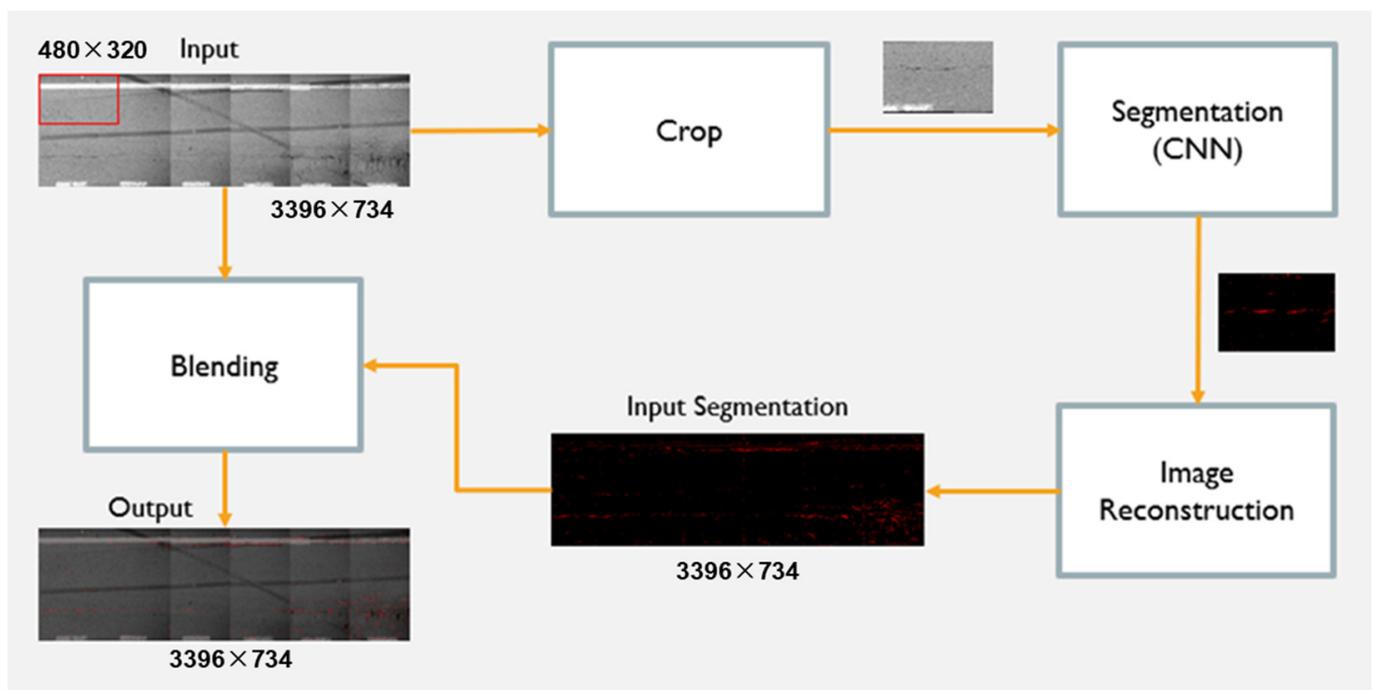


Figure 11. Road lane section processing during the segmentation step.

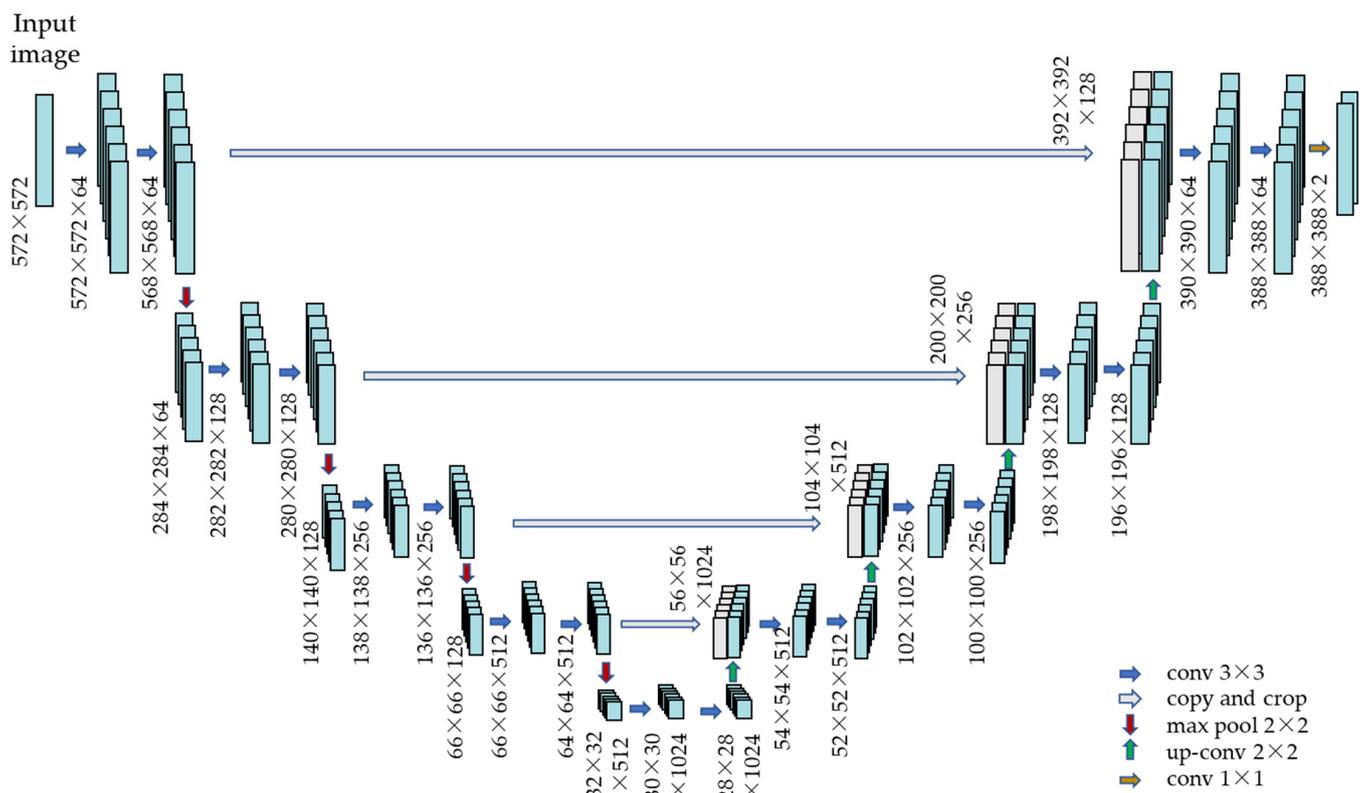


Figure 12. Illustration of the U-Net architecture proposed in [24].

In Figures 13–15, it is possible to observe a road lane section, the corresponding segmentation, and the merging of the two previous images, respectively.

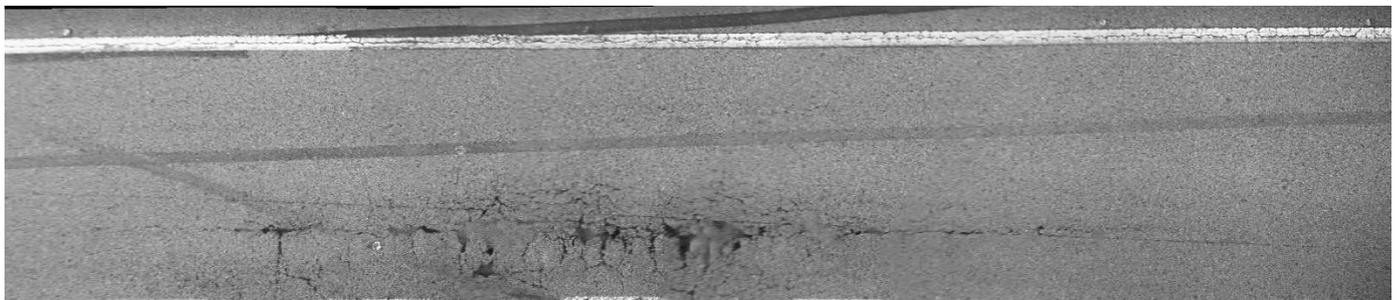


Figure 13. Sample of a road lane section.

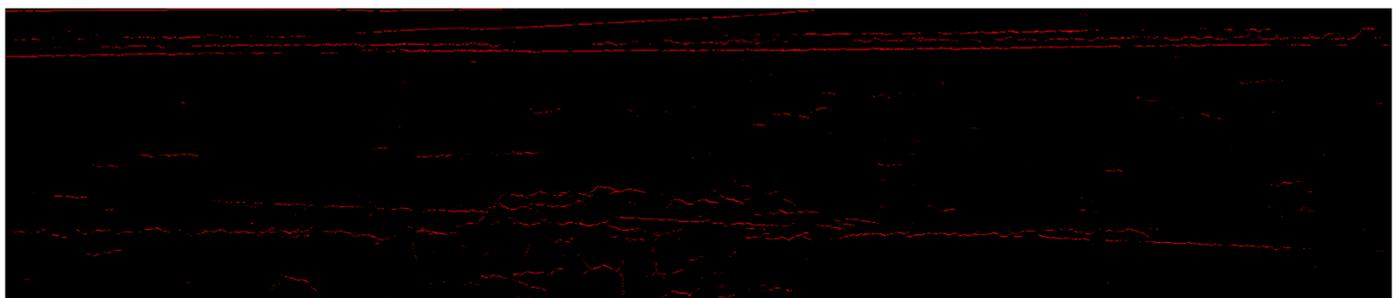


Figure 14. Road lane section segmented by the segmentation neural network.

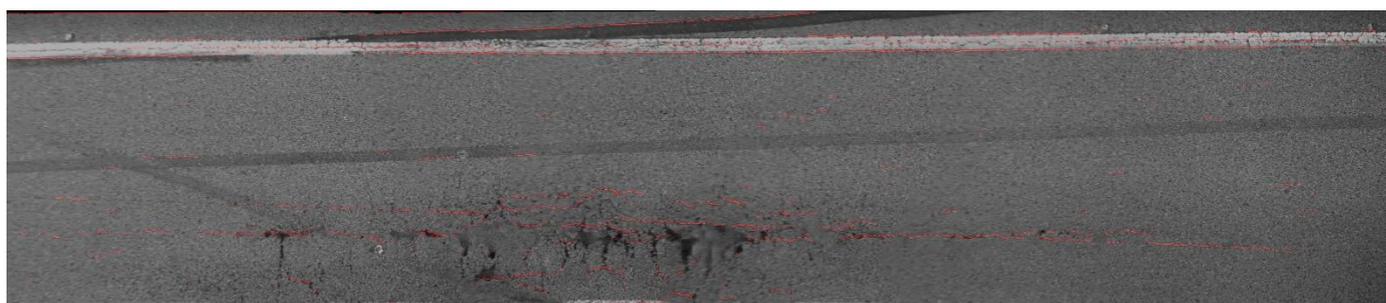


Figure 15. The result of merging the road section with its segmentation.

Figures 14 and 15 show that the segmentation method performs satisfactorily, distinguishing the cracks from the rest of the road pavement surface. However, the model is influenced by white lane markings and tire tracks on the pavement, segmenting some pixels that do not belong to cracks.

2.4. Classification

The classification step aims at label assignment regarding the type of crack present in each image coming from the pre-processing step. The core of this step is a multi-class classification neural network (Figure 16), whose proposed architecture is detailed in Table 3. The output of the classification model is a vector of probabilities. Each vector element contains the probability of a particular label being assigned to the image under analysis, and the one with the highest probability is chosen. The possible labels considered for classification are: (i) alligator crack; (ii) longitudinal crack; (iii) transverse crack; and (iv) non-crack. Figure 17 shows samples of each possible label.

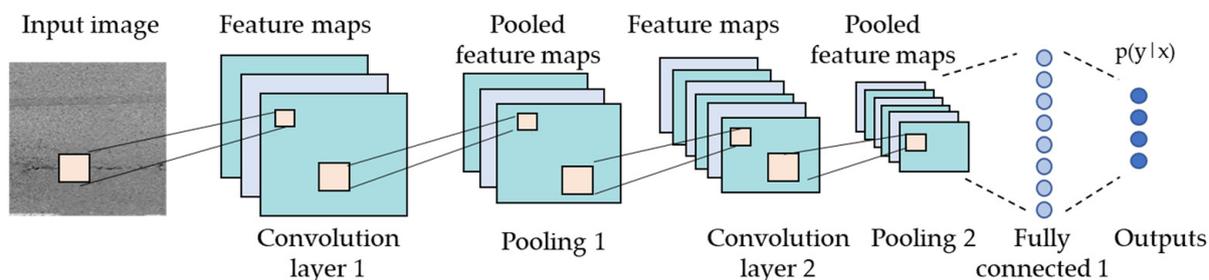


Figure 16. Generic architecture of a multi-class classification neural network.

Table 3. Proposed architecture for the classification network.

Layer	Kernel	Activation Function	Output
Convolutional	5×5	ReLU	(512, 512, 64)
MaxPooling	2×2	-	(256, 256, 64)
Convolutional	5×5	ReLU	(256, 256, 128)
MaxPooling	2×2	-	(128, 128, 128)
Convolutional	5×5	ReLU	(128, 128, 256)
MaxPooling	2×2	-	(64, 64, 256)
GlobalMaxPooling	-	-	256
Dense	-	ReLU	64
Dense	-	Softmax	4

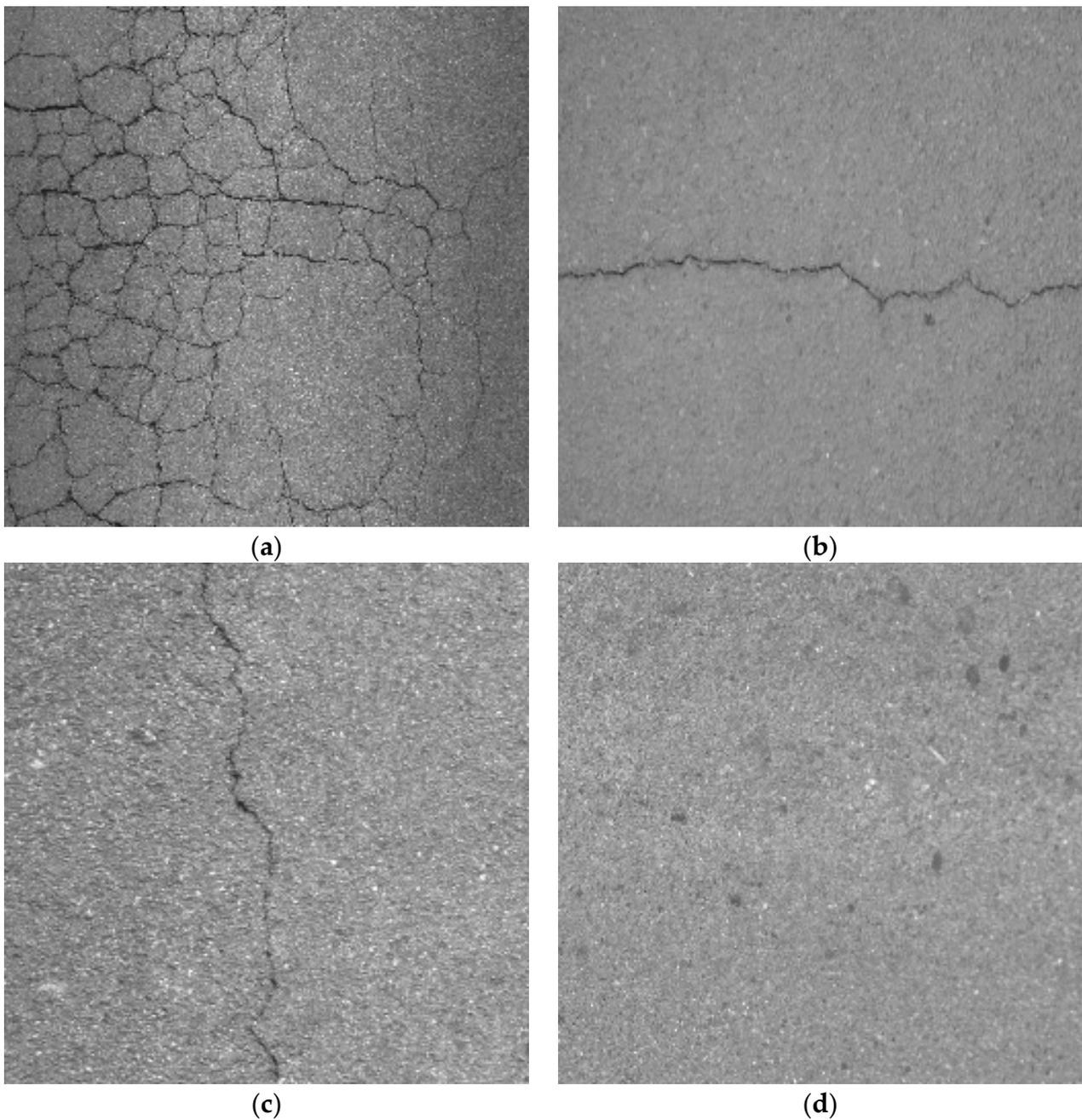


Figure 17. (a) Alligator crack; (b) longitudinal crack; (c) transverse crack; and (d) non-crack samples.

A balanced dataset was built to train the proposed classification neural network by collecting images from the five public datasets presented in Section 2 and by selecting some images captured during a survey of a major road in Portugal, the “IP3”, between the locations of “Souselas” and “Viseu” (a 2 km route section). Thus, 600 images were considered for training, where each of the four classes are represented by 150 images.

The parameters for training the classification model were number of epochs, 30; batch size, 20; optimizer Adam with a learning rate of 0.001; and Categorical Crossentropy Loss, which is described by Equation (1), where \hat{y}_i is the i th scalar value in the model output, y_i is the corresponding target value, and the output size is the number of scalar values in the model output:

$$Loss = - \sum_{i=1}^{outputsize} y_i \cdot \log \hat{y}_i. \quad (1)$$

2.5. Crack Percentage and Analysis of Results

The crack percentage is obtained by dividing the affected area by the total area of a road segment. The cracking level characterization adopted in this paper is structured in three severity levels, as described in Table 4, following the proposal of road analysis experts of the company.

Table 4. Cracking level characterization as proposed by the company’s road analysis experts.

Severity	Description	Affected Area
Level 1	Isolated but noticeable crack (<2 mm)	0.5 m × affected length
Level 2	Open and/or branched longitudinal or transverse cracks	2 m × affected length
Level 3	Alligator cracks	Road lane width × affected length

The first severity level (Level 1) was not considered because it is not guaranteed to detect cracks less than 2 mm in width on pavement surface images captured by the proposed system since a low-cost image sensor was used.

The classification model is also able to distinguish between longitudinal, transverse, and alligator cracks, assigning them a severity level according to the criteria listed in Table 5.

Table 5. Criteria for assigning crack severity labels according to the crack type.

Label	Severity	Affected Area
Longitudinal	Level 2	2 m × 3 m
Transverse	Level 2	2 m × 3.5 m
Alligator	Level 3	3.5 m × 3 m
Non-crack	-	0

Since the proposed system does not incorporate any methodology to measure the crack length, the size of the affected pavement surface area was considered according to the following: (i) if a crack is classified as longitudinal, its maximum length is equal to the length of the non-overlapping region of the image, i.e., 3 m; (ii) if a crack is classified as transverse, its maximum length is equal to the road lane width, i.e., 3.5 m; (iii) if a crack is classified as of alligator type, the affected area corresponds to the non-overlapping area of the image, i.e., 3.5 m wide by 3 m long. It should be noted that the cracking estimate obtained in this way tends to overestimate the actual crack value observed by visual inspection.

Finally, it should be noted that, if there is a large overlap between consecutive images, the same crack will be detected twice. To avoid counting the same crack twice, in a sequence of n classifications of a given type of cracking, only $n - 1$ are considered for the affected area calculation. As an example of calculating the crack percentage, consider the analysis of a 100 m road segment, represented by a sequence of four images containing a longitudinal cracking. The crack percentage contribution of this sequence is obtained by applying three times the longitudinal crack criteria presented in Table 5, where 2 m corresponds to the affected road width considered by the company when a longitudinal crack is detected, and 3 m is the length of the non-overlapping region in each image, as described above. Therefore, the affected area is $3 \times (2 \times 3) = 18 \text{ m}^2$ and dividing by the total area of the road segment ($3.5 \times 100 = 350 \text{ m}^2$) and dividing by the total area of the road segment ($3.5 \times 100 = 350 \text{ m}^2$), resulting in a crack percentage of $(18/350) \times 100 = 5.1\%$.

The overlap between consecutive images allows the implementation of an anomaly detection methodology and the association of an uncertainty value to the crack percentage result computed by the proposed system. For instance, a possible anomaly corresponds to

Table 7. Cont.

Road Segment [Start (m), End (m)]	FT 2 %		FT 3 %		Total %		S—Uncertainty (%)	S—Total with Uncertainty (%)
	T	S	T	S	T	S		
[76,200; 76,300]	41	26	0	3	41	29	4.7	[24.3; 33.7]
[76,300; 76,400]	54	5.1	0	0	54	5.1	1.7	[3.4; 6.8]
[76,400; 76,500]	0	1.7	0	0	0	1.7	1.7	[0; 3.4]
[76,500; 76,600]	0	7.1	0	0	0	7.1	5.4	[1.7; 12.5]
[76,600; 76,700]	18	7.1	0	0	18	7.1	3.4	[3.7; 10.5]
[76,700; 76,800]	57	15.7	0	3	57	18.7	7.1	[11.6; 25.8]
[76,900; 77,000]	18	27.7	0	0	18	27.7	3.4	[24.3; 31.1]

4. Discussion

By analyzing the results in Table 7, it was observed that the crack percentages presented by road analysis experts and the proposed system were similar in most road lane sections. In sections without cracking identified by road analysis experts, the proposed system gave a good response, presenting null or low percentages, as was the case of the following road lane sections: [75,900; 76,000], [76,000; 76,100], [76,400; 76,500], and [76,500; 76,600].

Then, by observing the five road lane sections with cracking identified by road analysis experts, for three of them, the proposed system came close to the percentage obtained by road analysis experts, namely in the following sections: [76,200; 76,300], [76,600; 76,700], and [76,900; 77,000]. The remaining two road lane sections with cracks, more specifically the sections [76,300; 76,400] and [76,700; 76,800], were critical cases in which the percentages obtained by the developed system presented a significant discrepancy, as in both cases the developed system erroneously presented a low percentage. Carrying out a more detailed analysis of these two road lane sections, with the help of the labels that the classification model assigned, it was concluded that the model did not identify the longitudinal cracking found within the white lane markings and sometimes in the road pavement. Hence, the classification model labeled images containing longitudinal cracking as non-crack images. Sample images corresponding to these two road lane sections are shown in Figures 18 and 19.

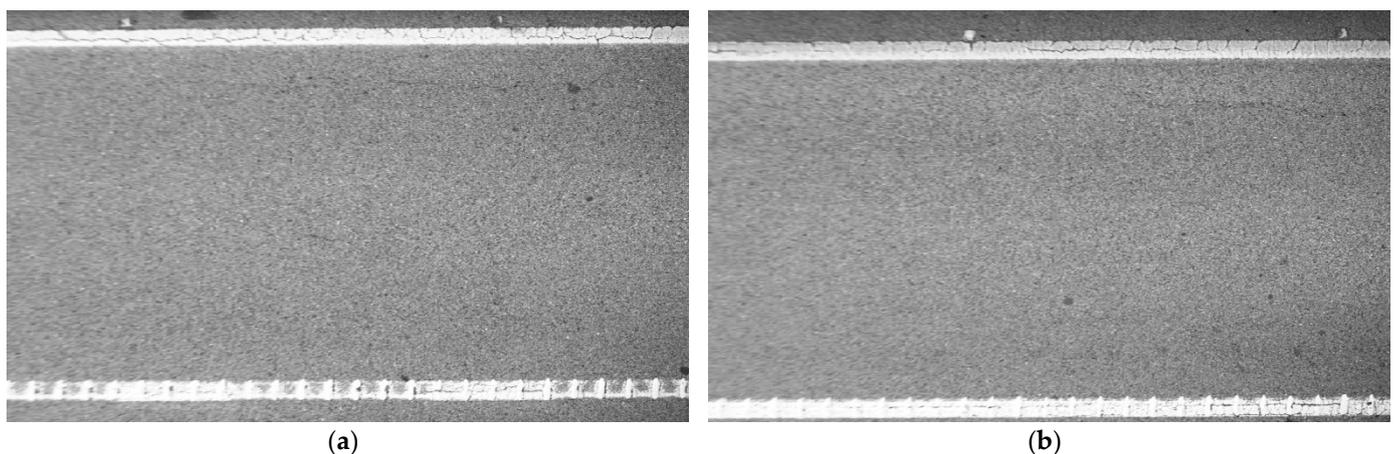


Figure 18. Images relating to the section [76,300; 76,400]: (a) longitudinal crack in the pavement and within the upper white lane marking (the system classified the image as non-crack); (b) longitudinal crack in the pavement (the system classified the image as non-crack).

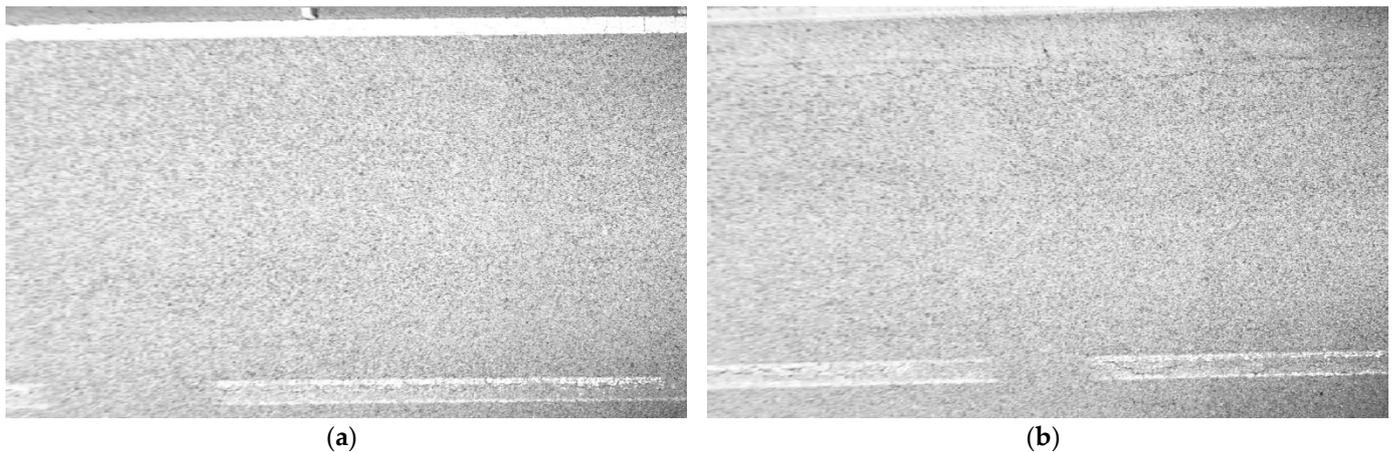


Figure 19. Images relating to the section [76,700; 76,800]: (a) longitudinal crack within the lower white lane marking (the system classified the image as non-crack); (b) longitudinal crack in the pavement and within the lower white lane marking (the system classified the image as non-crack).

By observing the images presented in Figures 18 and 19, it can be seen that the classification model has difficulties in identifying longitudinal cracks within the white lane markings. Sometimes, when the crack is within the white lane markings, it is at an early stage, and it can also be misinterpreted as flaws in the paint due to its wear and tear. Furthermore, the cracks in these images have a reduced width (less than 2 mm) and are barely visible, being below the detection threshold at which the classification model can detect cracking. One of the reasons why these cracks go unnoticed by the classification model is due to the resize procedure performed in the pre-processing step, as this transformation can attenuate the cracking present in the images. It should be noted that, although the classification model did not detect the longitudinal cracking in Figure 18, it was successfully detected by the segmentation model. Therefore, it would be interesting to associate the results of both networks to increase the system's robustness.

The situations highlighted in the previous analysis were cases marked by failures of the proposed system. However, it would also be interesting to compare these cases with some in which the system performed correct classifications to observe the differences. Therefore, in Figure 20, two sample images are presented, which were well classified by the classification model.

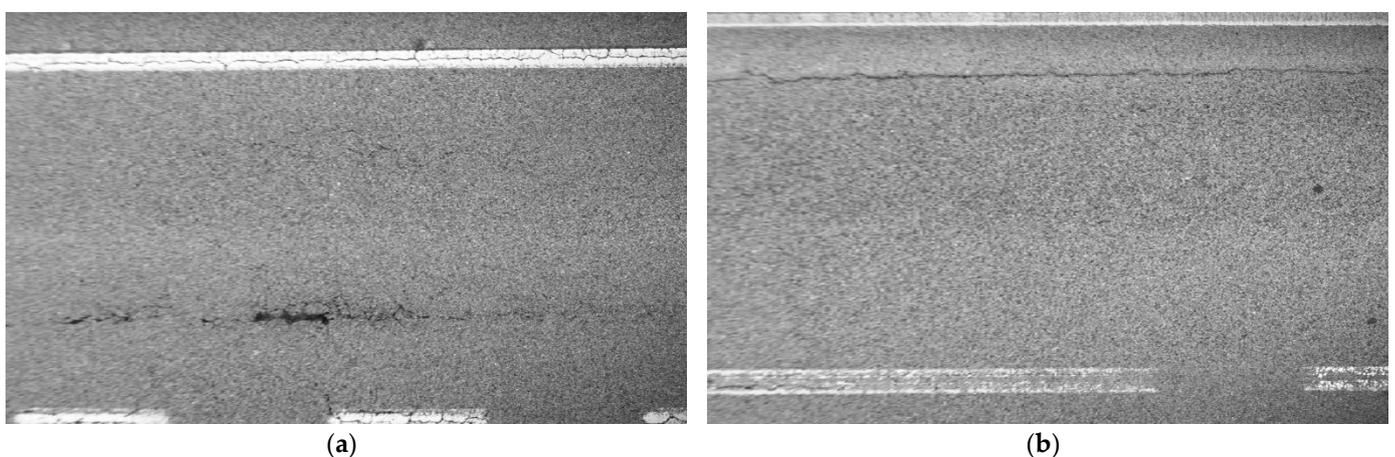


Figure 20. Correct classifications performed by the classification model: (a) image from section [76,200; 76,300] with a longitudinal crack in the pavement surface and within the white lane markings (the system classified the image as longitudinal crack); (b) image from section [76,900; 77,000] with a longitudinal crack in the pavement surface (the system classified the image as longitudinal crack).

Comparing the failures of the proposed system (samples shown in Figures 18 and 19) with cases in which the system correctly classified the longitudinal cracking present in the images (samples shown in Figure 20) verified that the cracking present in the latter sample is sharper and more noticeable. This comparison proves the existence of a detection threshold from which the classification model fails to detect less noticeable cracks, a system outcome that is related to the crack's width, i.e., thin cracks are less noticeable and thus less likely to be detected by the developed system than thick cracks.

Concluding the analysis, since the system is not totally reliable, the system's results should serve as a warning for the subsequent manual analysis of the road sections where possible problems were marked.

5. Conclusions

This paper proposed a low-cost automatic crack detection and classification system. By capturing images of road pavements, the proposed system provides the cracking percentage that characterizes the level of pavement surface degradation. Overall, the proposed system presented promising results, showing the potential to help the experts analyze and inspect road pavements. Moreover, correcting and improving some particularities of the system in its steps can become a very reliable and precise tool. With the use of the proposed system, the analysis and inspection of road pavements can give rise to faster procedures.

The proposed system has advantages and disadvantages. One of its advantages is that if the results are dubious, the system allows the user to visualize their origin. In this case, the system outputs a file with the labels assigned to the images corresponding to each road lane section, either in the case of classification or segmentation methods. This way, identifying errors and correcting the results becomes easier for the system's users. Another advantage is the processing time, which corresponds roughly to ten minutes per kilometer of the analyzed pavement surface.

In conclusion, this work proposed a low-cost and straightforward system that applies what was developed in the scientific community to obtain results from the image analysis of road pavement surfaces. It aimed to bridge the gap that exists between research work and what is demanded by the business world in the automatic detection of cracks in road pavement surfaces.

The proposed system has some margin for improvement as discussed in the following.

The formation of road lane sections could be improved through a more robust and controllable panorama creation function. Another alternative to improve the image concatenation could be to carry out preprocessing to standardize the pixel intensities presented by the images. Consequently, the transition between consecutive images would be smoother, minimizing the possibility of detecting false cracks in the union between concatenated images.

The robustness of the results could be improved by combination of the results provided by the segmentation neural network and those of the classification neural network, as the networks exploit different features, and therefore, the fusion of their results will lead to improved crack detection performance. Regarding image segmentation, adaptive learning approaches should be considered in further developments of the proposed system [35–38].

Another improvement for determining cracking percentages would be to incorporate the measurement of crack lengths. This measurement can easily be obtained from the crack segmentation result.

Finally, it would also be interesting to associate a global position system (GPS) with the acquisition system. In this way, it would be possible to match the GPS position of the acquisition vehicle to each captured image, allowing it to properly synchronize the vehicle speed with the consecutive camera's shutter releases. This strategy would guarantee that all the road lane sections would be correctly surveyed. Moreover, further developments of the system may consider a different image acquisition system, e.g., a camera mounted

on a drone supported by photogrammetric techniques, which allows for overcoming the problem of georeferencing the images.

Author Contributions: Conceptualization, D.I.; methodology, D.I. and P.C.; software, D.I.; validation, P.C. and P.O.; formal analysis, D.I, P.C., P.O. and H.O.; investigation, D.I. and P.C.; resources, P.O.; data curation, D.I. and P.O.; writing—original draft preparation, D.I.; writing—review and editing, P.C. and H.O.; visualization, P.C. and H.O.; supervision, P.C. and P.O.; funding acquisition, P.C. and P.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was partly funded by FCT/MEC under the project UIDB/50008/2020.

Data Availability Statement: Data are available from the authors upon reasonable request and with the permission of Pedro Oliveira (pedro.oliveira@tecnofisil.pt) from Tecnofisil.

Acknowledgments: The authors would like to thank Instituto de Telecomunicações and Tecnofisil, for the opportunity to work together with and for the help provided throughout this project.

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

References

- Mohan, A.; Poobal, S. Crack detection using image processing: A critical review and analysis. *Alex. Eng. J.* **2018**, *57*, 787–798. [\[CrossRef\]](#)
- Ouyang, A.; Luo, C.; Zhou, C. Surface distresses detection of pavement based on digital image processing. In *Computer and Computing Technologies in Agriculture IV*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 368–375.
- Tecnofisil. Tecnofisil—Consultores de Engenharia. Available online: <https://tecnofisil.pt/> (accessed on 21 July 2022).
- Oliveira, H.; Correia, P.L. Automatic Road Crack Detection and Characterization. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 155–168. [\[CrossRef\]](#)
- Cao, W.; Liu, Q.; He, Z. Review of Pavement Defect Detection Methods. *IEEE Access* **2020**, *8*, 14531–14544. [\[CrossRef\]](#)
- Waseem Khan, M. A Survey: Image Segmentation Techniques. *Int. J. Future Comput. Commun.* **2014**, *3*, 89–93. [\[CrossRef\]](#)
- Oliveira, H.; Correia, P.L. Automatic road crack segmentation using entropy and image dynamic thresholding. In Proceedings of the 17th European Signal Processing Conference (EUSIPCO 2009), Glasgow, UK, 24–28 August 2009; pp. 622–626.
- Ayenu-Prah, A.; Attoh-Okine, N. Evaluating Pavement Cracks with Bidimensional Empirical Mode Decomposition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 861701–861708. [\[CrossRef\]](#)
- Wu, Z.; Huang, N.E. A study of the characteristics of white noise using the empirical mode decomposition method. *Phys. Eng. Sci.* **2004**, *460*, 1597–1611. [\[CrossRef\]](#)
- Li, Q.; Zou, Q.; Zhang, D.; Mao, Q. FoSA: F* Seed-growing Approach for crack-line detection from pavement images. *Image Vis. Comput.* **2011**, *29*, 861–872. [\[CrossRef\]](#)
- Tarjan, R.E. *Data Structures and Network Algorithms*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1983.
- Zhang, L.; Yang, F.; Daniel Zhang, Y.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712.
- Aslan, O.D.; Gultepe, E.; Ramaji, I.J.; Kermanshachi, S. Using Artificial Intelligence for Automating Pavement Condition Assessment. In Proceedings of the International Conference on Smart Infrastructure and Construction 2019 (ICSIC), Cambridge, UK, 8–10 July 2019; pp. 337–341.
- Anand, S.; Gupta, S.; Darbari, V.; Kohli, S. Crack-pot: Autonomous Road Crack and Pothole Detection. In Proceedings of the 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 10–13 December 2018; pp. 1–6.
- David Jenkins, M.; Carr, T.A.; Iglesias, M.I.; Buggy, T.; Morison, G. A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2120–2124.
- Kang, K.; Wang, X. Fully Convolutional Neural Networks for Crowd Segmentation. *arXiv* **2014**, arXiv:1411.4464.
- Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
- Zhang, K.; Zhang, Y.; Cheng, H.-D. CrackGAN: Pavement Crack Detection Using Partially Accurate Ground Truths Based on Generative Adversarial Learning. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1306–1319. [\[CrossRef\]](#)
- Liu, C.; Zhu, C.; Xia, X.; Zhao, J.; Long, H. FFEDN: Feature Fusion Encoder Decoder Network for Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15546–15557. [\[CrossRef\]](#)
- Ma, D.; Fang, H.; Wang, N.; Zhang, C.; Dong, J.; Hu, H. Automatic Detection and Counting System for Pavement Cracks Based on PCGAN and YOLO-MF. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22166–22178. [\[CrossRef\]](#)
- Han, C.; Ma, T.; Huyan, J.; Huang, X.; Zhang, Y. CrackW-Net: A Novel Pavement Crack Image Segmentation Convolutional Neural Network. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 22135–22144. [\[CrossRef\]](#)

22. Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1525–1535. [[CrossRef](#)]
23. 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.
24. U-Net Architecture. Available online: <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/> (accessed on 23 July 2022).
25. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *162*, 94–114. [[CrossRef](#)]
26. Camera Teledyne Lumenera lt16059h. Available online: <https://www.znjtech.cn/resources/datasheets/teledyne-lumenera/lt16059h-datasheet.pdf> (accessed on 23 July 2022).
27. CrackForest Dataset. Available online: <https://github.com/cuilimeng/CrackForest-dataset> (accessed on 23 July 2022).
28. AigleRN. Available online: <http://telerobot.cs.tamu.edu/bridge/Datasets.html> (accessed on 23 July 2022).
29. CrackTree260. Dataset and Ground Truth. Available online: <https://1drv.ms/f/s!AittnGm6vRKLyiQUk3ViLu8L9Wzb> (accessed on 23 July 2022).
30. CRKWH100. Dataset, Ground Truth. Available online: https://1drv.ms/f/s!AittnGm6vRKLgLyfiCw_C6BDeFsP (accessed on 23 July 2022).
31. CrackLS315. Dataset, Ground Truth. Available online: <https://1drv.ms/u/s!AittnGm6vRKLg0HrFfjNhP2Ne1L5?e=WYbPvF> (accessed on 23 July 2022).
32. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *120*, 122–125.
33. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: <https://www.tensorflow.org/> (accessed on 23 July 2022).
34. Keras: Deep Learning for Humans. Available online: <https://github.com/fchollet/keras> (accessed on 23 July 2022).
35. Guo, P.; Zhang, W.; Li, X.; Zhang, W. Adaptive Online Mutual Learning Bi-Decoders for Video Object Segmentation. *IEEE Trans. Image Process.* **2022**, *31*, 7063–7077. [[CrossRef](#)] [[PubMed](#)]
36. Kerdvibulvech, C. A methodology for hand and finger motion analysis using adaptive probabilistic models. *J. Embed. Syst.* **2014**, *2014*, 18. [[CrossRef](#)]
37. Yang, F.; Yuan, X.; Ran, J.; Shu, W.; Zhao, Y.; Qin, A.; Gao, C. Accurate Instance Segmentation for Remote Sensing Images via Adaptive and Dynamic Feature Learning. *Remote Sens.* **2021**, *13*, 4774. [[CrossRef](#)]
38. Bourouis, S. Adaptive variational model and learning-based SVM for medical image segmentation. In Proceedings of the ICPRAM 2015–4th International Conference on Pattern Recognition Applications and Methods, Lisbon, Portugal, 10–12 January 2015; pp. 149–156.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.