applied sciences

MDPI

Article

An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment

Gilberto Ochoa-Ruiz [1,*], Andrés Alonso Angulo-Murillo [2], Alberto Ochoa-Zezzatti [3], Lina María Aguilar-Lobo [2], Juan Antonio Vega-Fernández [2] and Shailendra Natraj [4]

[1] School of Engineering and Sciences, Tecnológico de Monterrey, Guadalajara 45201, Mexico
[2] Maestría en Cs Computacionales, Universidad Autónoma de Guadalajara, Zapopan 45129, Mexico; alonsoangulom@outlook.com (A.A.-M.); lina.aguilar@edu.uag.mx (L.A.-L.); javega@edu.uag.mx (J.A.V.-F.)
[3] Doctorado en Tecnología, Universidad Autónoma de Ciudad Juárez, Ciudad Juárez 32315, Mexico; alberto.ochoa@uacj.mx
[4] Vidrona LTD, Edinburgh HMGP+42, Didcot OX11 0QX, UK; shailendra@vidrona.com
* Correspondence: gilberto.ochoa@tec.mx; Tel.: +52-11-622-8559

Received: 13 March 2020; Accepted: 22 April 2020; Published: 8 June 2020

Featured Application: The proposed solution is intended to serve as the acquisition system for a physical asset management tool for big data analytics and smart cities applications.

Abstract: The analysis and follow up of asphalt infrastructure using image processing techniques has received increased attention recently. However, the vast majority of developments have focused only on determining the presence or absence of road damages, forgoing other more pressing concerns. Nonetheless, in order to be useful to road managers and governmental agencies, the information gathered during an inspection procedure must provide actionable insights that go beyond punctual and isolated measurements: the characteristics, type, and extent of the road damages must be effectively and automatically extracted and digitally stored, preferably using inexpensive mobile equipment. In recent years, computer vision acquisition systems have emerged as a promising solution for road damage automated inspection systems when integrated into georeferenced mobile computing devices such as smartphones. However, the artificial intelligence algorithms that power these computer vision acquisition systems have been rather limited owing to the scarcity of large and homogenized road damage datasets. In this work, we aim to contribute in bridging this gap using two strategies. First, we introduce a new and very large asphalt dataset, which incorporates a set of damages not present in previous studies, making it more robust and representative of certain damages such as potholes. This dataset is composed of 18,345 road damage images captured by a mobile phone mounted on a car, with 45,435 instances of road surface damages (linear, lateral, and alligator cracks; potholes; and various types of painting blurs). In order to generate this dataset, we obtained images from several public datasets and augmented it with crowdsourced images, which where manually annotated for further processing. The images were captured under a variety of weather and illumination conditions and a quality-aware data augmentation strategy was employed to filter out samples of poor quality, which helped in improving the performance metrics over the baseline. Second, we trained different object detection models amenable for mobile implementation with an acceptable performance for many applications. We performed an ablation study to assess the effectiveness of the quality-aware data augmentation strategy and compared our results with other recent works, achieving better accuracies (mAP) for all classes and lower inference times (3× faster).

Keywords: asphalt damage; dataset; deep learning; object detection; asset management

1. Introduction

Research on damage detection of road surfaces using artificial intelligence techniques has seen an increased interest in recent years [1–4], particularly with the arrival of new digital transformation paradigms (i.e., smart cities). Road maintenance efforts and follow ups had been difficult to automate in the past, and yet, the problem is of paramount importance, as demonstrated by the widespread push and adoption of protocols and standards in many countries to tackle these issues.

Nonetheless, the automated inspection and follow up of road damages remains a complex problem, as most efforts do not consider effective digitization strategies, and thus governments and agencies lack mechanisms to maintain accurate and up-to-date databases of the road conditions. Another issue is the lack of experts that can assess the state and spread of several structural damages, as oftentimes, the evaluation can be very subjective. To make matters worse, traditional methods for collecting samples in the field are time-consuming and cost-intensive. Therefore, several research and commercial efforts have been conducted to aid government agencies to automate the road inspection and sample collection process, making use of technologies with varied degrees of complexity [5].

Initial approaches made use of mobile measurement system (MMS), which combined various sensors (i.e., inertial profilers, scanners), in tandem with sophisticated imaging techniques for automating the data collection process (although not necessarily identifying particular types of damages in real time). The information gathered by MMS is usually fed to artificial intelligence methods as georeferenced information to produce a "digital twin" representation of the collected data, fostering big data approaches for road damage prevention and prognosis.

Although these approaches have shown promising results [6], many of them are targeted by design to specific types of damages (i.e., cracks) and made use of rather small datasets. In many instances, this was because of the lack of large and diverse datasets and to sub-optimal feature extraction and machine learning algorithms, leading to solutions that did not perform well in very complex scenarios (i.e., for images under varying degrees of illumination, camera perspective, among others). Nonetheless, recent efforts in the domain have seen a surge in performance owing to the adoption of deep learning-based computer vision acquisition systems. Such artificial intelligence-based systems are in general less costly than other technological choices and, if implemented properly, they can represent a cost-effective solution for agencies or governments with low budgets.

Furthermore, these AI-based approaches can be augmented if they are coupled with mobile and cloud computing components; the former can be exploited for implementing lightweight mobile acquisition systems, while the latter can be used for storing the information captured and processed by these edge sensors for carrying out further big data analyses. For instance, as the captured information is geo-localized and digitized, the road damages can be tracked over time if implemented within an asset management tool for data analytics (i.e., planning, allocation of budgetary resources, among others). This approach suits particularly well with modern digital transformation paradigms and can be readily deployed as business model and a means for improving governmental decisions. For instance, the United Kingdom launched the Digital Roads Challenge to attain this goal.

In this article, an initial step towards a road damage monitoring and management tool is introduced. The research presented herein is an extension of previous work in road damage detection; in this article, we introduce and make use of a new and wider-reaching dataset, consisting of various types of asphalt damages (see Figure 1: longitudinal cracks, alligator cracks, potholes, bumps, patches). To the best of our knowledge, this database is one of the most complete in the recent literature.

Some of the most representative road damages are shown on Figure 1. These samples depict mainly structural damages on the asphalt, but do not take into consideration wear on traffic signs, as other works in the infrastructure management literature. Previous works on road damage identification have tackled the detection and classification of individual types and classes, and only very recently, some works have addressed the problem of detecting multiple classes and instances in real time, which usually requires the use of modern Deep Learning (DL)-based detectors or semantic segmentation schemes.

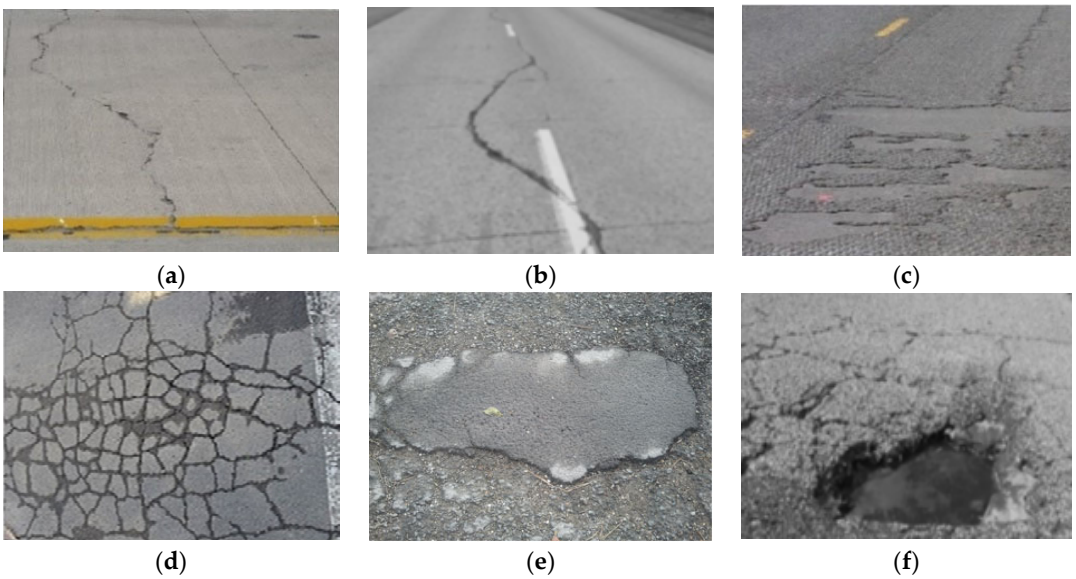


Figure 1. Some samples of road damages tackled by road inspection systems and agencies: (**a**,**b**) linear cracks, (**c**) peeling, (**d**) alligator cracks, (**e**) patches, and (**f**) potholes.

In this work, we made use of a very large road damages dataset [7], which, as with other recent works, was used to train a state-of-the-art generic object detector. As briefly mentioned above, in order to be practically useful in the field, the deep learning-based road damage acquisition system should run in real time, preferably in constrained devices (i.e., smartphones). We carried out extensive experiments using various real-time object detectors to evaluate the best compromise between model complexity and performance (in terms of accuracy) in all the classes in Figure 1, which is one of the most comprehensive in the literature to the best of our knowledge.

The rest of this paper is organized as follows. In Section 2, we analyze recent works in asphalt damage assessment that make use of different imaging modalities, but we pay particular attention to approaches that deploy modern artificial intelligence algorithms (DL-based detectors). In Section 3, we present our road damage database and the first stage for asphalt physical management: the image acquisition process via detection; we also discuss the methodology for validating our method (including several ablation studies). Afterwards, in Section 4, we analyze these results in comparison with other works found in the literature and we discuss some promising avenues of future research.

2. Related Works

The automation of any task using computer vision depends on the use of robust real-time computational methods with a great level of intelligence. However, in the context of road damage inspection systems, most of the initial approaches were originally devised as a mechanism for aiding the road inspectors as simple acquisition instruments using image processing techniques, whereas other systems relied upon heavily instrumented MMSs. Both approaches were rarely real-time and were constrained by capabilities of the machine learning algorithms of the time. For a more comprehensive survey of the various acquisition techniques, there exist excellent studies in 3D imaging [8], remote sensing [9], and computer vision and image processing in general [10,11].

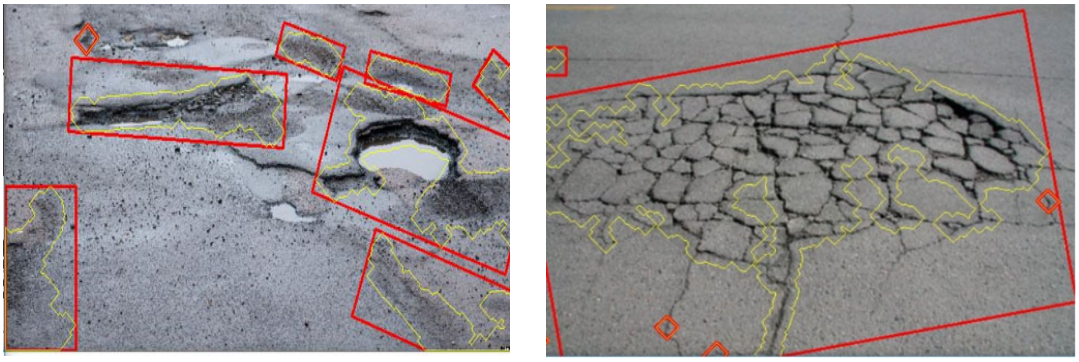
It has been recognized for some time that computer vision (CV) approaches were among the best technological choices for implementing asset management tools, but several technical limitations hampered their widespread adoption. In particular, the biggest hurdle for automating asphalt damage detection and classification is to consistently achieve high classification accuracies (and high recalls) for a typically large number of classes, in very challenging environmental conditions (i.e., wide changes in illumination and camera pose, weather conditions, among others). Despite these issues, various computer vision-based methods for automatically classifying individual structural damages can be found in the literature.

This lack of more comprehensive approaches stemmed from the shortage of large road damage datasets (although the Kitti [4] dataset has been used by some researchers). This problem has been addressed in recent years by academics working in the domain, although it remains an open problem, as it is common to find unbalanced classes for some types of structural damages. Regardless of the aforementioned problems, these datasets were deployed by many studies to implement and test artificial intelligence-based methods, using conventional image features and machine learning models such as support vector machines for the identification of cracks [12] and potholes [13].

As larger sets of images became more common, DL-based methods to address the classification of road damages have become the norm in the field. It is widely recognized [14] that algorithms based on such architectures have obtained outstanding results in several image analysis and classification problems. In this sense, several studies have been carried out that aim to take advantage of the capabilities of neural networks' architectures for improving the detection and classification of road asphalt damages, attaining very competitive results for certain classes such as asphalt cracks [15,16].

An additional benefit of using such CV-based methods over other approaches (i.e., MMS) is that cheaper imaging devices can be used for acquiring the samples. Technologies based on mobile devices were explored previously, which made use of traditional algorithms such as Haar-cascade classifiers [17,18], but these works did not make comparisons with DL-based acquisition systems.

Despite the good results of these approaches, they suffer from the fact that custom datasets were employed and the results are not scalable or easily replicable, as the tool was intended for "static analyses" (simple image capture and classification). For instance, Figure 2 shows examples of image acquisition of road damage instances using an LPB-based approach on OpenCV.



(a) (b)

Figure 2. Examples of classification of two types of damages using a simple computer vision (CV)-based LBP detector. A high number of false positives (FPs) can be observed in the examples.(**a**) potholes (**b**) alligator cracks

Conversely, recent strides in deep learning techniques have enabled the deployment of very efficient and resource-constrained object detection systems in mobile devices [19, 20] and have started to be deployed on road damage acquisition systems [21,22]. These progresses were possible owing to the introduction single stage detectors such as YOLO or RetinaNet (for a detailed discussion in this topic, see [18] for an excellent survey). Two-stage detectors such as Fast-RCNN could achieve high accuracies, but with the cost of high inferences times, making them unsuitable for deployment on resource-constrained devices [21].

On the other hand, one-stage detectors suffered from low accuracies, but offered very low inference times; this issue has steadily been tackled by the DL community and recent works (both academic and commercial) have made use of these object detectors for implementing real-time acquisition systems on mobile devices [23–25], as depicted in Figure 3a,b.

(**a**) (**b**)

Figure 3. Two different types of acquisition systems implemented using mobile devices: (**a**) road acquisition system by the University of Tokyo [25] and (**b**) integrated platform by RoadBotics [26].

However, the introduction of large road damage datasets cannot be underestimated, as they are unavoidable for training data hungry DL models. For instance, authors at the University of Tokyo [5] introduced one the first large road damage datasets; although it contains several classes, some are still underrepresented, which has limited the attainable performances in previous works [23,25]. Thus, one of the contributions of our work is to complement the original dataset with more samples per class, in particular for the potholes and alligator cracks classes, which are very common in tropical regions and countries like Mexico, and thus essential for a great variety of case uses.

Table 1 presents a summary of the works found in the literature, identifying potential areas of opportunity. For road damages classes found in previous works, we made use of the road damage classification proposed by Maeda et al. [25], shown in Table 2. The IDs are deployed in Japan to identify types of structural damages and have been introduced by the researchers in the study to categorize road instances in their dataset. The University of Tokyo dataset is one of the largest available in the literature and, therefore, it has been widely used in road damage classification algorithms, making use of both traditional models and deep learning architectures (using Single Shot Detectors (SSDs) such as RetinaNet in [23] and MobileNet in [24]).

Table 1. Summary of the state of the art and classification. ML, machine learning; AI, artificial intelligence; SVM, support vector machine, RF, Random Forest; DL, Deep Learning; LBP, Locally Binary Patterns.

Work	Supported Classes	ML–AI Algorithm	Mobile Deployment	Real-Time Inference
Oliveira [2]	D10	Clustering		
Radopoulu [3]	D40	Texture–SVM		
Seung-Ki [6]	D40	Img. Processing		
Hoang [12]	D10	SVM–RF		
Hoang [13]	D40	SVM–RF		
Cha [17]	D10	DL–Custom		
Tedeschi [17]	D10, D20, D40	LBP Detector	✓	
Siribor [17]	D10, D20, D40	Texture–SVM	✓	
Pereira [24]	D40	DL–Custom		
Ale [23]	All	DL–RetinaNet	✓	✓
Maeda [25]	All	DL–MobileNet	✓	✓

Table 2. Asphalt damage classes IDs used in [25] and their corresponding definitions.

Damage Type		Description	Class Name	Num Images
Crack	Linear	Car wheel mark	D00	2678
		Cement joint	D01	3789
		Zebra crossings	D10	742
		Cement or asphalt construction joint	D11	636
	Alligator	Partial, overall pavement	D20	2541
		Rutting, bump, peeling or pothole	D40	409
Other damages		White lane wearing	D43	817
		Zebra cross wearing	D44	3733

As mentioned above, one of the disadvantages of this and other datasets is a general lack of representation for some important classes such as potholes. We consider this an important issue that we aim to solve though the introduction of an enlarged dataset that was created using a quality aware image filtering process [27]. In what follows, we will discuss how the dataset was created and we will analyze how the used strategy has improved the results over those in the state-of-the-art; in order to validate if any ameliorations obtained were by following our method, we carried out several ablation studies, which will be discussed in detail too. Finally, we will see some important performance metrics that we use to evaluate the advantages of the solution proposed in our work.

3. Materials and Methods

As described in previous work [28], the integration of artificial intelligence algorithms into digital transformation approaches involving tools such as data analytics for physical asset monitoring and management was only a question of time, and many governmental agencies and commercial initiatives have been launched for exploiting the information collected by road damage acquisition systems and stored digitally into databases in various ways.

The rationale for doing so is twofold, as depicted in Figure 4. First, to leverage the recent advances made in several disciplines (artificial intelligence and Internet of Things mainly) for conceiving novel platforms to facilitate the labor of road inspectors during the inspection on a day-to-day basis. For instance, the computer vision algorithms discussed in Section 2 can run on edge devices (either smartphones or mounted in drones or cars) for implementing lightweight mobile measurement systems that can be used as well as gateways for sending georeferenced info to the cloud. Second, this information can be exploited for creating large datasets or digital representations of a road or street conditions (but in a streamlined fashion as metadata, instead of images), from medium to large cities, effectively creating what is known as "digital twins" (Figure 3b).

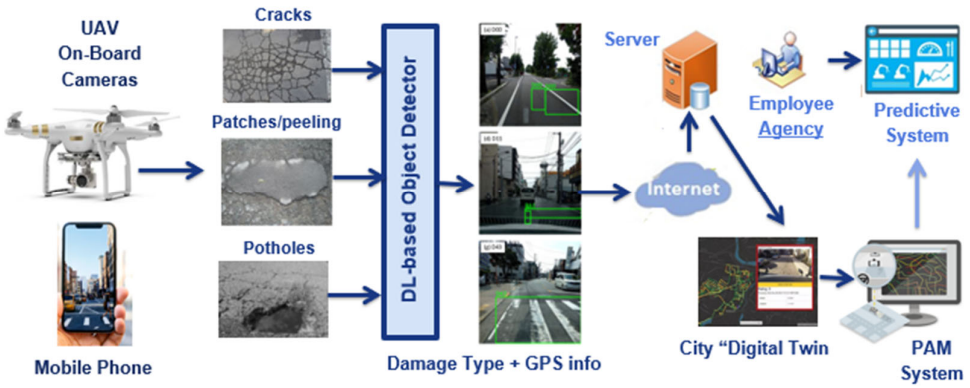


Figure 4. Architecture of a prototypical road damage acquisition and physical asset management system based on artificial intelligence (AI) and big data technologies.

The information can be aggregated over time for bridging the digital transformation gap that is currently plaguing the road inspection process. These platforms could help to reduce the burden for

the municipalities and federal government, with the benefit of providing actionable insights about their assets. Platforms with these capabilities have emerged as commercial solutions (RoadBotics [26]), but their service is more oriented towards a general assessment of the street qualities.

In this paper, we focus on the creation of the dataset and training the AI algorithm in charge of performing the road damage detection and identification; the road damage acquisition system is already in use as a mobile app (similar to the one on the left side of Figure 4). The other components of the system will be described in a subsequent work as they are currently under development.

3.1. Sample Collection and Characteristics of the Dataset

The dataset proposed in [25] (for which the classes are summarized in Table 2) is one of the largest in the literature, but some important instances are combined with other very similar samples and, furthermore, they are poorly represented; this is the case of the potholes class, which stems from the fact that these damages are quickly repaired in Japan. We trained several generic object detectors using this data and we quickly realized that they failed to detect potholes in a reliable manner.

This represents a major issue for the follow up of road damages of this kind (at several rates of deterioration), which is indeed one of the most pervasive types of structural damages found in roads and needs to be tracked effectively. Therefore, we contribute to the state-of-the-art by proposing an extended dataset that incorporates more samples for some of the classes' IDs introduced by the authors in [25]. The main idea is to mitigate the class imbalance present in their work, as depicted in Table 3, which compares the number of class instances in each dataset. A sample of the expanded dataset containing multiples class labels in shown in Figure 5. We collected more examples of the D00, D10, and D11 classes (longitudinal cracks of various types) and for D40 (potholes), for which several hundred examples have been added, as shown in the third row of Table 3. The samples were obtained from several Mexican cities using conventional mobile cameras and annotated using the Open Label Master to train classifiers such as YOLO, MobileNet, and RetinaNet. In some cases, it was necessary to modify the annotations using scripts to conform with the format required for the algorithms.

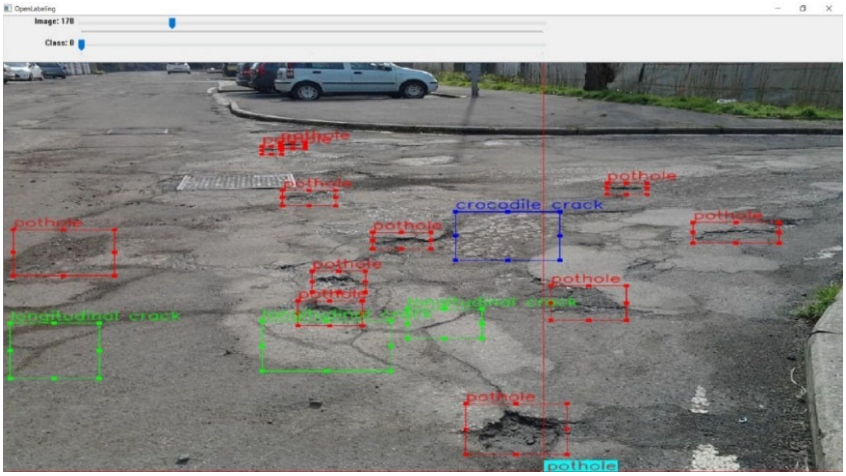


Figure 5. An example of a training instance in our dataset, labels were manually annotated using Open Label Master; note the prominence of the pothole class in this image.

Table 3. Comparison of the between the dataset presented in [26] and our extended database in terms of the number of instances per class.

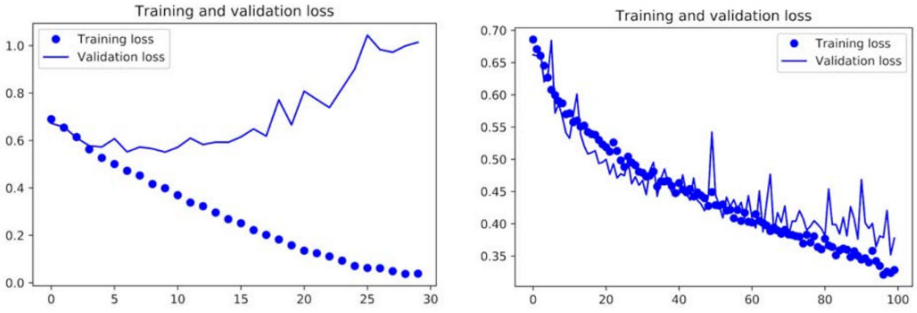
Dataset/Class	D00	D01	D10	D11	D20	D40	D43	D44
Maeda [25]	2678	3789	742	636	2541	409	817	3733
Extended [26]	**2978**	3789	**1042**	**1036**	3341	**1609**	817	3733

The choice of the acquisition system was dictated for both technical and tactical reasons; it must be noted that in many jurisdictions, installing an imaging device on top of a car is considered violation of the law [25], and thus it was avoided in this work. We performed a thorough comparative quantitative analysis with other studies in the state-of-the-art to assess if any improvements in performance were attainable using our extended dataset. In what follows, we will describe the methodology used for training and testing a set of "object detectors", whose performance is of special interest as it represents the foundation of any minimally viable CV-based MMS. The object detectors were chosen for their low memory footprints, an absolute must in edge devices.

3.2. Data Pre-Processing and Filtering

The approach presented here is a multi-class classification problem, in which we take the eight classes of the proposed dataset and train an optimized object detector model (MobileNet, RetinaNet, but it could be others) to implement a road damage acquisition system. However, training deep learning-based object detectors is not a trivial task; one of the major issues of deep learning, when working with relatively small datasets such as the one used in this study, is the problem of overfitting.

Data augmentation (DA) is a manner for "synthetically" generating novel training data from the existing examples in a dataset. They are generated by submitting images in the latter to several image processing operations that yield believable looking images. The process of DA can significantly reduce the validation loss, as shown in Figure 6, where the plots for the training and validation for two identical ConvNets are shown—one without using data augmentation and the other using the strategy. Using data augmentation clearly combats the overfitting problem, the effect of which can be observed in the left side of the figure (the validation loss decreases and then increases again, while the training loss decreases for the training set). The right side of the figure shows how the validation loss can be regularized using data augmentation, among other strategies.



(**a**) (**b**)

Figure 6. Train and validation loss for two identical DNNs, (**a**) without using data augmentation and (**b**) with data augmentation. The loss clearly shows how data augmentation (DA) prevents overfitting.

The data augmentation approach has proven to be a successful method in several computer vision applications, but there are some major considerations when applying these transformations for tasks such as image classification and object detection. The main issue is related to the potential poor quality of many of the training examples, which can lead to suboptimal models. Such models could introduce a great number of false positives or, even worse, misclassify a new training example, as thoroughly documented by the authors in [27].

The authors of that study carried out a battery of tests in which they evaluated the performances of some well-known ConvNets (i.e., VGG16 and Google Net, among others) using images corrupted with different kinds of image quality distortions: compression, low contrast, blurring, and Gaussian noise. The authors found that the networks are resilient to compression distortions, but are greatly affected by the other three.

These findings have led to a great deal of research in this domain, with some works concentrating on the effect of such image quality distortions for several computer vision tasks based

on deep neural nets [29]. To the best of our knowledge, no image quality assessments have been reported by previous works with the road damages datasets, which typically include a great deal of poor-quality images, as depicted in Figure 7. Some of the images have very low contrast, while others present blurring or an overall lack of sharpness, and others suffer extreme saturation.

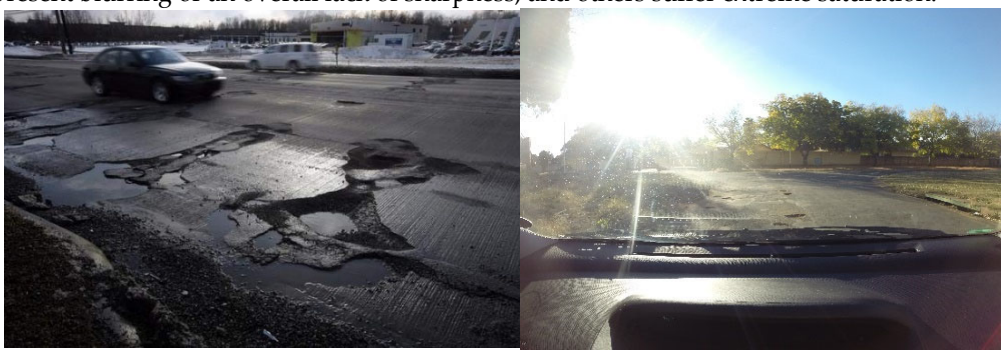


Figure 7. Examples of low-quality images deemed not useful for classification, which are automatically filtered during the pre-processing stage.

Therefore, in addition to the data augmentation strategy outlined above, the second contribution of this paper is the implementation of an image quality assessment to determine if a training example can be considered a good candidate for the data augmentation phase and for further processing (i.e., for the ConvNet training process). Some of the considerations for evaluating the image quality, as well as some of the image preprocessing algorithms to determine it, will be discussed as follows.

As mentioned above, the performance of the ConvNet models can be severely affected by the quality of the input images; the authors in [26] performed individual tests on for Gaussian noise, blur, compression, and low contrast. Individually, even for moderate amounts of noise and blur, the accuracy of the network decreased significantly, and one can only assume that the combination of various of these image distortions can yield even poorer results.

Therefore, in this paper, we implement a "filtering" or sample preselection process based on several traditional image processing metrics [30] to estimate whether an image can be considered for further processing or not, namely, the signal-to-noise ratio (SNR), the blurring index, and the variance of the image (for determining the contrast). Typically, the following parameters are correlated: an image with high levels of noise present and low contrast and low blurring.

The images in the extended dataset were then first pre-processed using traditional image processing techniques. (1) We calculate the blurring index using a spectral blur estimation operator. (2) For estimating the contrast, we make use of the variance as a measure of the quality of the image, using the Michelson Contrast Ratio, as shown in Figure 8 for an example in a previous work in the context of medical imaging [31], filtering out images with high variance. (3) Additionally, we performed an SNR test using the methods proposed by Lagendijk and Biemond [30] to determine whether or not an image is well suited for the data augmentation and training phases.

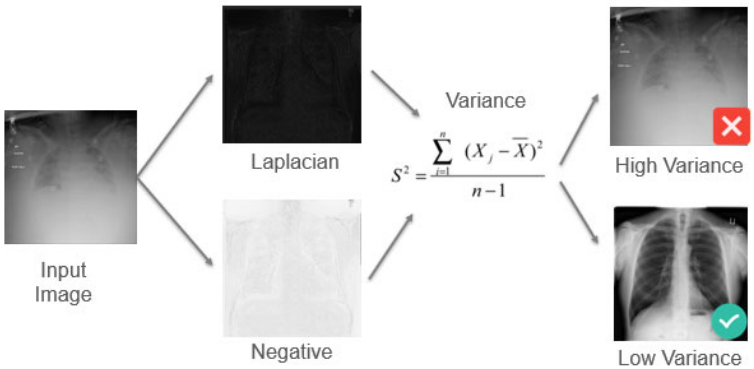


Figure 8. Process for discarding images with low contrast via the calculation of the variance [31].

Once these tests have been carried out, we perform image restoration procedures for removing image distortion like blurring (using Wiener Convolution) and low contrast (using CLAHE, contrast limited adaptive histogram equalization) and repeating the SNR test to determine if the corrected images can be used for further processing.

This process was applied to the 18,345 images (divided in the eight classes in Table 2) yielding a reduced set of filtered images shown in the third row of Table 4. This set of images undergoes the data augmentation process, applying various geometric (flip, rotation), photometric (histogram equalization, changes in hue), and noise (saturation, blurring) transformations. The augmented dataset is comprised of 251,250 images.

Table 4. Quantity of road damage samples in each class for the dataset presented in [25], our extended dataset [28], and the filtered and augmented datasets.

Dataset/Class	D00	D01	D10	D11	D20	D40	D43	D44
Maeda [25]	2678	3789	742	636	2541	409	817	3733
Extended [28]	2978	3789	1042	1036	3341	1609	817	3733
Filtered	2950	3240	1024	1026	2940	1550	760	3260
Augmented	**44,250**	**48,600**	**15,360**	**15,390**	**44,100**	**23,250**	**11,400**	**48,900**

From the table, it can be observed that many instances of the non-extended classes were discarded. This is because of the fact that many of these images were corrupted by intense illumination (i.e., sun light) or were under-exposed; our data collection process was more careful, and thus the filtering for the D00, D10, D11, and D40 was less severe.

3.3. Training

We have already discussed the advantages of using various recently developed generic object detection methods based on deep learning architectures; in what follows, we will discuss how we trained these different algorithms, and later, we will compare them in terms of performance. In order to compare our results with other works in the state-of-the-art, we have decided to implement various generic object detectors, focusing our efforts on methods amenable for mobile and embedded implementations: an LBP-based object detector (as in [17]), but also modern, deep learning algorithms, RetinaNet (as in [23]) and MobileNet (as in [25]).

For the former model, as discussed thoroughly in [21], the architecture supports the training with different feature extraction backbone networks (for instance, RestNet or VGG). The choice of backbone serves as a means for exploring trade-offs between accuracy versus inference time. In general, RetinaNet yields better results other one stage detectors (i.e., YOLO), while producing models that are efficient for embedded or mobile implementations. According to our experiments, this detector model, using VGG19 for feature extraction, has a memory footprint of 115.7 MB and attains an inference time of 500 ms, a low enough latency for most road damage acquisition systems.

We carried out the training of the deep learning models described above using the following methodology. The filtered dataset was randomly shuffled and split into two disjoint sets: 60% of the images were used for training and 30% for validation. We took great care to avoid overfitting in our models; in order to do so, we integrated regularization techniques such as dropout and sample augmentation over the proposed extended dataset, as discussed above.

The model description, training, and validation were all done using Keras and Tensorflow. The training and all of the other experiments were executed on the collaboratory tool provided by Google, making use of a Tesla K80 GPU. The tested architectures were trained using the following parameters: a batch size of 24 and number steps per epoch of 300. We optimized the search of the model parameters using Adam, modifying the learning rate according to the approach proposed in [24]. As the results to be discussed next showcase, our RetinaNet-based approach yielded significant improvements, particularly for the classes poorly represented in the dataset in [25].

4. Results and Discussion

In what follows, we will discuss the implementation and experimental results obtained using the augmented dataset on several recent state-of-the-art object detectors. First, in Section 4.1, we will compare the performance of three of these models using four classes from the augmented dataset (most recent works consider only single classes or few of them, with the exception of the work in [25]). Afterwards, we analyze the results of an ablation study performed to assess the impact of the quality aware data augmentation strategy described in Section 4.2 compared with the baseline. In Section 4.3, we compare our best performing model (based on the RetinaNet object detector) against the model in [25] under various metrics. Finally, in Section 4.4, we compare both models in terms of real-time performance and model size, important metrics for an effective implementation of deep learning models on constrained devices.

As mentioned above, we carried all our experiments on Google Colab, a platform that enabled us to train our models using cloud computing resources. Moreover, we are able to host our databases on Google Cloud, avoiding any computational burden. The testing phase of the chosen models was also carried out remotely, importing the trained models and the validation set. Additionally, we performed tests on real-time video captured from smartphones in order to assess the capabilities of the proposed approach in different scenarios, using the Cartucho/mAP evaluation tool [32].

4.1. Performance Comparison between Real-Time Object Detectors

Table 5 summarizes the performance results, in terms of accuracies obtained using the chosen "generic object detectors" in our tests. As discussed above, we included tests using Haar-based classifier as in [17] to see if any improvement could be obtained, as we included part of their dataset for the testing phase of our models.

In the table, we also included metrics for performance for the best performing object detectors in the state-of-the-art, which made use of RetinaNet [21] and MobileNet [22]. The reported models attained in general better results than previously using the same configurations ([17,21,25], respectively), owing to the use of the extended and quality aware dataset. As we will see next, this was accompanied with comparable or better precisions (and lower recalls), as well as reduced latencies.

Table 5. Comparison between different object detectors reported in the literature making use of our extended dataset, significant improvements were obtained for all, but RetinaNet was chosen owing to its performance.

Metric/Class	D01 (Long Crack)	D11 (Lat. Crack)	D20 (Alligator)	D40 (Pothole)
LPB-cascade	0.7080	0.7182	0.7625	0.7476
MobileNet	0.8124	0.9032	0.9234	0.9675
RetinaNet	**0.9148**	**0.9511**	**0.9522**	**0.9823**

In Table 5, we chose to include a comparison of only four of most studied classes (common in works in the state-of-the-art, as they are the most prevalent); we believe that, otherwise, the comparisons will be less fair and more difficult to appreciate. We have to stress that other methods analyzed in Section 3 yielded good results for individual types of asphalt damages, but in general, those models were not amenable for constrained mobile implementations (as shown in Table 1).

The sole reference in the state-of-the-art to implement a full-fledged acquisition system for all eight classes in Table 2 and to carry out a thorough analysis is the work presented by the researchers at the University of Tokyo [25]. The authors designed and tested two deep learning-based object detectors (using Inception V2 and MobileNet, respectively). They reported that the best results were obtained with the latter, but in recent years, other object detectors have been developed (i.e., RetinaNet), which yielded better results in our experiments, as can observed in Table 6. For the sake of completeness, we show some instances of the test dataset that were correctly identified and detected in Figure 9. From the figures, it can be verified that, apart from detecting the instances with high confidence, the extent of the structural damages is well localized.

Figure 9. Some examples of detection results for the D01, D20, and D43 road damage classes.

Table 6. Mean average precisions obtained through the ablation study; note that the results for both the UT and augmented dataset are better than those in [25], and the quality aware model outperforms them by about 10%.

Metric/Class	D01 (Long Crack)	D11 (Lat. Crack)	D20 (Alligator)	D40 (Pothole)
UT Dataset	0.8030	0.8225	0.8351	0.9233
Augmented	0.8420	0.8550	0.8759	0.9534
Quality Aware	**0.9148**	**0.9511**	**0.9522**	**0.9823**

4.2. Ablation Studies for Assessing the Effect of the Image Quality Aware Data Augmentation Strategy

We conducted an ablation study using the different datasets in Table 3. For this study, we trained only the RetinaNet object detector following the procedure described in Section 4.3, monitoring both learning curves and, specifically, the losses or the training and validation sets.

The results of this study are shown in Figure 10, which depicts the training loss curves for the models trained with the augmented and quality aware datasets (pink and green lines) and the validation losses for the augmented (blue line) and quality aware (red line) datasets.

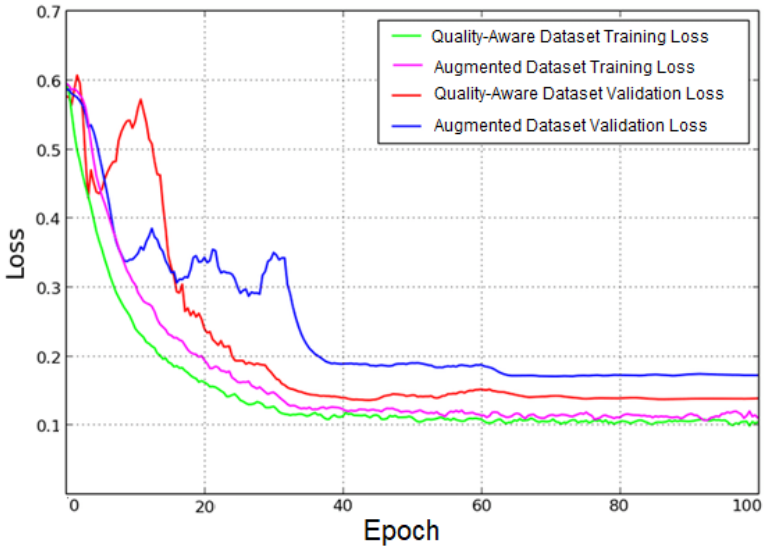


Figure 10. Training and validation loss curves for the augmented and quality aware datasets.

As can be observed, both strategies significantly reduce the validation loss, but a clear improvement can be seen with the latter, which is also reflected in the performance of the trained detectors, as summarized in Table 6. The table also shows the results for the model using the original dataset proposed in [25], for which very similar (but slightly superior) results were obtained (in the first row of the table). These results are consistent with other works that make use of RetinaNet [23], as this object detector was designed to focus on "hard examples", making it more robust, particularly in terms of recall, as we will see in Section 4.3 (see Table 7).

Table 7. Comparison with the best method previously published in the state-of-the-art [24] for all of the classes in Table 2; we show the results for recall precision and accuracy for both methods.

Metric	**Road Damage Instance**							
Maeda et al. [25]	**D00**	**D01**	**D10**	**D11**	**D20**	**D40**	**D43**	**D44**
Recall	0.40	0.89	0.20	0.05	0.68	0.02	0.71	0.85
Precision	0.73	0.64	0.99	0.95	0.68	0.99	0.85	0.66
Accuracy	0.81	0.77	0.92	0.94	0.83	0.95	0.95	0.81
Best Model (Ours)								
Recall	0.60	0.90	0.40	0.40	0.76	0.70	0.80	0.70
Precision	0.87	0.70	0.89	0.92	0.92	0.88	0.87	0.82
Accuracy	**0.91**	0.81	0.92	**0.95**	**0.95**	**0.98**	0.95	**0.84**

4.3. Comparison in All Classes with the State-of-the-Art

We performed a comparative assessment to evaluate whether any improvements were obtained by making use of the extended, quality aware generated datasets. The comparison was made between our best performing model, based on RetinaNet, and the best model in [25], which used MobileNet. The results of this analytical exercise are summarized in Table 7. The RetinaNet-based model performed consistently better than the MobileNet implementation, showing better accuracies and higher recall for all of the classes (D11 and D40 in particular, as expected).

This fact is important as they are the classes that we are more interested in identifying for the purpose of road damage monitoring applications, where these classes must be detected with high reliability. These results can be correlated with the better representation of these training instances in our dataset, which was an issue in previous research [25]. The researchers have cited the class D43 (lane marking degradations) as an exception, given that the samples for this class are limited, but we believe that such simpler samples are by definition easier to classify, which might explain the results.

For our implementation, we made use of VGG19 for feature extraction as in [21,22], but we report the full metrics for all eight classes in Table 2, whereas the authors in [21] use a combined single metric (a mAP of 0.8279), instead of the usual individual per-class comparison. From all of the previous results (summarized in Tables 5–7), we can confidently assert that the quality aware augmentation strategy implemented in this study was indeed effective.

4.4. Real-Time Performance Evaluation and Model Size

In Table 8, we summarize some of the most important characteristics of our RetinaNet-based implementation, contrasted with other two models in the recent literature. As can be observed from the table, the memory footprint is relatively low (125.5 MB), which enables the model to attain a low inference time (500 ms), while achieving a very competitive precision (mAP). These factors make our model one of the most amenable for implementation in constrained devices in the literature. As we mentioned in Section 2, the inference time of the detector depends largely on the choice of the backbone or feature extraction layers (various configurations of ResNet or VGG), which can range from 200 to 500 milliseconds in our experiments. The configuration used here makes use of VGG19, which represents a good compromise in terms of memory footprint and accuracy, at the cost of a slightly larger inference time. However, the acquisition process is done at 40 km/h (11 m/s) and, given

the large line of sight, 500 milliseconds is more than sufficient to acquire the samples. It must be noted that the obtained inferences using GPU are much lower, ranging from 20 to 60 milliseconds, but their use is simply not possible in this context (but it could work in an ADAS context).

Table 8. Metrics comparison for our best model against the state of the art.

Model/Metric	Memory Footprint	Latency	Best Accuracy
RetinaNet [22]	115.7 MB	500 ms	0.8279
MobileNet [24]	Not reported	1500 ms	Not reported
RetinaNet (ours)	125.5 MB	500 ms	0.91522

5. Conclusions

In this paper, we introduced the foundational block for implementing an acquisition system for a physical asset management tool. The detection system is geared road damage identification and makes use of a very recent generic object detector (RetinaNet) trained on a quality aware filtered dataset. The proposed method can detect various asphalt structural damages from video with high accuracy and low inference times, effectively serving as a stepping stone for implementing sophisticated road damage acquisition systems.

We conducted a thorough comparison between various object detectors and several ablation studies, for both detectors based on traditional computer vision algorithms and others based on two- and one-stage objects detectors. For validating our models, we made use of a very large dataset of structural damages initially proposed in [25], which we augmented with road damage samples from Mexican and Italian roads [17]. The main rational for this approach was to mitigate the class imbalance present for some road damage types such as potholes and alligator cracks, which are not well represented in the previous dataset.

Furthermore, we proposed and made use of an image quality aware data augmentation strategy for discarding road damage samples that could hamper the performance of the trained detectors, and we carried out an ablation study to demonstrate the effectiveness of such an approach for training deep learning-based object detectors. The ablation study clearly demonstrated that data augmentation is beneficial for training the generic object detectors, but its effectiveness can be hampered owing to the poor quality of the images, as documented in the literature. As demonstrated by our experiments, a simple, yet powerful "image filtering" technique can ameliorate the proceedings and lead to a more optimal model.

Among the various tested models, the results demonstrated that RetinaNet outperforms other recent architectures for object detection, while producing a model that it is efficient for smartphone implementation (in terms of memory footprint) and attaining acceptable inference rates (500 ms in our RetinaNet model). An additional advantage of RetinaNet is that it presents less jitter in the detection, owing to improved non-maximum suppression strategies and leading to a better performance.

The obtained inference times are lower than those in previous research (500 ms in contrast to 1500 ms in [24]); because the detector is to be implemented on mobile phones attached to a car, low inference times are of paramount importance, and although 500 ms might seem not sufficiently low, there is also a compromise to be made in terms of accuracy and memory footprinted, as discussed in Section 4.4. Nonetheless, previous research [25] demonstrated that less than 1500 ms is sufficient for imaging road damages for a vehicle moving at 40 km/s (10 m/s), while avoiding information duplication, as the line of sight is sufficiently large for most purposes.

As a continuation of this research, we plan to introduce the generated detector and mobile application into a tool for prognosis and big data analysis applied to road asset management as a cloud service. The main goal is taking full advantage of the capabilities of a heterogeneous architecture for follow up operations in road asset management by municipal or regional organizations. If coupled with a data analytics and prognosis model, this geo-referenced road damage database could lead to a better decision-making process in road management, in both the allocation of resources for maintenance as well as the required periodicity.

Author Contributions: Conceptualization, A.A.-M. and G.O.-R.; methodology, A.O.-Z. and G.O.-R.; software, A.A.-M. and J.A.V.-F.; validation, S.N.; resources, L.A.-L.; data curation, A.A.-M. and J.A.V.-F.; writing—original draft preparation, G.O.-R.; writing—review and editing, L.A.-L. and A.O.-Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: We appreciate the support of Dr. Benedetto from Universita de Roma 3 (Italy) for his support with some road images to augment the UT dataset (Japan). We would like to extend our appreciation to our colleagues in Vidrona LTD (especially to Dr. Ashutosh Natraj for involving us is the project) and to other team members for helping us to delimit the problem, especially to Shailendra Natraj.

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

References

1. Koch, C.; Asce, A.M.; Jog, G.M.; Brilakis, I. Automated Pothole Distress Assessment Using Asphalt Pavement Video Data. *J. Comput. Civ. Eng.* **2013**, *27*, 370–378.
2. Oliveira, H.; Correia, P.L. Automatic Road Crack Detection and Characterization. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 155–168.
3. Radopoulou, S.C.; Bralakis, I. Patch Detection for pavement assessment. *Autom. Constr.* **2015**, *53*, 95–104.
4. Geiger, A.; Lenz, P.; Stiller, C.; Urtasun, R. Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* **2013**, *32*, 1231–1237.
5. Medina, R.; Llamas, J.; Zalama, E.; Gómez-García-Bermejo, J. Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 27–30 October 2014; pp. 778–782.
6. Seung-Ki, R.; Taehyeong, K.; Young-Ro, K. Image-Based Pothole Detection System for ITS Service and Road Management System. *Math. Probl. Eng.* **2015**, *2015*, 968361.
7. Angulo-Murillo, A.; Ochoa-Ruiz, G. Dataset: "Road Surface Damages". *IEEE DataPort* **2019**, doi:10.21227/nbdy-r451.
8. Mathavan, S.; Kamal, K.; Rahman, M. A Review of Three-Dimensional Imaging Technologies for Pavement Distress Detection and Measurements. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2353–2362.
9. Schnebele, E.; Tanyu, B.F.; Cervone, F.; Waters, G. Review of remote sensing methodologies for pavement management and assessment. *Eur. Transp. Res. Rev.* **2015**, *7*, 7.
10. Koch, C.; Giorgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A review of computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210.
11. Mohan, A.; Poobal, S. Crack detection using image processing: A critical review and analysis. *Alex. Eng. J.* **2018**, *57*, 787–798
12. Hoang, N.D.; Nguyen, Q.L. A novel method for asphalt pavement crack classification based on image processing and machine learning. *Eng. Comput.* **2018**, *35*, 487–498.
13. Hoang, N.D. An Artificial Intelligence Method for Asphalt Pavement Pothole Detection Using Least Squares Support Vector Machine and Neural Network with Steerable Filter-Based Feature Extraction. *Adv. Civ. Eng.* **2018**, *2018*, 7419058.
14. Pouyanfar, S.; Sadiq, S.; Yan, Y.; Tian, H.; Tao, Y.; Presa-Reyes, M.; Shyu, M.L.; Chen, S.C.; Iyengar, S.S. A Survey on Deep Learning: Algorithms, Techniques, and Applications. *ACM Comput. Surv.* **2018**, *51*, 92.
15. Zhang, L.; Yang, F.; Zhang, Y.D.; 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.
16. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Network. *Comput. Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378.
17. Tedeschi, A.; Benedetto, F. A real time pavement crack and pothole recognition system for mobile Android-based devices. *Adv. Eng. Inform.* **2017**, *32*, 11–25.
18. Siriborvornratanakul, T. An Automatic Road Distress Visual Inspection System Using an Onboard In-Car Camera. *Adv. Multimed.* **2018**, *2018*, 2561953.
19. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Liu, X.; Pietikäinen, M. Deep Learning for Generic Object Detection: A Survey. *arXiv* **2016**, arXiv:1809.02165v2.

20. Huang, J.; Rathod, V.; Sun, C.; Zhu, M.; Korattikara, A.; Fathi, A.; Fischer, I.; Wojna, Z.; Song, Y.; Guadarrama, S.; et al. Speed/accuracy trade-offs for modern convolutional object detectors. *arXiv* **2018**, arXiv:1611.10012v3.
21. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
22. Lin, T.; Goyal, P.; Girshick, R.B.; He, K.; Dollár, P. Focal loss for dense object detection. *arXiv* **2017**, arXiv:1708.02002.
23. Ale, L.; Zhang, N.; Li, L. Road Damage Detection Using RetinaNet. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; pp. 5197–5200.
24. Pereira, V.; Tamura, S.; Hayamizu, S.; Fukain, H. A Deep Learning-Based Approach for Road Pothole Detection in Timor Leste. In Proceedings of the 2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Singapore, 31 July–2 August 2018; pp. 279–284.
25. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiyama, T.; Omata, H. Road Damage Detection Using Deep Neural Networks with Images Captured Through a Smartphone. arXiv 2018, arXiv:1801.09454.
26. RoadBotics. Available online: https://www.roadbotics.com/company/ (accessed on April 1st 2020).
27. Dodge, S.; Karam, L. Understanding how image quality affects deep neural networks. In Proceedings of the 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX), Lisbon, Portugal, 6–8 June 2016.
28. Angulo, A.; Vega-Fernández, J.A.; Aguilar-Lobo, L.M.; Natraj, S.; Ochoa-Ruiz, G. Road Damage Detection Acquisition System Based on Deep Neural Networks for Physical Asset Management. In Mexican International Conference on Artificial Intelligence; Martínez-Villaseñor, L., Batyrshin, I., Marín-Hernández, A., Eds.; Springer: Cham, Switzerland, 2019; Volume 11835.
29. Diamond, S.; Sitzmann, V.; Boyd, S.; Wetzstein, G.; Heide, F. Dirty Pixels: Optimizing Image Classification Architectures for Raw Sensor Data. arXiv 2017, arXiv:1701.06487.
30. Lagendijk, R.L.; Biemond, J. Basic Methods for Image Restoration and Identification. Available online: www-inst.cs.berkeley.edu/~ee225b/sp10/handouts/Image_Restoration_99.pdf (accessed on Day April 1st 2020).
31. Ortiz-Preciado, A.A.; Vega-Fernandez, J.A.; Ochoa-Ruiz, G. Thoracic Disease Classification using Deep Learning and Quality Aware Data Augmentation in the ChestX-ray8 Dataset. Res. Comput. Sci. 2019, 148, 41–53.
32. Cartucho, J. Mean Average Precision. 2020. Available online: https://github.com/Cartucho/mAP (accessed on 1 April 2020).