Llopis-Castelló, D.; Paredes Palacios, R.; Parreño-Lara, M.; García-Segura, T.; Pellicer, E. (2021). Automatic Classification and Quantification of Basic Distresses on Urban Flexible Pavement through Convolutional Neural Networks. Journal of Transportation Engineering, Part B: Pavements. 147(4):1-8. https://doi.org/10.1061/JPEODX.0000321

AUTOMATIC CLASSIFICATION AND QUANTIFICATION OF BASIC DISTRESSES ON URBAN FLEXIBLE PAVEMENT THROUGH CONVOLUTIONAL NEURAL NETWORKS

David Llopis-Castelló, **Ph.D.**, Assistant Professor, dallocas@upv.es; Highway Engineering Research Group, Universitat Politècnica de València (UPV), Camino de Vera, s/n. 46022 – Valencia, Spain. ORCiD: 0000-0002-9228-5407 (Corresponding author)

Roberto Paredes, Ph.D., Associate Professor, rparedes@dsic.upv.es; Pattern Recognition and Human Language Technology Research Center, UPV, Spain. ORCiD: 0000-0002-5192-0021

Mario Parreño-Lara, Ph.D. Student, maparla@phrlt.upv.es; Pattern Recognition and Human Language Technology Research Center, UPV, Spain. ORCiD: 0000-0002-4539-3288

Tatiana García-Segura, Ph.D., Assistant Professor, tagarse@upv.es; School of Civil Engineering, UPV, Spain. ORCiD: 0000-0002-7059-0566

Eugenio Pellicer, Ph.D., M.ASCE, Professor, pellicer@upv.es; School of Civil Engineering, UPV, Spain. ORCiD: 0000-0001-9100-0644

ABSTRACT

Pavement condition assessment is a critical step in road pavement management. In contrast to the automatic and objective methods used for rural roads, the most commonly used method in urban areas is the development of visual surveys usually filled out by technicians that leads to a subjective pavement assessment. While most previous studies on automatic identification of distresses focused on crack detection, this research aims not only to cover the identification and classification of multiple urban flexible pavement distresses (longitudinal and transverse cracking, alligator cracking, raveling, potholes, and patching), but also to quantify them through the application of Convolutional Neural Networks. Additionally, this study also proposes a methodology for an automatic pavement assessment considering the different stages developed in this research. This methodology allows for a more efficient and reliable pavement assessment, minimizing the cost and time required by the current visual surveys.

INTRODUCTION

Road maintenance is crucial to minimize transportation costs and vehicle emissions (Bull 2003; Setyawan and Kusdiantoro 2015) as well as to avoid cost overruns associated with late maintenance that leads to pavement reconstruction (Hajj et al. 2010). Therefore, it is essential to promote proper pavement management that provides suitable pavement conditions for road users at the lowest life cycle cost (AASHTO 2011). In fact, this is why various highway agencies have developed Pavement Management Systems (PMS) which could superintend the life cycle of existing road infrastructures (Hicks et al. 2011). These systems perform a variety of functions in which pavement condition assessment plays a critical role.

In general, these systems employ two types of road condition surveys: (i) automatic and (ii) visual. While automatic surveys employ objective indexes such as the International Roughness Index (IRI) and standard procedures, in visual surveys the assessment is always carried out by individual technicians whose personal interpretations are subjective and solely justified based on previous experience in the field.

Nonetheless, still the most commonly used method in urban areas has seen to be the visual survey, primarily due to the difficulty of obtaining objective indexes in such environments characterized by numerous intersections, high speed variability, and important traffic flow changes (Reggin et al. 2008; Wang et al. 2013). In addition, the most common urban pavement distresses and their influence on road safety and operation differ from those on rural roads.

To reduce the subjectivity associated with the visual method, a survey should be based on simple and straightforward criteria (Kraemer et al. 2004). Although there is yet no standard guideline to classify and quantify pavement distresses based on their type, severity, and extent, different researchers and agencies have proposed guidelines for use in a certain region or country (Miller and Bellinger 2003). The Long-Term Pavement Performance (LTPP) program of the Federal Highway Administration (FHWA) aims to collect data about pavement condition and maintenance and rehabilitation activities in the United States and Canada (Perera and Kohn 2001). The highway administrations of France and Switzerland have also developed systematic approaches to identify pavement distresses, and the Irish highway administration has incorporated an approved procedure of pavement condition assessment into their national road design guideline (Ragnoli et al. 2018).

In the last decade, image processing has been practicing by researchers to make pavement condition assessment more and more objective. The relative studies can be categorized based on their approach to image processing: (i) histogram analysis, (ii) mathematical morphological tools, (iii) Machine Learning techniques, (iv) filtering, and (v) analysis of a model (Chambon and Moliard 2011). Nevertheless, these studies focused primarily on crack detection and, in addition, did not provide a second algorithm or method for an automatic quantification of the predicted distresses.

In this context, the development of an automated procedure that can simultaneously identify and quantify the most common distresses affecting urban flexible pavement is of great interest in pavement management. Such a procedure would allow pavement engineers to perform more efficient and reliable pavement assessments and thus reduce the cost and time required by the current visual evaluation methods.

LITERATURE REVIEW

The first researchers addressing the automatic identification of cracks quickly discarded the use of Machine Learning techniques because of the computational and economic costs; instead, they proposed combining techniques from histogram analysis, mathematical morphological tools, and filtering procedures (Chambon and Moliard 2011). The Minimal

Path Selection (MPS) algorithm (Zou et al. 2012) and the CrackTree method (Amhaz et al. 2016) were the most promising techniques to automatically detect cracks in asphalt pavements.

Thanks to innovations and technological progress in the field of computational image processing, several studies have recently considered Machine and Deep Learning techniques to identify not only cracks but also other types of pavement distresses (Oliveira and Correia 2012; Radopoulou and Brilakis 2017; Shi et al. 2016; Hadjidemetriou et al. 2018). The most common Machine Learning techniques applied to the identification and classification of pavement distresses are: (i) Support Vector Machines, (ii) Decision Trees, (iii) Random Forest, and (iv) Neural Networks (NN).

CrackIT, the integrated system for automatic crack detection developed by Oliveira and Correia (2012), was one of the first methods based on Machine Learning techniques. This model relies on unsupervised learning and consists of two stages: (i) crack detection and (ii) crack characterization. Although its performance showed good results generally, it had difficulty detecting thin cracks (< 2 mm) and returned many false positives.

Radopoulou and Brilakis (2017) developed a low-cost method based on Decision Trees to identify longitudinal and transverse cracks, patches, and potholes from images gathered by those video cameras that assist in parking individual cars. Their method uses the Semantic Texton Forests (STF) algorithm as a supervised classifier and achieves an overall accuracy greater than 82%.

Shi et al. (2016) developed a new method called CrackForest to identify and classify pavement cracks based on Random Structured Forests that led to an even more accurate classification comparing to CrackTree, CrackIT, and MPS. Moreover, Hadjidemetriou et al.

(2018) proposed a method for patch identification and quantification based on Support Vector Machine (SVM) classification techniques.

Deep Learning for automatic crack detection by applying Convolutional Neural Networks (CNNs) has become very popular in the last few years (Zhang et al. 2016; Wang and Hu 2017; Xia 2018; Jenkins et al. 2018; Carr et al. 2018; Maeda et al. 2018; Park et al. 2019). A Convolutional Neural Network is a deep neural network with two or more hidden layers that operates in two stages: (i) base model or feature extraction and (ii) top model or classification (Figure 1). The main objective of the convolutional layer is to reduce the computational workload of the system by reducing the number of elements and detecting certain characteristics that can be useful when analyzing the image. This process is carried out by applying a filter to the input image and storing the result in the activation matrix. The pooling layer aims to decrease further the computational workload while characterizing the image by obtaining and locating its predominant features. Finally, the full connected layers deal with classification, i.e., indicate the probability that the input image displays a specific pavement distress.

Most of the studies that applied deep learning to the identification of pavement distresses focused on crack detection and used images collected from standard smartphones and cameras (Coenen and Golroo 2017). These studies identified differences in image resolution as well as the size of the blocks into which each image was divided (Table 1). Although the architecture of the Convolutional Neural Networks also differed among studies, the application of this technique resulted in more accurate findings than those achieved by methods based on other machine learning techniques, such as Support Vector Machine or Random Forest (Zhang et al. 2016; Carr et al. 2018; Park et al. 2019).

The deep neural network proposed by Maeda et al. (2018) was trained to identify not only cracks but also rutting, bumps, and potholes; however, the distresses identified in each image were labeled as cracks or other corruptions, which clustered the rest of the pavement distresses. In contrast, Xia (2018) did identify different road distresses (longitudinal, transverse, and alligator cracks; seam breaks; and potholes) through the application of CNNs. This research found the vehicle's speed during data collection to be a critical factor. As this speed increased from 10 to 60 km/h, the accuracy were observed to be decreased by 10 to 15%, depending on the type of road distress.

Mohanraj et al. (2018) proposed a methodology to enhance the accuracy of crack detection in "noisy" conditions, which were introduced through the image gathering system (e.g., shadows, painted signs, or tire marks). They divided their method into three steps: (i) image preprocessing by applying filters, (ii) feature extraction, and (iii) detection and classification using k-mean clustering.

GOAL OF THE RESEARCH

Unlike most previous studies on automatic identification of distresses that focused on crack detection, this research aims not only to identify and classify multiple urban flexible pavement distresses (longitudinal and transverse cracking, alligator cracking, raveling, potholes, and patching), but also to quantify them through the application of Convolutional Neural Networks. Specifically, this study proposes a two-step method consisting of two concatenated CNNs, the first one for distresses identification and the next one to quantify the predicted distresses. It also proposes an innovative data collection methodology using video cameras located at the rear of passenger cars to assess pavement condition automatically.

RESEARCH METHOD

The research method used for this study consisted of five steps: (i) data collection, (ii) video filtering and image preprocessing, (iii) data augmentation, (iv) Convolutional Neural Network training, and (v) Convolutional Neural Network training validation. A video camera at the rear of a passenger car conducted data collection in an urban environment. Researchers filtered the collected videos to extract the images of the pavement and processed them to remove areas of each image that did not contain relevant information for the detection and classification of pavement distresses and to divide each image into three blocks of 256 x 256 pixels. These processes created a database of images classified by types of pavement distresses. Researchers considered different data augmentation techniques because of the large amount of data required for training and decided on a two-step procedure consisting of two concatenated Convolutional Neural Networks (CNN).

Data collection

To obtain a comprehensive database of pavement distresses, researchers collected data along a total of 140 km of streets in the city of Valencia (Spain) under favorable weather conditions using a Garmin Virb Ultra 30 video camera. This video camera was mounted on the rear of a passenger car by a camera gripper suction system that ensured a zenithal position – at 1.4 m high – to avoid image distortion (Figure 2).

Traveling at 50 km/h maximum speed (as required in the urban areas of the city) and using “1080p” recording mode and “wide” view angle, researchers recorded approximately 350 minutes of video at 24 pixels per inch resolution (1920 x 1080 pixels).

Video filtering and image preprocessing

The video filtering stage extracted certain frames of the recorded video. Given that the width of each filtered image was approximately one meter (w) and the maximum traveling speed was 50 km/h, the required video frame rate was approximately 15 frames per second (fps)

($f = v/w$). However, the frame rate during the data collection was 30 fps, so researchers extracted one of every two frames for a total of 361,381 images.

Researchers cropped each image to remove the part of the vehicle captured in the frame and split each image into three blocks, resized to 256 x 256 pixels each (Figure 3). This last step minimized the likelihood of having different types of distress in each block and consequently enhanced the accuracy of the classification procedure. Automated scripts programmed in Python carried out both the video filtering and image preprocessing stages.

To ensure as objective a classification as possible, three experts identified the diverse road distresses included in each image considering the following categories: (i) longitudinal cracking, (ii) transverse cracking, (iii) alligator cracking, (iv) raveling, (v) potholes, (vi) patching, (vii) road markings, (viii) manholes, and (ix) no pavement distress. The resulting supervised database contained 29,846 images and the number of images in which each type of distress was present was: 5,697 for longitudinal cracking, 3,467 for transverse cracking, 595 for alligator cracking, 622 for raveling, 1,231 for potholes, 3,420 for patching, 9,047 for road marking, 1,149 of manholes, and 18,002 with no evidence of pavement distress – including images with only road markings and/or manholes –. In this way, the number of images with only one category of distress was 9,186. The amount of images containing two, three, four, and five types of distress was 2,202, 388, 62, and 6, respectively.

Data Augmentation

A common problem for applying deep learning techniques is the large amount of data required for training to arrive at a particular level of generalization. To address this problem, this study used various data augmentation techniques. These techniques consist of making slight transformations to the training images to create new images that were similar enough to the original ones to maintain the original class of pavement distress but that also introduced some variations that might not be significantly represented in the original dataset. These transformations are performed by iterating over the training dataset. These modified images were used for network learning and then discarded.

The employed data augmentation techniques included: (i) Random Crop, which consists of selecting a random part of the original image; (ii) Horizontal Flip, which flips the image horizontally; and (iii) Color Jitter, which introduces slight modifications in brightness, contrast, and saturation to simulate lighting variations. The random crop technique produced images of 224 x 224 pixels, the size required by the ImageNet configuration, which is an ongoing research effort to provide researchers around the world with an easily accessible image database (Russakovsky et al. 2015). The other data augmentation techniques did not modify the size of the images.

Convolutional Neural Network

The study employed two complemented methods for image characterization and damage quantification:

- A Convolutional Neural Networks (CNN1) that identified all types of distress included in an image (longitudinal cracking, transverse cracking, alligator cracking, raveling, potholes, and patching) (Figure 4).
- Four CNNs (CNN2) that quantified the severity of each type of distress classified by CNN1. These CNNs determine the geometric features – area, length, and width – of longitudinal cracks, transverse cracks, potholes, and patches. In the case of alligator cracking and raveling, the whole image was considered as damaged because these types of distress usually took up a large area of the image (>80%).

Both methods are based on a ResNet architecture introduced by He et al. (2016) and applied to a variety of image classification problems with excellent results (Carr et al. 2018). During the training stage, researchers considered the Adam optimization algorithm, which calculates an exponential moving average of the gradient and the squared gradient, while having the parameters beta1 and beta2 to control the decay rates of these moving averages, providing an optimization algorithm that can handle sparse gradients on noisy problems (Kingma and Ba 2015). Cross-entropy loss function, which increases as the predicted probability diverges from the ground-truth label, is used. The pre-trained weights of the ResNet were used as initial values for training on ImageNet, which is an ongoing research effort to provide researchers around the world with an easily accessible image database (Russakovsky et al. 2015). Images were then resized to 256 x 256 pixels to meet the requirements of ImageNet. A learning rate of 0.0001 was selected and a 5-fold cross-validation procedure was used to estimate the accuracy of the model on new data.

Particularly, this study used the ResNet34 architecture. The first step on the ResNet before entering the common layer behavior is a block (Conv1) consisting of a convolution, batch normalization, and max pooling operation based on a kernel size of 7 and a feature map size of 64. Taking these parameters into account, the output size of that operation will be a (112×112) volume. Since each convolution filter (of the 64) is providing one channel in the output volume, the result is (112x112x64) output volume. The next step is the batch normalization, which is an element-wise operation and therefore does not change the size of the volume. In this way, a (3x3) max pooling operation with a stride of 2 is applied.

The ResNet consists of four blocks following the same pattern. Each block performs a 3x3 convolution with a fixed feature map dimension (64, 128, 256, 512) respectively, bypassing the input every 2 convolutions. Furthermore, the width and height dimensions remain constant during the entire layer. The process yields 32 hidden layers, which along with the initial 2 hidden layers total 34 hidden layers. Finally, an average pooling layer and a dense layer are used to extract the final features and classify the image. The total number of floating-point operations per second (FLOPS) is 3.6 x 10^9.

CNN1 addressed a multi-label classification problem that aimed to identify all distress instances that appear in a single image. To do so, all images were labeled using one-hot encoding using ones to denote the presence of determined distress. The Network was trained using binary cross entropy loss. After training the network, researchers selected a desirable threshold rate for false positives by using the Receiver Operating Characteristic (ROC) curve. A ROC curve displays the performance of a classification model at all classification thresholds. This curve plots two parameters, the True Positive Rate (TPR) and the False Positive Rate (FPR), from which researchers selected a threshold aimed at minimizing FPR and ensuring a high TPR. After the multi-label classification provided by CNN1, the second CNN (CNN2) assessed only those images with distresses as the objective of this stage was to measure the severity of each damage in terms of geometric dimensions – length, width, and area –.

This study implemented all experiments using the Pytorch framework and two NVIDIA RTX 2080 GPUs. The Python code, models, and full results are available at https://github.com/MarioProjects/MnMsCardiac (Parreño-Lara et al. 2021).

RESULTS

Training

Deep Learning requires a great amount of data to train a model so that removing a part of the database for validation poses a problem of underfitting. By reducing the training data, a risk of losing important patterns/trends in data set exists, which in turn increases error induced by bias. In this context arises K-Fold cross-validation which is a method that provides a wide dataset for training the model and also leaves a wide dataset for validation. In K-Fold cross validation, the data is divided into k subsets. The holdout method is repeated k times, such that each time, one of the k subsets is used as the test set/validation set and the other k-1 subsets are put together to form a training set. The error estimation is averaged over all k trials to get total effectiveness of the model.

This study employed a 5-fold cross-validation procedure to obtain more reliable results. To conduct this procedure, researchers split the available sample into five parts, using four parts for training and the fifth part for validation. Excluding the validation fold from the training task, the researchers calculated the global metrics by averaging the results of varying the validation fold among the five possible combinations.

The accuracy of the trained Convolutional Neural Networks was assessed through four parameters: (i) *Precision*; (ii) *Recall*; (iii) *F1 score*; and (iv) *Intersection over Union* (IoU). Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (Eq. 1), whereas Recall is the ratio of correctly predicted positive observations to all observations (Eq. 2). F1 score is the weighted average of Precision and Recall, considering, therefore, both false positives and false negatives (Eq. 3). Finally, Intersection over Union (IoU) is an evaluation metric used to measure the accuracy of an object detector on a particular dataset. This parameter compares the ground-truth pixels (i.e., the hand labeled pixels from the testing set that specify where in the image the distress is) and the predicted pixels (Eq. 4).

$$Precision = TP/(TP + FP) \tag{1}$$

$$Recall = TP/(TP + FN) \tag{2}$$

$$F1\ score = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision} \tag{3}$$

$$IoU = \frac{Common\ pixels\ between\ Ground - truth\ and\ Prediction}{Ground - truth\ pixels + Predicted\ pixels} \tag{4}$$

where True Positives (TP) are the correctly predicted images with distresses; True Negatives (TN) are the correctly predicted images without distresses; False Positives (FP) are images predicted as damaged but actually there is no distress; and False Negatives (FN) are images with distresses predicted as without damage.

The first multi-label classification network (CNN1) yielded results of 0.9317 precision, 0.9252 recall, and 0.9262 F1 score. Figure 5 shows the ROC curve for CNN1. This result led to the choice of a threshold of 0.8% for false positives, which meant that four of every 1,000 images that did not contain distresses were classified with at least one "damage", leaving the precision for true positives still very high at 92.35%. CNN2, which focused on distress quantification via image segmentation, yielded a global intersection over union (IoU) of 0.6821 for longitudinal cracks, 0.6709 for transverse cracks, 0.8760 for patches and 0.6870 for potholes.

Table 2 shows the results for each class for CNN1, where the precision for each of the different types of distresses is higher than 0.9. These high values are a result of the quality of the data: the collected images for training have the same lighting conditions, the video camera was set at a constant position, and damages were sufficiently distinct. Pavement inspections can replicate this uniformity easily by conducting data collection under the proper weather and lighting conditions. Doing so produces a high-quality dataset with minimal variability among the classes, resulting in highly accurate and reliable classifications.

Among the incorrect classifications, 1.6% of transverse cracking images, mainly presenting wide crack widths, were mislabeled as potholes, while some images initially classified as raveling and potholes were actually longitudinal cracks. These latter improper classifications usually occurred when the distress was at the edges of the image. Additionally, some potholes that occupied a large area of the image were wrongly classified as raveling and patching. Researchers also measured fold training times and inference times. The average Resnet34 inference time was 0.005 seconds, and the training time per fold was 2.5 hours, while this training time for distress quantification was around 1.25 hours per fold.

Validation

The performance of the proposed method was assessed considering an unseen dataset, i.e., images not used in training. This dataset, consisting of a total of 12,788 images, was obtained from a 4.262 km road section of the data collection that was not considered for training. Particularly, these images were used to validate only CNN1 that aims to identify and classify the diverse types of distress existing in an image or block. To determine the accuracy of the proposed method, the parameters of Precision, Recall, and F1 score were estimated resulting in 0.9733, 0.9146, and 0.9431, respectively. Comparing these values with those obtained in training, it can be concluded that the method proposed in this study is able to accurately identify and classify urban pavement distresses.

DISCUSSION

Currently, urban pavement assessment is performed through visual inspections that leads to a subjective evaluation of pavement condition. To minimize this subjectivity and achieve a more efficient pavement assessment, various studies have recently considered using image processing techniques to identify distresses.

The most accurate, reliable, and efficient methods are those based on Deep Learning, particularly through the application of Convolutional Neural Networks (Zhang et al. 2016; Wang and Hu 2017; Jenkins et al. 2018; Carr et al. 2018; Park et al. 2019), but most of these studies focused only on crack detection. By contrast, this research presents a two-step procedure consisting of two concatenated Convolutional Neural Networks to automatically identify and quantify not only longitudinal and transverse cracks, but also alligator cracking, raveling, potholes, and patching. As a result, the proposed procedure allows pavement engineers to identify and classify pavement urban distresses with an precision of more than 0.93 on average.

This study also used a total of 29,846 pavement images, many more than the number of images considered in previous studies (Table 1). This research required a greater number of pavement images because it included more types of distresses, and this high-quality and extensive dataset led to a highly accurate classification of the different urban pavement distresses.

The findings of this research yielded a new methodology for assessing urban pavement condition automatically (Figure 6). This procedure consists mainly of the four steps of the research method defined in this study and an additional stage focused on pavement condition estimation. The main strength of this methodology is that it allows pavement engineers to perform a more efficient and reliable pavement assessment, minimizing the cost and time required for the current visual surveys.

A recent pilot study in the city of Valencia surveyed a total of 50 km of urban roads. The data collection took approximately 5 hours, and the data reduction and distress classification and quantification took 18 hours (corresponding with stages 1-4). This automatic inspection took a full-time expert approximately three days, while a manual visual inspection of these same urban roads might have taken the same full-time expert up to four weeks.

The first stage of the proposed methodology is the automatic pavement data collection through video cameras installed on a specific vehicle or even on public vehicles, such as police cars or public transport buses. It is required to record the pavement from a zenithal position – at 1.4 m high – to avoid image distortion and, additionally, to set the camera to 30 fps to ensure the recording of the entire road length at the maximum urban speed – 50 km/h –. The second stage involves processing the video to extract the pavement images and dividing each frame into blocks of 256 x 256 pixels. A script programmed in Python performs this stage automatically, removing duplicated images when the vehicle stops – e.g., at traffic lights –. The third stage consists of identifying and classifying pavement distresses by analyzing each block through the first CNN (CNN1) proposed in this study. This technic is able to predict all types of distress included in each image, i.e., it consists in a multi-labeling task. After their identification, the different distresses must be quantified – length and width of cracks and area of raveling, potholes, and patching –. This task is developed by the second CNN trained in this study (CNN2). Finally, the last step of the proposed methodology aims at estimating the condition of the pavement. Among the diverse indexes proposed in the literature, the Pavement Condition Index (PCI) (ASTM 2003) is the most commonly used. Although this index depends on a total of 19 types of distresses, it might be used to estimate urban pavement condition because the method proposed in this study is able to identify the most common urban pavement distresses. Nevertheless, other indexes such as the Urban Pavement Condition Index (UPCI) (Osorio et al., 2014) might be used. Regardless the used index, the most important contribution of this research is that the proposed method provides the data needed to estimate these indexes, i.e., pavement distress identification and quantification.

CONCLUSIONS AND FURTHER RESEARCH

A critical step in managing road pavement is the assessment of pavement condition. Although diverse objective indexes (e.g., IRI) are collected automatically to assess pavement condition on rural roads, the most commonly used method in urban environments is the visual survey conducted by a technician, which introduces a certain degree of subjectivity into the assessment.

To minimize this subjectivity, this study proposes a new two-step procedure to identify and quantify road distresses through the application of Convolutional Neural Networks. While most previous research focused only on crack detection, this method can classify the most common urban road distresses (longitudinal, transverse, and alligator cracks, raveling, potholes, and patches) and quantify their severity – geometric features –.

This procedure consists of two concatenated CNNs. The first one identifies all urban pavement distresses that an image contains with 0.9317 precision, 0.9252 recall, and 0.9262 F1 score. The second CNN quantify the severity of each predicted distress in those images containing longitudinal cracks, transverse cracks, potholes, and/or patches. As a result, the IoU of the distresses quantification is more than 0.65 for all cases. Therefore, the proposed two-step procedure consisting of two concatenated convolutional neural networks yields a highly accurate and reliable classification of urban pavement distresses.

Finally, a new methodology to automatically assess pavement condition is proposed, consisting of five stages: (i) pavement data collection by video cameras installed at the rear of a vehicle to ensure a zenithal position, (ii) automatic video filtering and image preprocessing to produce 256-x-256-pixel images, (iii) classification of pavement distresses, (iv) quantification of pavement distresses, and (v) an estimation of pavement condition.

A pilot study based on stages i-iv of this methodology showed that this approach allows pavement engineers to perform a more efficient and reliable pavement assessment, minimizing the cost and time required by the current visual surveys.

Although the findings of this study are encouraging, further research is needed to cover some limitations of this study. The trained Convolutional Neural Networks can only identify those distresses associated with changes in the image properties, i.e., texture, contrast, brightness, and so on. However, those distress types that do not present a deterioration on the pavement surface, such as shoving or rutting, cannot be detected by using the proposed procedure focused on two-dimensional images. Three-dimensional image processing and the analysis of z-acceleration experienced by the vehicle may help to solve this issue. One interesting solution might be to embed an accelerometer in the video camera (Coenen and Golroo 2017). The images used for training in this study were collected under favorable weather and lighting conditions, but underexposed or overexposed images might not be classified properly. These images usually occur when the vehicle moves from darkness into light and vice versa, yielding a sudden large lighting variation. Therefore, researchers need to collect additional images under non-favorable conditions to extend the testing and application of the proposed procedure. Applying this methodology to rural roads would require the use of a more sophisticated video camera to avoid blurred images during pavement data collection.

DATA AVAILABILITY STATEMENT

The pavement deterioration image dataset that supports the findings of this study is available from the corresponding author upon reasonable request, whereas the code of the Concatenated Convolutional Neural Network is available at https://doi.org/10.5281/zenodo.4738913.

ACKNOWLEDGMENTS

The study presented in this paper is part of the research project titled “SIMEPU – Sistema Integral de Mantenimiento Eficiente de Pavimentos Urbanos”, funded by the Spanish Ministries of Science and Innovation and Universities, as well as the European Regional

Development Fund under Grant RTC-2017-6148-7. The authors also acknowledge the support of partner companies “Pavasal Empresa Constructora, S.A.” and “CPS Infraestructuras, Movilidad y Medio Ambiente, S.L.” and the Valencia City Council.

REFERENCES

AASHTO (2011). *Transportation Asset Management Guide: A Focus on Implementation*. American Association of State Highway and Transportation Officials.

ASTM (2003). *Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys*. Publication No. ASTM D6433-11, American Society for Testing and Materials. West Conshohocken.

Amhaz, R., Chambon, S., Idier, J., and Baltazart, V. (2016). “Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection.” *IEEE Transactions on Intelligent Transportation Systems*, *17*(10), 2718-2729.

Bull, A. (2003). *Traffic Congestion: The Problem and How to Deal with It*. United Nations Publications, 87.

Carr, T. A., Jenkins, M. D., Iglesias, M. I., Buggy, T., and Morison, G. (2018). “Road crack detection using a single stage detector based deep neural network.” In *2018 IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS)*, 1-5. IEEE.

Chambon, S., and Moliard, J. M. (2011). “Automatic road pavement assessment with image processing: review and comparison.” *International Journal of Geophysics*, 1-20.

Coenen, T. B., and Golroo, A. (2017). “A review on automated pavement distress detection methods.” *Cogent Engineering*, *4*(1), 1374822.

Hadjidemetriou, G. M., Vela, P. A., and Christodoulou, S. E. (2018). “Automated pavement patch detection and quantification using support vector machines.” *Journal of Computing in Civil Engineering*, *32*(1), 04017073.

Hajj, E. Y., Loria, L., and Sebaaly, P. E. (2010). “Performance evaluation of asphalt pavement preservation activities.” *Transportation Research Record*, *2150*(1), 36-46.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

Hicks, R. G., Simpson, A. L., and Groeger, J. L. (2011). *Pavement Management Practices in State Highway Agencies: Madison, Wisconsin Peer Exchange Results* (No. FHWA-HIF-11-035). United States. Federal Highway Administration. Office of Asset Management.

Jenkins, M. D., Carr, T. A., Iglesias, M. I., Buggy, T., and Morison, G. (2018, September). “A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks.” In *2018 26th European Signal Processing Conference (EUSIPCO)*, 2120-2124.

Kingma, D. P., and Ba, J. (2014). “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980*.

Kraemer, C., Pardillo, J. M., Rocci, S., Romana, M., Sánchez, V., and del Val, M.A. (2004). *Ingeniería de Carreteras. Vol. II*. Editorial McGraw Hill.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., and Omata, H. (2018). “Road damage detection using deep neural networks with images captured through a smartphone.” *arXiv preprint arXiv:1801.09454*.

Miller, J. S., and Bellinger, W.Y. (2003). *Distress Identification Manual for the Long-term Pavement Performance Project* (No. FHWA-RD-03-031). Federal Highway Administration, United States Department of Transportation, Washington, DC.

Mohanraj, V., Huang, L., and Asghari, H. (2018, October). “Improved automatic road crack detection and classification.” In *2018 International Conference on Image and Video*

Processing, and Artificial Intelligence, *10836*, 108360A. International Society for Optics and Photonics.

Oliveira, H., and Correia, P. L. (2012). "Automatic road crack detection and characterization." *IEEE Transactions on Intelligent Transportation Systems*, *14*(1), 155-168.

Osorio, A., Chamorro, A., Tighe, S., and Videla, C. "Calibration and validation of condition indicator for managing urban pavement networks." *Transportation Research Record*, 2014, 2455, 28–36.

Park, S., Bang, S., Kim, H., and Kim, H. (2019). "Patch-based crack detection in black box images using convolutional neural networks." *Journal of Computing in Civil Engineering*, *33*(3), 04019017.

Parreño-Lara, M., Paredes, R., Llopis-Castelló, D., García-Segura, T., and Pellicer, E. *Convolutional Neural Networks for the identification and quantification of basic urban pavement distresses*. https://doi.org/10.5281/zenodo.4738913

Perera, R. W., and Kohn, S. D. (2001). *LTPP data analysis: Factors Affecting Pavement Smoothness*. Washington, DC, USA: Transportation Research Board, National Research Council. (http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w40-a.pdf)

Radopoulou, S. C., and Brilakis, I. (2017). "Automated detection of multiple pavement defects." *Journal of Computing in Civil Engineering*, *31*(2), 04016057.

Ragnoli, A., De Blasiis, M. R., and Di Benedetto, A. (2018). "Pavement distress detection methods: A review." *Infrastructures*, *3*(4), 58.

Reggin, A., Shalaby, A., Emanuels, R., and Michel, G. (2008). "Urban considerations for using road roughness to manage road networks." In *7th International Conference on Managing Pavement Assets* (p. 8).

Ronneberger, O., Fischer, P., and Brox, T. (2015). “U-net: Convolutional networks for biomedical image segmentation.” In *International Conference on Medical Image Computing and Computer-assisted Intervention*, 234-241. Springer, Cham.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... and Berg, A. C. (2015). “Imagenet large scale visual recognition challenge.” *International Journal of Computer Vision*, *115*(3), 211-252.

Setyawan, A., and Kusdiantoro, I. (2015). “The effect of pavement condition on vehicle speeds and motor vehicles emissions.” *Procedia Engineering*, *125*, 424-430.

Shi, Y., Cui, L., Qi, Z., Meng, F., and Chen, Z. (2016). “Automatic road crack detection using random structured forests.” *IEEE Transactions on Intelligent Transportation Systems*, *17*(12), 3434-3445.

Wang, H., Chen, Z., and Sun, L. (2013). “Pavement roughness evaluation for urban road management.” In *ICTE 2013: Safety, Speediness, Intelligence, Low-Carbon, Innovation* (pp. 2709-2713).

Wang, X., and Hu, Z. (2017, August). “Grid-based pavement crack analysis using deep learning.” In *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, 917-924. IEEE.

Xia, W. (2018, July). “An approach for extracting road pavement disease from HD camera videos by deep convolutional networks.” In *2018 International Conference on Audio, Language and Image Processing (ICALIP)*, 418-422. IEEE.

Zhang, L., Yang, F., Zhang, Y. D., and Zhu, Y. J. (2016, September). “Road crack detection using deep convolutional neural network.” In *2016 IEEE International Conference on Image Processing (ICIP)*, 3708-3712. IEEE.

Zou, Q., Cao, Y., Li, Q., Mao, Q., and Wang, S. (2012). "CrackTree: Automatic crack detection from pavement images." *Pattern Recognition Letters*, *33*(3), 227-238.

Table 1. Studies using CNNs for pavement crack detection

Study	Camera	Resolution (pixels)	# images	Block (pixels)
Zhang et al. (2016)	Smartphone	3264 x 2448	500	99 x 99
Wang and Hu (2017)	iPhone 6	960 x 704	510	32 x 32 64 x 64
Jenkins et al. (2018)	iPhone 5	480 x 320	118	572 x 572
Carr et al. (2018)	iPhone	4000 x 3000	118	480 x 340
Maeda et al. (2018)	LG Nexus 5X	600 x 600	9,053	600 x 600
Park et al. (2019)	Black Box Camera	1920 x 1080	664	40 x 40

Table 2. Per-class multi-label metrics

Distress	Precision	Recall	F1 Score
Road Marking	0.9872	0.9824	0.9833
Manhole	0.9562	0.9504	0.9513
Longitudinal cracking	0.9124	0.9111	0.9115
Transverse cracking	0.9189	0.9177	0.9179
Alligator cracking	0.9200	0.9194	0.9194
Raveling	0.9056	0.9044	0.9046
Potholes	0.9134	0.9122	0.9122
Patching	0.9128	0.9125	0.9128

FIGURE CAPTIONS

Fig. 1. Convolutional Neural Network structure.

Fig. 2. Instrumented vehicle for data collection.

Fig. 3. Image preprocessing: (a) raw image; (b) filtered image; (c) division of the image into blocks.

Fig. 4. Two-step CNN procedure.

Fig. 5. ROC Curve for classification task.

Fig. 6. Methodology for pavement condition assessment.