VOL.40 Nº4 2023
https://doi.org/10.22491/rmct.v40i4.12481.en

Identification and classification of roadway distress using an integrated system with convolutional neural network

Kaio Gefferson de Almeida Mesquita*[1], Yasmin Pereira de Brito Barroso[1], Thiago de Sousa Tostes[1], Luan Pablo de Holanda Barros[1], Thiago Prezotte Reis[1], Frederico Rodrigues[1]
[1]**Imtraff Engenharia e Mobilidade**
Av. Cristiano Machado, 640 - sala 1106 - Sagrada Família, Belo Horizonte (MG), Brasil
***e-mail: kaio@det.ufc.br**

ABSTRACT: Road transportation is essential to Brazil's economy, dominating goods and passenger movements. However, the country faces challenges in road quality, with only a fraction being paved and many exhibiting structural issues. Accurate recognition of these problems is crucial for effective road interventions. Manual approaches prove costly, slow, and error-prone, leading to the emergence of convolutional neural networks (CNNs) as a solution. This study introduces a multi-label CNN method for real-time distress detection on roads, aiming to reduce errors between distress classes in both aggregated and disaggregated categories. The results show clear advantages for the model with aggregated pathology categories, demonstrating improved accuracy, recall, and average precision. Comparative analysis highlights enhanced accuracy with category aggregation and exclusion of poorly represented categories, resulting in an average error of just 7%. In conclusion, the operational model significantly contributes to optimizing the definition of road interventions.

KEYWORDS:Computer vision. Highways. Distress Detection. Convolutional Network. Flexible Pavement.

RESUMO: No Brasil, o transporte rodoviário desempenha um papel crucial na economia, responsável por movimentações de mercadorias e passageiros. No entanto, as estradas enfrentam desafios significativos, com grande parte não pavimentada e muitas apresentando problemas estruturais, como fissuras, trincas e buracos. A detecção precisa dessas patologias é essencial para a intervenção adequada nas rodovias. Este estudo propõe um método inovador utilizando redes neurais convolucionais (CNNs), um subconjunto do aprendizado profundo, para a detecção de patologias nas estradas. O foco é reduzir erros entre classes considerando categorias agregadas e desagregadas. O método mostrou vantagens claras, com melhorias na acurácia, recall e média de precisão ao usar categorias agregadas. A análise comparativa revelou que a agregação de categorias e a exclusão de categorias pouco representadas aprimoraram a precisão do modelo. Com um erro médio de 7%, o modelo já está operacional, contribuindo de maneira prática para otimizar processos de intervenções em estradas.

PALAVRAS-CHAVE: Visão Computacional. Rodovias. Detecção de Patologias. Rede Convolucional. Pavimentos Flexíveis.

1. Introduction

Transportation infrastructure far exceeds most other in-demand assets in most countries. Developed and well-maintained road networks are crucial to develop the economy of a nation, especially for those with a road-based transportation matrix, such as Brazil. Brazilian roads transport about 61% of its goods and 95% of its passengers. Although Brazilian highways span 1,720,700 km, only 12.4% of these roads are paved — of which 59% were considered inadequate in at least one criterion (geometry, signage, or pavement) — and 52.4% of the evaluated highways had some defect in their pavement [6].

Studies indicate that visible pavement distresses show signs of structural wear [3] — openings with varying directions and shapes — that can cause accidents. Detecting and properly diagnosing distresses are of paramount importance. Repairing them or monitoring their behavior in cases in which repair is impossible minimizes damage or completely solves the problem. This necessitates regular monitoring and inspections for maintenance and failure prevention. Traditional manual visual inspections by engineers or specialized technicians have limitations, such as subjective evaluations, high cost and execution time, and difficulties addressing the scale and frequency necessary to adequately monitor extensive road networks. The manual approach is falling out of favor due to its subjectivity, time, and high financial and human resource costs. Moreover, manual inspection is becoming impractical due to the required scale and

frequency. Technological advances have increased the trend for automated surveys to evaluate pavement with equipment ranging from cameras to technologies such as sensors and computer vision [16].

Over the past eight years, convolutional neural networks (CNNs) — a type of neural network widely known as deep learning — has gained remarkable popularity in many industries, especially in computer vision [11]. Convolutional neural networks lie within the deep learning universe. They have been developed to classify images as they can constantly analyze data and recognize patterns. A CNN can automate processes, but it also offers a particularly suitable tool to detect pavement distresses. Its singular effectiveness in detecting edges in images by applying convolution operations outperforms other deep learning models. This specific capability makes identifying paving problems quick and simple and improves result accuracy and reliability, making CNNs essential in applications in which accurate detection is crucial, such as road infrastructure maintenance. In addition to automating processes, CNNs are particularly suitable for detecting pavement distresses, outperforming other deep learning models due to their unique efficiency in detecting edges in images stemming from convolution operations. However, the application of deep learning techniques in pavement engineering is constantly evolving, with advances in hardware and software. A review of the literature on the application of deep learning to pavement maintenance found that most studies focused on road distresses, showing that attention has predominantly revolved on classifying, detecting, and segmenting cracks [10][14], rather than on furthering the detection of other types of less frequent distresses and on reducing detection errors, the focus of this study. Thus, manually assessing distresses loses accuracy due to human errors, requires more time than field collection itself, and presents methodological difficulties. Therefore, this study hypothesizes that models to identify distresses may lose accuracy during training due to their large number of types, being unable to generalize novel cases well and causing overfitting [9].

In this context, this research aims to propose a method to train and apply a convolutional neural network with multiple labels (creating boxes during training) to identify distresses on highways in real time, reducing aggregate errors between classes and generalizing new cases. The process seeks to automate the registration and diagnosis of highway pavement, expanding the pavement management system by comparing two models: one trained with an aggregated specific distress set and a generic one with various types of disaggregated distresses, according to the specifications of the DNIT standard (2003). This study also compared these models with manual counting and classification, using sample data from several stretches of a Brazilian highway with three lanes per route and comparing these results with performance indicators to evaluate the adequacy of the models to distress types.

2. Literature review

A road infrastructure is essential for any society. Structural defects due to aging and environmental variations can negatively affect its durability, which highlights the importance of monitoring and management strategies. Such infrastructure management generally has the preservation and extension of the useful life of long-term structures as its fundamental objective [4]. To maintain the integrity of structures, it is important to detect the beginning of any defect. Fissures, holes, and cracks constitute the most common distresses in pavements and the surface of structures; such structural degradation may threaten safety and reduce pavement life [23].

The first and most common procedure to assess the health of a structure is visual inspection, which is expensive and laborious. Moreover, manual inspection requires highly trained experts and is subject to their judgment. These limitations have motivated research in industry and academia on automated distress detection approaches [5]. Image-based approaches improve cost effectiveness due to the widespread availability of cameras and smartphones [13]. Computer vision techniques have effectively automated the identification on distresses by images. Their application has grown in recent decades [23].

Image-based distress detection studies usually fall into (i) manual methods and (ii) automatic approaches based on the extraction of characteristics. Early

studies focused on methods such as edge detectors [1], morphological operations, and thresholding [7]. Manually extracted characteristics are fed into classic machine learning algorithms, such as support vector machines, random forests, and neural networks. Several distress detection studies have been proposed based on deep learning. They are categorized by image classification, object recognition, and semantic segmentation [12].

Deep learning methods focus on learning hierarchies of features based on higher levels that combine lower-level features. Learning features at various abstraction levels can train complex functions that directly map inputs to outputs from the data [10][14].

Convolutional neural networks (CNNs) were inspired by the functioning of the visual cortex for image classification [16]. CNNs can receive an input image and assign filters that can be learned to the objects in the image, differentiating one from the other [2]. The pre-processing CNNs require is much lower than in other classification algorithms, making processing computationally manageable since the network reduces the received images without losing information.

A CNN typically requires many tagged images to obtain high prediction accuracy. However, it is often difficult to collect thousands of images and label them manually. Transfer learning can build models with fewer inputs. Instead of starting the learning process from scratch, transfer learning chooses a model that has been trained on a larger dataset (such as ImageNet) to address a similar issue. This model uses pre-trained weights and a learning rate that adapts as needed [22].

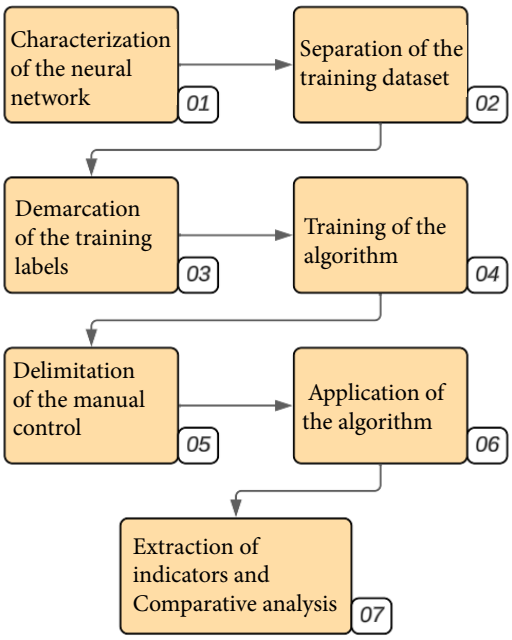
Learning rates consist of a tuning parameter in an optimization algorithm that determines step size in each iteration, moving it toward minimal loss function [17]. [20] stresses that it should vary within a range of values, rather than operating under a fixed or exponentially decreasing value. Its minimum and maximum limits are fixed, and the learning rate varies cyclically between these limits. That study proposed training for some epochs to estimate the minimum and maximum limit values.

3. Methods

A seven-stage development was proposed: (i) characterization of the neural network; (ii) separation of the training dataset; (iii) demarcation of the training labels; (iv) training of the algorithm; (v) delimitation of the manual control; (vi) application of the algorithm; and (vii) extraction of indicators and comparative analysis (Figure 1). The entire process was carried out in Python 3.10 and developed on Yolo v7 (You Only Look Once); a framework implemented with deep learning from a darknet. Based on this algorithm, implementations were carried out to optimize the speed and detection of several distresses in the same image, as will be described.

In total, three hypotheses were tested in this study: (i) training with disaggregated distress categories can negatively impact detection accuracy; (ii) distresses with more training samples tend to obtain more detections in new datasets, whereas those with fewer samples may decrease model accuracy; and (iii) the models to identify distresses may lose accuracy during training due to the large number of types, being unable to generalize new cases well.

Figure 1. - Methodology to detect distresses with a personalized neural network.

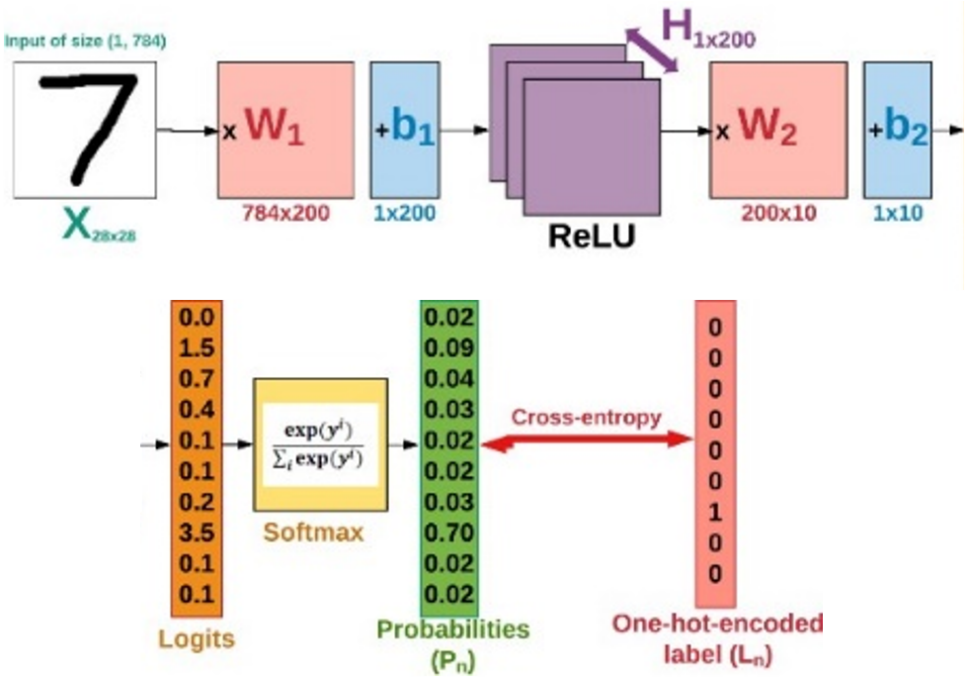


Source: Authors.

The first stage consists of characterizing and describing the convolutional neural network, as per the topic below in this section. Then, it is necessary to separate sample images for each road distress category under analysis. The highway analyzed in this study has three lanes per route, and information was collected from all lanes. Due to the Brazilian General Data Protection Law [15], the location of the highway in this case study will remain undisclosed as this section will only focus on the method to identify distresses. Then, the labels for each category were first manually then optimally demarcated according to the identification standard. The used tool was developed by the authors on Python and the pyqt5 framework to optimize this process.

The algorithm was trained on a 16-core 32-GB RAM GPU, enabling parallel processing and indicator visualization during training. Training calibrates the weights of each variable (or category) for each activation function in each layer of the neural network. The process is evaluated by a loss function comparing the original probability distribution with the trained distribution and decreasing the error with each sample picture epoch. The weights are updated by backpropagation, in which the derivative of each activation function is calculated in relation to each weight of the function to try to reduce the error in the next epoch since the derivative tends to zero when a local or global optimal point is found (Figure 2).

Figure 2 - Structure of the process of identifying objects in an image by CNN.



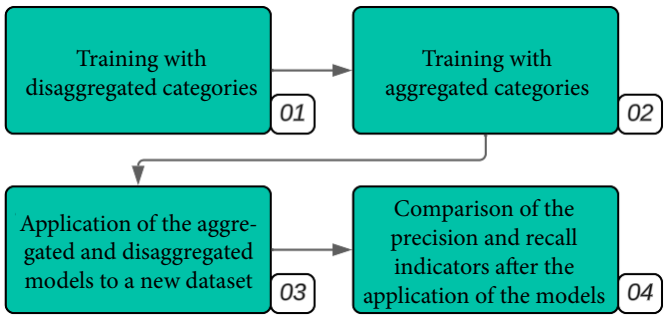
Source: [9].

A sample manual identification control was chosen for each applied model to compare human perception and artificial intelligence results according to error variation. Finally, after training the algorithms, the recall, accuracy, and mean precision indicators and the variation of these parameters between the aggregated (10 classes) and segregated distress (25 categories) models were compared. This variation stems from variations in some categories that are often difficult to identify (even manually). Thus, some categories were added, and the results were compared to test hypothesis I.

The darknet structure was chosen as it can process images in up to 0.05 seconds under a 30 frames-per-second rate. A contribution of this study involves its use of pre-trained initial weights to identify distresses, even before the new training process. The non-random initialization of the weights of each layer optimizes the training process and can ensure it avoids stagnating in an optimal place and reach the global optimal point of the required activation function.

Finally, Figure 3 objectively describes how this work intends to test the suggested hypotheses. First, the disaggregated and aggregated models (same database) were separately trained and applied to a new dataset (Step 3). The first hypothesis is tested in the training stage and the second one, in the application stage. In the fourth stage, in which the indicators are extracted, model overfitting is evaluated, thus testing the third and final hypothesis.

Figure 3 - Methodology for testing hypotheses.



Source: Authors.

3.1. Dataset and creation of labels to train distress identification models

The image database consists of photographic records of videos of a Brazilian three-lane highway. The recordings were carried out vertically with a GoPro Hero 10 Black 1,920-x-1,080-pixel resolution camera attached to the rear of a vehicle, enabling the visualization of the pavement due to its perpendicular relation to it. Thus, these images only show sections of the pavement and, at a fixed height and angle, can improve distress visualization and dimensioning. Figure 4 exemplifies the process of drawing the labels and the manual definition of the categories.

The groups demarcated for training in each model and the percentage of images per category are shown in Table 1. Only the acronyms of the distresses are shown in the tables to improve visualization. Their description is given below: W – Wear; FI – Fissures; A – Alligator Cracking without Erosion; AE – Alligator Cracking with Erosion; PH – Pot Holes; P – Patching; STC – Short Transverse Cracks; LTC – Long Transverse Cracks; SLC – Short Longitudinal Cracks; LLC – Long Longitudinal Cracks; BC – Block Cracking without Erosion; BCE – Block Cracking with Erosion; U – Upheaval; S – Shoving; Ra – Raveling; SC – Slippage Cracks; Ru – Rutting. An important observation for understanding defect types is that "fissures" configure capillary distresses that are perceptible to the naked eye at a distance equal to or below 1.50 m from the pavement. Cracks, on the other hand, show greater thicknesses than those of fissures and are thus, easily visible to the naked eye. Cracks fall into three classifications according to their thickness and erosion: FC-1, FC-2, and FC-3: those with openings greater than those of fissures (as mentioned above) but smaller than 1 mm; those with openings greater than 1 mm but without erosion at their edges; and cracks with openings greater than 1 mm with erosion at their edges, respectively. Moreover, the classification of FC-1 cracks applies only to isolated cracks, whether longitudinal, transverse, or shrinkage [8]. It is important to stress that the aggregate model compiled some categories — as explained in Table 1 — and removed those showing low occurrence to test the influence of these categories on class separation accuracy.

Figure 4 - Demarcation of the labels on the dataset.

Source: Authors.

Table 1 - Summary of training data on the aggregated (a) and disaggregated (b) models

Aggregate model	Images	Proportion
W	120	3.1%
FI	341	8.8%
A	496	12.8%
AE	61	1.6%
PH	304	7.8%
P	619	15.9%
STC	790	20.3%
LTC	751	19.3%
SLC	369	9.5%
LLC	35	0.9%

Disaggregated Model	Images	Proportion
W	120	3.1%
FI	338	8.7%
A	496	12.7%
AE	61	1.6%
PH	304	7.8%
P	619	15.8%

continue

Disaggregated Model	Images	Proportion
STC-FC1	416	10.6%
STC-FC2	348	8.9%
STC-FC3	26	0.7%
LTC-FC1	399	10.2%
LTC-FC2	312	8.0%
LTC-FC3	40	1.0%
TLC-FC1	120	3.1%
SLC-FC2	239	6.1%
SLC-FC3	10	0.3%
LLC-FC1	4	0.1%
LLC-FC2	24	0.6%
LLC-FC3	7	0.6%
BC	2	0.1%
BCE	-	-
U	1	0.0%
S	11	0.3%
Ra	5	0.1%
SC	1	0.0%
Ru	4	0.1%

Source: Authors.

3.2. Neural network training

Convolutional neural network training includes mathematical optimization by a cost function (loss function) to automatically adjust for neural network biases and weights. In this procedure, the training dataset with expected labels is introduced to the optimization procedure to find a global minimum loss function.

Training the CNN consumed a lot of computational time, and the use of non-random initial weights is of paramount importance to speed up this process. An alternative to GPUs is field-programmable gate arrays or application-specific integrated circuits. As

for the hyperparameters of the models, this study decided to retain them in both in order to evaluate the influence of the training data without external factors. Thus, a 4 lot and 350 epochs were defined.

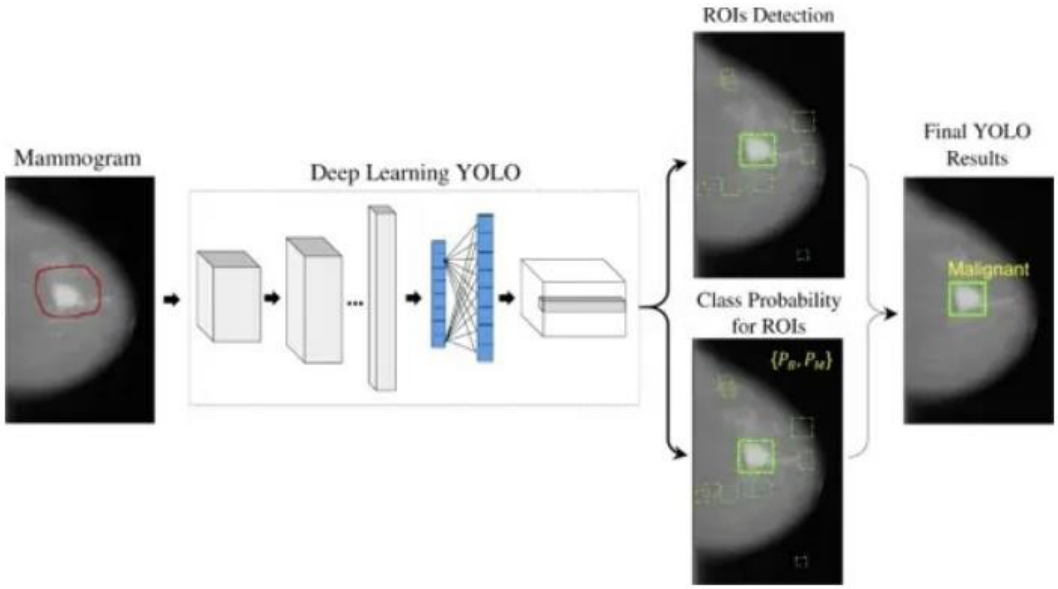
3.3. The algorithm in detail

First, the algorithm converts the images into an S--x-S bounded grid. A 19-x-19 grid was used in this study, as in [9]. Each cell is responsible for predicting five bounding boxes in case that cell has more than one object. The confidence score — which indicates how accurate the algorithm is that a given bounding box has an object — is also retrieved.

Cells also predict a class for each box, thus functioning as a classifier: a probability value is provided for each possible class. The confidence value for the bounding box and the prediction of the classes are combined into a final score, which informs the likelihood of that box containing a specific object. The 19-x-19 grid in this study resulted in 361 cells. In total, five bounding boxes are detected by each cell, totaling 1,805 boxes. Since most boxes will have an extremely low confidence value, only those with a score equal to or greater than 50% were considered. This threshold classification can be manipulated for model testing, but only a single threshold was considered in this study.

The use of Yolo stems from its effective prediction capacity in a single network epoch (Figure 5). Detection by dividing an image into several parts and executing a classifier in each such part would be necessary before this version, which would necessitate running the same classifier thousands of times over the same image [19].

Figure 5 - Example of single-epoch detection with Yolo.



Source: [19].

Building a pre-trained CNN could adequately identify pavement distresses [18]. Thus, the upper layers of the network often require huge changes in parameters, whereas the deeper layers — which have been well trained in detecting basic features (such as edges and contours) — need small modifications. The CNN is adjusted in two stages since transfer learning is used in the network architecture. The first step freezes the first layer group to avoid updating any of its parameters. The first layer group is thawed in the second step. Training resumes from the previous training state.

3.4. Indicators

The models were compared after the neural network was applied to evaluate the predictive capacity of the classification models based on test data and evaluate whether the models really "learned," i.e., if they could satisfactorily extend their predictions to unknown data sets. Precision, recall, mean average precision (MAP), and accuracy were used as indicators to evaluate the prediction level of the classification models. Recall measures the ability of a model to find all positive examples. It is calculated as the ratio of correct positive predictions to the total number of positive examples in the dataset. This metric aims to minimize false negatives. Accuracy measures the ability of a model to avoid labeling a negative example as a positive one, which is calculated as the ratio of correct positive predictions to the total number of positive predictions. Since different classes of objects can have different levels of difficulty in detection, MAP calculates the average precision for all classes in the dataset, offering a general view of the performance of a model across all classes. In addition to the indicators above, model accuracy in relation to manual identification for each category was also tested [21].

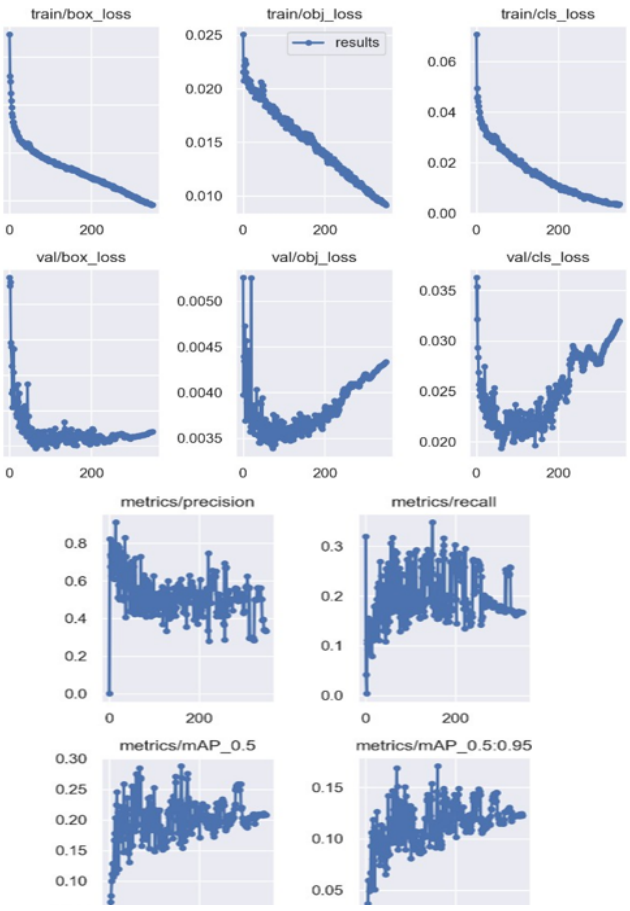
4. Results

Table 1 shows the data used to train each model and Table 2, the results of the model with the aggregated classes, that with the disaggregated classes, and the manual count. It is worth noting that the process limited itself to an 80.5-km highway stretch. Finally, this study compared the extracted results with each other, validating them with the control and analyzing the measurement error in distress detection.

4.1. Training results

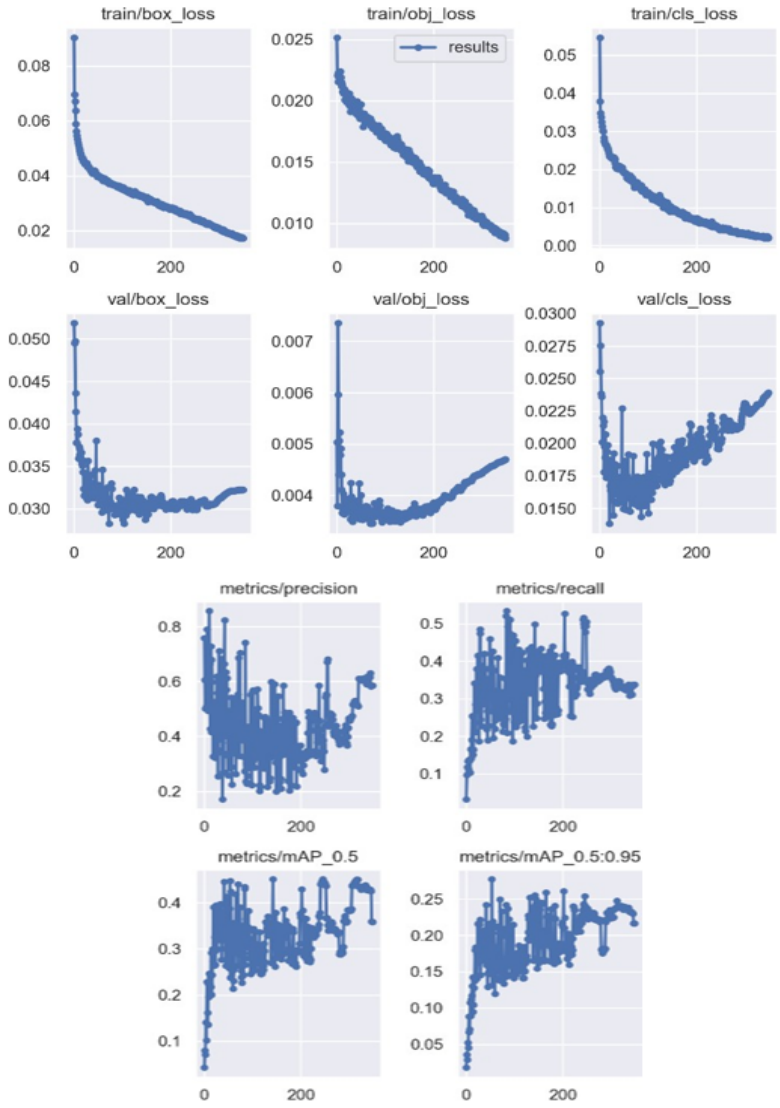
The following figures refer to the disaggregated and aggregated models, respectively. This study generated them during training and express the metrics to assess distress identification quality. Note that, reinforcing the hypothesis of improvement of the model with aggregated classes, the metrics proved to be generally superior in this model than in that with disaggregated classes. After 300 epochs, recall begins to stabilize at around 0.18 for the disaggregated model and at 0.34 for the aggregate one, indicating an 88% improvement. MAP consolidates itself around 0.12 for the former model and around 0.23 for the latter, indicating a 91% improvement, i.e., the capacity for positive events increases under a smaller possibility of the same class occurring due to its subclasses. Applying this model in practice requires the manual separation of the appropriate categories focusing on a certain class at a time, thus reducing the perception error due to wear. Regarding accuracy, the number of epochs failed to stabilize this metric in any case. Even so, the graph of the aggregate model indicates a trend of improvement in its final portion.

Figure 6 - Training result – disaggregated model.



Source: Authors.

Figure 7 - Training result – aggregated model.



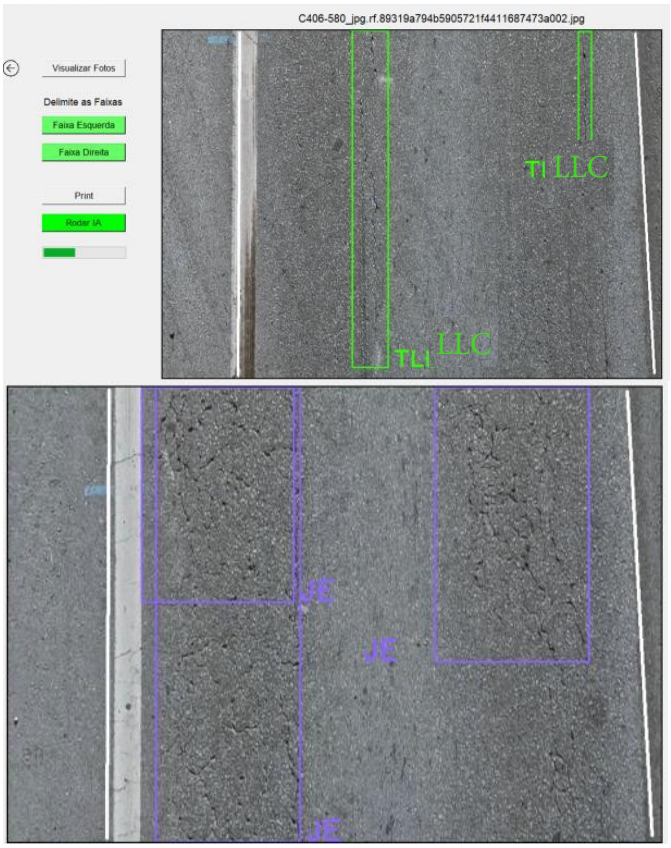
Source: Authors.

4.2. Model Applications

Figure 8 exemplifies the practical results of distress identification as explained in the method section. Note that the identification for aggregate categories showed higher accuracy in all classes, except for alligator cracking with erosion. In these cases, the disaggregated model showed a 38% lower error rate than the control. The transverse cracks and longitudinal cracks categories obtained a 57 and 62% success rates, respectively.

Table 2 reflects the differences in metrics due to the practical application of the chosen models, comparing their results with the control group. Note the evidence validating the hypothesis that aggregating categories improves model accuracy since the mean error in the aggregated model only totaled 7%, totaling 407% in the disaggregated model, making its application unfeasible. Regarding hypothesis 2 (on the relation of sample data and model improvement), this study found no noticeable increase in indicators and identifications, requiring new analyses.

Figure 8 - Application of the (a) aggregated and (b) disaggregated models.



Source: Authors.

Table 2 - Comparative prediction results

Distresses	Model Aggregated	Manual
W	1	2
FI	2	2
A	230	404
AE	58	10
PH	26	4
P	116	81
STC	9	3
LTC	3	4
SLC	66	105
LLC	125	67
Distresses	**Disaggregated Model**	**Manual**
W	415	2
FI	13	2
A	343	404
AE	36	10
PH	43	4
P	1447	81
STC-FC1	4	1
STC-FC2	10	1
STC-FC3	1	1
LTC-FC1	13	2
LTC-FC2	0	1
LTC-FC3	4	1
SLC-FC1	218	94
SLC-FC2	48	10
SLC-FC3	1	1
LLC-FC1	100	55
LLC-FC2	65	11
LLC-FC3	1	1
BC	0	1
BCE	0	1

continue

Distresses	Model Aggregated	Manual
U	0	1
S	1	0
Ra	6	1
SC	0	0
Ru	0	0

Source: Authors.

Thus, the model with aggregated classes showed a significant improvement in relation to that with disaggregated ones according to the total number of distresses and excluding those that occurred only in the disaggregated model up to 682 manually counted instances. This study highlights the deviation between instances of model identification and manual counts: 384 aggregated and 2,213 disaggregated counts, respectively.

5. Final considerations

In summary, the aggregate model showed an average error of only 7%, a recall of 0.34, and a MAP of 0.23. On the other hand, the disaggregated model obtained an alarming average error of 407%, a recall of 0.18, and a MAP of 0.12, thus validating the hypothesis that aggregating classes and removing poorly represented ones could improve the accuracy of a model since, only with these methods, the results of the aggregate model trained with the same photos of the disaggregated model proved to be superior.

The influence of the network architecture and hyperparameters may influence training as much as the amount of data since variations on activation functions, loss functions, and learning rates may directly impact the identification of edges, silhouettes, and shapes. However, this hypothesis is yet to be evaluated.

Our future studies will use a more robust database encompassing the main Brazilian highways, with better distribution of distress classes and in portions with varying pavement since distresses can have completely different appearances in different scenarios. It is of great importance for academia and the market to offer the different models of identification by computer vision to Brazilian highways since those near the North and Northeast tend to show greater wear and tear.

References

[1] ABDEL-QADER, I.; ABUDAYYEH, O.; KELLY, M. E. Analysis of Edge-Detection Techniques for Crack Identification in Bridges. Journal of Computing in Civil Engineering, v. 17, p. 255-263, 2003.
[2] Academy, D. S. Deep learning book. [S. l.]: [S. n.], 2019.
[3] ALI, L.; ALNAJJAR, F.; JASSMI, H. A.; GOCHO, M.; KHAN, W.; SERHANI, M. A. Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures. Sensors, v. 21, n. 5, 2021.
[4] CAGLE, R. F. Infrastructure Asset Management: An Emerging Direction; AACE International Transactions: Morgantown, WV, USA, 2003.
[5] CHA, Y. J.; CHOI, W.; SUH, G.; MAHMOUDKHANI, S.; BUYUKOZTUK, O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Computer-Aided Civil and Infrastructure Engineering, v. 33, p. 731-747, 2017.
[6] CNT - CONFEDERAÇÃO NACIONAL DO TRANSPORTE. Pesquisa CNT de Rodovias 2019. Brasília, DF: CNT, 2019.
[7] CORD, A.; CHAMBON, S. Automatic Road Defect Detection by Textural Pattern Recognition Based on AdaBoost. Computer-Aided Civil and Infrastructure Engineering, v. 27, p. 244-259, 2011.
[8] DNIT - DEPARTAMENTO NACIONAL DE INFRAESTRUTURA DE TRANSPORTES. Norma DNIT 005/2003 - TER. Defeitos nos pavimentos flexíveis e semirrígidos Terminologia. Rio de Janeiro: DNIT, 2003.
[9] GÉRON, A. Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, v. 1. p. 856, 2019.

[10] GOPALAKRISHANAN, K.; KHAITAN, S. K.; CHOUDHARY, A.; AGRAWAL A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr-Build Mater, v. 157, p. 322-330, 2017.
[11] GOPALAKRISHANAN, K. Deep learning in data-driven pavement image analysis and automated distress detection: A review. Data, v. 3, n. 3, 2018.
[12] HSIEH, Y. A.; TSAI, Y. J. Machine Learning for Crack Detection: Review and Model Performance Comparison. Journal of Computing in Civil Engineering, v. 34, 04020038, 2020.
[13] JAHANSHAHI, M. R.; KELLY, J. S.; MASRI, S. F.; SUKHATME, G. S. A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. Structure and Infrastructure Engineering, v. 5, p. 455-486, 2009.
[14] KHANDELWAL, R. Computer vision: Instance segmentation with maskr-cnn. [S. l.]: [S. n.], 2019. Disponível em: https://towardsdatascience.com/computer-visioninstance-segmentation-with-mask-r-cnn 7983502fcad1. Acesso em: 14 jul. 2023.
[15] BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2020. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2020/lei/l14020.htm. Acesso em: 11 ago. 2023.
[16] MASSUCATTO, J. D. P. Aplicação de conceitos de redes neurais convolucionais na classificação de imagens de folhas. Dissertação (Mestrado) - Universidade Tecnologica Federal do Paraná, Curitiba, 2019.
[17] MURPHY, K. P. Machine Learning: A Probabilistic Perspective. [S. l.]: [S. n.], 2012.
[18] NIE, K.; WANG, Pavement Distress Detection Based on Transfer Learning, 2018 5th Int Conf Syst Informatics, ICSAI 2018, no. Icsai, p. 435-439, 2018.
[19] REDMON, J. You Only Look Once Unified Real Time Object Detection. Washington, D.C.: University of Washington, 2016.
[20] SMITH, L. N. Cyclical learning rates for training neural networks. Proc - 2017 IEEE Winter Conf Appl Comput Vision, WACV 2017, no. April, p. 464-472, 2017.
[21] SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. Information Processing & Management, v. 45, n. 4, p. 427-437, 2009.
[22] TAJBAKHSH N.; SHINA, J. Y.; GURUDU, S. R.; HURST, R. T.; KENDALL, C. B.; GOTWAY, M. B.; JIANMIN, L. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? Computer Vision and Pattern Recognition, v. 35, n. 5, p. 1299-1312, 2016.
[23] YANG, F.; ZHANG, L.; YU, S.; PROKHOROV, D.; MEI, X.; LING, H. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. IEEE Transactions on Intelligent Transportation Systems, v. 21, p. 1525-1535, 2020.

VOL.40 Nº4 2023
https://doi.org/10.22491/rmct.v40i4.12481.en