A computer vision application for automatic detection of defects in urban asphalt pavement in Brazil

1st Marcos Augusto Borges
Graduate Program in Computer Science
State University of Western Paraná (UNIOESTE)
Cascavel, Brazil
marcos.borges2@unioeste.br

2nd Fabio Alexandre Spanhol
Graduate Program in Computer Science
State University of Western Paraná (UNIOESTE)
Cascavel, Brazil
Federal Technology University of Paraná (UTFPR)
Toledo, Brazil
faspanhol@utfpr.edu.br

***Abstract*—This work aims to develop an automated system for detecting damage in asphalt pavements by leveraging computer vision and deep learning techniques. It addresses key challenges such as the accurate identification and classification of various types of pavement distress. As a central contribution, we present a publicly available dataset of high-resolution digital images of urban pavements, structured in accordance with Brazilian standards and accompanied by per-image annotations curated by a technical committee. Based on this dataset, the study also presents initial experiments that implement a methodology for the automated diagnosis of road conditions. The model trained with YOLOv5 achieved an mAP@50 of 84.4%. The proposed approach is intended to support public administration in decision-making processes related to road maintenance and intervention planning, ultimately contributing to improvements in the quality, efficiency, and safety of urban road infrastructure.**

***Index Terms*—asphalt pavement defect, computer vision, deep learning, YOLO**

I. Introduction

The rapid urban growth in Brazil has increased the need for efficient road maintenance, as many urban pavements show deterioration such as potholes, cracks, and deformations.

Traditionally, defect assessment follows The Brazilian National Department of Transport Infrastructure (DNIT), an agency under the Ministry of Transport of the Federal Government, establishes in Standard DNIT 008/2003 [1], relying on manual visual inspection conducted from moving vehicles.

Recent advances in computer vision and deep learning have enabled automated detection of pavement defects from vehicle – or drone-mounted cameras, offering more precise and scalable evaluations.

In this context, the present work proposes using these technologies to automate defect detection, introducing a labeled dataset based on Brazilian standards and training deep learning models for road condition analysis.

The paper is structured as follows: Section II presents the foundational concepts; Section III reviews related studies on pavement defect detection; Section IV details the proposed methodology; Section V discusses the experimental results; and Section VI concludes the study and outlines future work.

II. Background

A. Asphalt Pavement

Pavements are vital for mobility and economic activity, with significant government investment directed toward their maintenance and expansion. In Brazil, asphalt pavements—composed of layered structures using Petroleum Asphalt Cement (CAP)—predominate. Effective maintenance is essential to extend their lifespan, ensure safety, and minimize repair and reconstruction costs.

B. Defects in Asphalt Pavement

In pavement engineering, asphalt distresses are structural or functional failures that compromise safety, comfort, and pavement durability. Potholes typically form from small cracks caused by traffic loads, which allow water infiltration, weakening the base layers and leading to progressive structural degradation and surface deformation.

The Brazilian National Department of Transport Infrastructure (DNIT) Standard 005/2003 [2] classifies pavement surface defects into the following categories: cracks, deformations (including depressions, rutting, shoving), surface defects (such as ravelling, bleeding), potholes, and patches. Each of these defects can be associated with different causes, such as vehicle traffic, climatic conditions, and the quality of materials used in pavement construction. Different types of defects in asphalt pavement are illustrated in Figure 1.

III. Related Works

Recent research has advanced pavement condition assessment through deep learning and computer vision. Early methods relied on traditional image processing (e.g., color thresholding, Gabor filtering, morphology), which were effective in controlled environments but sensitive to noise and lighting [4],

Fig. 1. Defects in Asphalt Pavement, adapted from [3].

[5]. With the emergence of Convolutional Neural Networks (CNNs), studies [6], [7] demonstrated significant gains in defect detection accuracy and efficiency using models like Single-Shot Detector (SSD), Inception V2, and MobileNet.

Subsequent works introduced semantic segmentation [8] and hybrid models combining CNNs with genetic algorithms [9] to improve precision. The YOLO family became prominent for real-time detection, with enhancements through attention mechanisms enabling deployment on embedded systems [10], [11]. More recent approaches leverage Transformer-based architectures, such as SegFormer [12], and loss functions like Cross-Entropy and Dice to refine crack segmentation. Overall, the field has evolved from basic image analysis to sophisticated deep learning frameworks, though challenges remain regarding data availability and generalization.

IV. Methodology

A. Br-AsPaveDam Dataset

We present**Br-AsPaveDam** (Brazilian Asphalt Pavement Damage Dataset) [13], a novel dataset containing 2,167 images of urban asphalt pavement damage, encompassing 3,918 annotated defects.

The dataset was created from videos captured with a bumper-mounted camera under diverse lighting, weather conditions (excluding rain), viewing angles, and traffic scenarios. Recordings were collected on roads representing a wide range of pavement conditions. RGB frames (1152 $\times$ 2048 pixels) were extracted in PNG format and manually filtered to remove low-quality images. Pavement specialists then annotated the dataset following DNIT guidelines, using bounding boxes to accurately identify and label defects in each image.

The original classes followed the nomenclature and definitions established by DNIT, namely: *Fissures*, *Deformation*, *Ravelling*, *Pothole*, *Patch*. Images and annotations were labeled based on the predominant defect, using severity to break ties, and grouped into five classes. The "Bleeding" class was removed due to potential confusion with rainwater stains, and the similar classes "Depression," "Slippage," and "Shoving" were merged into a single "Deformation" class to improve consistency and reduce annotation ambiguity. Table I shows the distribution of images among the different classes.

TABLE I
Number of labels (bounding boxes) per class in the original high-resolution dataset.

Class	# of bounding boxes
Fissures	1843
Deformation	14
Ravelling	1097
Pothole	125
Patch	839
Total	**3918**

Figure 2 shows representative examples of the Br-AsPaveDam dataset classes, illustrating the visual characteristics and diversity within each category.

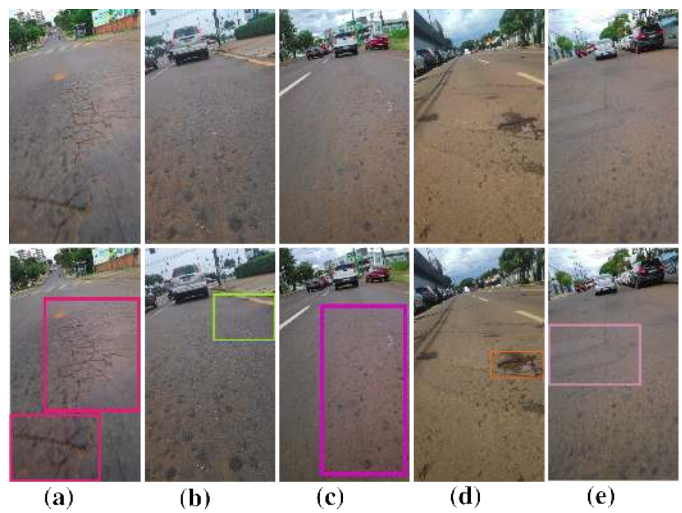

Fig. 2. Illustrative examples from the Br-AsPaveDam dataset: original images (top row) and corresponding annotated bounding boxes (bottom row). (a) Fissure, (b) Deformation, (c) Ravelling, (d) Pothole, (e) Patch.

B. Deep Learning Model

The model adopted in this study was YOLOv5-s developed by Ultralytics [14]. Like every member of the YOLO family [15], YOLOv5 is a one-stage detector: object localisation and classification are completed in a single forward pass, avoiding the proposal stages required by two-stage approaches.

C. Training Approach

All experiments were conducted in Google Colab Pro using an NVIDIA L4 GPU (24 GB VRAM). The model was trained with a mini-batch size of 16 and a fixed input resolution of 640 $\times$ 640 pixels, ensuring stable gradients, consistent spatial dimensions, and efficient GPU use while preserving fine details essential for small defect detection.

In the training pipeline, video frames were downsampled to about 640 pixels on the longer side to reduce computation while retaining detail. Preprocessing included removing EXIF metadata and ensuring consistent upright orientation before augmentation.

To address class imbalance and enhance generalization, an offline class-balanced augmentation was applied using geometric, photometric, and noise-based transformations (e.g., rotation, brightness adjustment, noise, contrast-limited adaptive histogram equalization, shadows) while preserving label integrity. Synthetic samples were generated to equalize class frequencies, and the balanced dataset was split using stratified sampling (80% training, 20% test), maintaining class proportions without requiring on-the-fly oversampling.

The model was fine-tuned from the RDD2020-IMSC [16] checkpoint, replacing the original eight-class detection head with a new output branch for five target classes (*Fissures*, *Deformation*, *Ravelling*, *Pothole*, *Patch*) while keeping the backbone and neck unchanged. This domain-specific initialization leveraged prior knowledge of asphalt textures and defects, improving convergence and generalization.

D. Evaluation Metrics

The YOLO model's performance in detecting and classifying pavement defects was evaluated using standard classification metrics from the scikit-learn library, providing a comprehensive assessment of its accuracy, robustness, and reliability across different defect types.

True Positives (TP) are correctly detected defects, False Positives (FP) are incorrect defect predictions, and False Negatives (FN) are missed defects. Using TP, FP, and FN, Precision, Recall, and F1-score are calculated to evaluate the model's classification performance, according to Equations 1, 2, and 3, respectively.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

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

Standard computer vision metrics, including Intersection over Union (IoU) and Mean Average Precision (mAP), were used to evaluate the model's spatial accuracy, measuring bounding box overlap and overall detection performance across all pavement defect classes.

For each class c, the Average Precision is computed as (4):

$$\text{AP}_c = \int_0^1 p_c(r)\, dr \tag{4}$$

where $p_c(r)$ is the precision as a function of recall r for class c.

The mean of these values across all C classes yields the mAP (5):

$$\text{mAP} = \frac{1}{C} \sum_{c=1}^{C} \text{AP}_c \tag{5}$$

Equation 6 defines the IoU, where B_p is the predicted bounding box, B_{gt} is the ground-truth bounding box, $\cap$ denotes the intersection (overlapping area), and $\cup$ denotes the union (total area covered by both boxes).

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{|B_p \cap B_{gt}|}{|B_p \cup B_{gt}|} \tag{6}$$

V. Initial Results

In addition to reporting the conventional mAP@50 metric, we also evaluate the detector using the COCO-style IoU thresholds, denoted as mAP@50:95. This metric computes the average precision over ten IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05, offering a more stringent and comprehensive assessment of the model's localization performance. The best-performing checkpoint, obtained at epoch 76, achieved a mAP@50 of 84.4% and a corresponding mAP@50:95 of 73.0%. At this stage, the model also reached a Precision of 83.9%, a Recall of 80.7%, and an overall F1-score of 82.3% on the complete validation set. These quantitative results are summarized in Table II.

TABLE II
DETECTION PERFORMANCE OF THE BEST MODEL (EPOCH 76).

Metric	**Value (%)**
mAP@50	84.4
mAP@50:95	73.0
Precision	83.9
Recall	80.7
F1-score	82.3

Table III presents the class-wise detection performance of the model on the validation set, evaluated at an Intersection over Union (IoU) threshold of 0.5. Among the five target classes, the highest average precision (AP) was achieved for Deformation (91.3%), followed closely by Ravelling (87.2%) and Patch (85.2%). Pothole detection yielded an AP of 83.1%, while Fissures exhibited the lowest performance at 72.3%, likely due to their subtle appearance and elongated structure, which makes them more challenging to detect reliably. The overall mean Average Precision (mAP) across all categories reached 83.4%, indicating robust and consistent detection performance across diverse types of pavement distress.

The objectness loss exhibited a smooth and consistent convergence throughout training, decreasing from an initial value of 0.148 to 0.097 by the final epoch. This steady decline indicates stable optimization dynamics and the absence of overfitting, as no sudden fluctuations or divergence were observed. These results reinforce the effectiveness of

TABLE III
DETECTION PERFORMANCE AT IOU 0.5 (VALIDATION SET).

Fissures	Deformation	Ravelling	Pothole	Patch	**mAP**
72.3	91.3	87.2	83.1	85.2	**83.4**

the RDD2020-based initialization strategy, which contributes to both robust classification confidence and accurate spatial localization across the five target distress categories (Figure 3). The low final objectness loss further suggests that the model has learned to effectively distinguish relevant features from background noise, enhancing its reliability in real-world deployment scenarios.

After 100 training epochs, the object detector achieved a mean Average Precision of **83.4%** at an **IoU of 0.5 (mAP@50)**, indicating a high level of detection accuracy.

VI. CONCLUSION

The proposed approach showcases the potential of computer vision to automate road inspection, modernizing traditional pavement monitoring methods. By releasing a publicly labeled dataset aligned with national standards, this work fills a key gap in the literature and enables reproducible research. It also supports sustainable, data-driven urban management by facilitating large-scale infrastructure monitoring and informing preventive maintenance and mobility policies.

Fig. 3. Sample predictions.

REFERENCES

[1] DNIT, "Levantamento visual contínuo para avaliação da superfície de pavimentos flexíveis e semi-rígidos procedimento," 2003. [Online]. Available: https://shorturl.at/EaOlr

[2] ——, "Defeitos nos pavimentos flexíveis e semi-rígidos terminologia," 2003. [Online]. Available: https://shorturl.at/fvPTF

[3] L. Fan, D. Wang, J. Wang, Y. Li, Y. Cao, Y. Liu, X. Chen, and Y. Wang, "Pavement defect detection with deep learning: A comprehensive survey," *IEEE Transactions on Intelligent Vehicles*, vol. 9, no. 3, pp. 4292–4311, 2024.

[4] Jo, Ryu, "Pothole detection system using a black-box camera," *Sensors*, 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4701334/

[5] A. Mohan and S. Poobal, "Crack detection using image processing: A critical review and analysis," *Alexandria Engineering Journal*, vol. 57, no. 2, pp. 787–798, 2018.

[6] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3708–3712.

[7] H. Maeda, Y. Sekimoto, and T. Seto, "Lightweight road manager: smartphone-based automatic determination of road damage status by deep neural network," in *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*, ser. MobiGIS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 37–45. [Online]. Available: https://doi.org/10.1145/3004725.3004729

[8] W. Yang, X. Zhang *et al.*, "Enhanced crack detection using improved convolutional neural network with attention modules," *Journal of Infrastructure Systems*, vol. 26, no. 3, p. 04020025, 2020. [Online]. Available: https://doi.org/10.1061/(ASCE)IS.1943-555X.0000578

[9] Q. Li, D. Zhang, Y. Chen, M. Cao, L. He, and B. Zhang, "An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection," *Image and Vision Computing*, vol. 57, pp. 130–146, 2022.

[10] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022.

[11] K. Guo, C. He, M. Yang, and S. Wang, "A pavement distresses identification method optimized for yolov5s," *Scientific Reports*, vol. 12, p. 3542, 2022.

[12] H. Li, H. Zhang, H. Zhu, K. Gao, H. Liang, and J. Yang, "Automatic crack detection on concrete and asphalt surfaces using semantic segmentation network with hierarchical transformer," *Engineering Structures*, vol. 307, p. 117903, 2024.

[13] M. A. Borges and F. A. Spanhol, "BR-AsPavDam – brazilian asphalt pavement damage dataset," 2025. [Online]. Available: https://doi.org/10.5281/zenodo.16291115

[14] Ultralytics, "Ultralytics YOLOv5: Real-time object detection," https://github.com/ultralytics/yolov5, 2023, commit f4ff536, accessed 25 Jul 2025.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in *Proc. 30th IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[16] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, A. Mraz, and Y. Sekimoto, "RDD2020: An image dataset for smartphone-based road damage detection and classification," 2021. [Online]. Available: https://doi.org/10.17632/5ty2wb6gvg.1