2025 7th Experiment@ International Conference (expat'25)
September 3-5, 2025, University of the Azores, Faial, Azores, Portugal

Enhancing Shared Vehicle Management with AI: A Study on Interior Damage Detection Models

Sandra Dixe
Computer Graphics and Vision
DTx-CoLAB
Guimarães, Portugal
0000-0003-4595-3828

Cátia Loureiro
Computer Graphics and Vision
DTx-CoLAB
Guimarães, Portugal
0000-0002-8092-5630

Beatriz Miranda
Computer Graphics and Vision
DTx-CoLAB
Guimarães, Portugal
0009-0005-4859-1449

Diana Ferreira
LASI/Algoritmi
University of Minho
Braga, Portugal
0000-0003-2326-2153

Tiago Aston
Electronics and Connectivity (EC)
Ceiia
Matosinhos, Portugal
tiago.aston@ceiia.com

Duarte Fernandes
Computer Graphics and Vision
DTx-CoLAB
Guimarães, Portugal
0000-0001-9736-5812

Hugo Peixoto
LASI/Algoritmi
University of Minho
Braga, Portugal
0000-0001-6289-2719

2025 7th Experiment@ International Conference (exp.at'25) | 979-8-3315-7663-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/EXP.AT2565440.2025.11348472

***Abstract*—Sustainable urban mobility, promoted by shared vehicles, has gained attention as a solution to reduce carbon emissions and facilitate transportation in cities. However, managing these vehicles presents challenges, such as internal damage caused by vandalism or improper use. This study evaluates the effectiveness of different Artificial Intelligence (AI) approaches for automatic interior damage detection in shared vehicles, considering their specific characteristics, such as compact and urban-oriented designs. The research also investigates whether the public dataset used, consisting of images of vehicle interiors with and without damage, is suitable for this scenario, given that shared vehicles may differ from conventional cars. Convolutional Neural Network (CNN) approaches, such as YOLOv5, YOLOv7, YOLOv8, YOLOv10 and RTMDet, as well as transformer-based models like RT-DETR and Dino, and hybrid approaches incorporating attention mechanisms, such as YOLO-NAS, were analyzed. The key question is whether damage detection in shared vehicles can generalize to highly variable images in this dataset. The diversity of vehicle interiors, including differences in materials, lighting conditions, and structural configurations, presents a significant challenge for AI. Ensuring accurate damage detection despite this variability is crucial for practical applications. Therefore, it is essential to assess not only model accuracy but also their ability to generalize to new vehicle configurations not represented in the dataset. Furthermore, this study provides insights into the necessary next steps for developing a robust and effective damage detection service, contributing to better management and maintenance of shared vehicle fleets.**

***Index Terms*—Shared Mobility, Interior Damage Detection, Sustainable Mobility, Deep Learning, Computer Vision.**

I. Introduction

The adoption of shared vehicles has been increasingly promoted as a solution for sustainable urban mobility, contributing to the reduction of carbon emissions and optimizing transportation efficiency. Shared mobility services, such as carsharing, bikesharing, and ridesharing, represent a transformative shift in transportation paradigms, offering environmental, economic, and social benefits. However, as these services grow, they introduce new operational challenges, especially in terms of vehicle condition monitoring and maintenance. Traditional visual inspection methods were highly based on manual analysis or rule-based techniques [1], which have limitations in dynamic and varied environments, such as the interior of shared vehicles.

However, maintaining these vehicles presents significant challenges, particularly in automatic detection of interior damage caused by vandalism or misuse. Detecting not only natural wear but also acts of intentional misuse is essential for ensuring service reliability and user safety, especially given the high turnover and exposure of shared vehicles to diverse user behaviors. Beyond identifying natural wear and tear, it is crucial to address acts of vandalism or intentional misuse during the journey, enabling a more effective response to mitigate these impacts. Automated monitoring and assessment of vehicle interior conditions are essential for ensuring service quality, user satisfaction, and operational efficiency.

Recent advancements in computer vision and Machine Learning (ML) offer promising solutions for automated interior damage detection, leveraging deep learning techniques to analyze vehicle conditions in real time. While numerous studies have explored damage detection in industrial applications, in-vehicle damage remains an underexplored field, especially when considering the unique characteristics of shared vehicles, such as their compact design and increased exposure to both wear and intentional misuse. These features require specialized algorithms capable of accurately detecting and classifying different types of interior damage. However, as this study will show, even the most recent and powerful architectures struggle to achieve consistent performance when faced with highly variable and under-representative data. This highlights the fact that architectural advances alone may not overcome fundamental data limitations, which remain a major obstacle to effective generalization in real-world applications.

This study evaluates multiple state-of-the-art AI models, including convolutional neural network (CNN)-based ap-

979-8-3315-7663-9/25/$31.00 ©2025 IEEE

2025 7th Experiment@ International Conference (exp.at'25) | 979-8-3315-7663-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/EXP.AT2565440.2025.11348472

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.

proaches such as YOLOv5, YOLOv7, YOLOv8, YOLOv10, and RTMDet, as well as transformer-based architectures like RT-DETR and Dino. Furthermore, hybrid models incorporating attention mechanisms, such as YOLO-NAS, are analyzed to assess their effectiveness in detecting damage within shared vehicle interiors.

Additionally, the research investigates the suitability of the public dataset MoLa-VI [2] for this specific task, examining how it may or may not address the unique challenges of detecting damage in shared vehicles.

We specifically examine the potential of the MoLa-VI dataset to detect signs of vandalism, focusing on its ability to capture intentional misuse across a range of vehicle types and interior settings.

A key concern is the dataset's high variability—images come from multiple vehicles and perspectives, featuring different types of damage. This, combined with design and usage differences between the dataset vehicles and actual shared vehicles, may hinder the models' ability to generalize effectively to shared mobility contexts.

The study also takes into account inference times in milliseconds (ms) on CPU and GPU test sets, as well as the model sizes evaluated in megabytes (MB). Performance metrics such as Recall, Precision, mAP50, and mAP50-95 are utilized to compare these models.

Through this comparative analysis, the study aims to identify an optimal balance between accuracy, efficiency, and real-time applicability in shared mobility scenarios. By conducting a comparative analysis of these models based on these metrics, this study aims to identify the most efficient approach for real-time damage detection, ultimately contributing to more effective maintenance strategies and enhancing the sustainability of shared mobility services.

The rest of the paper is organized as follows: Section II presents the state of the art (SoTA) for different methods in damage detection, as well as the models used in this study. The dataset used is described in Section III. The implementation of the models is presented in Section IV, and the results and discussion are provided in Section V. Finally, Section VI concludes the paper.

II. Related Work

The automated detection of damages in shared vehicle interiors has gained significant attention with the advancement of computer vision and machine learning techniques. While automated inspection of defects in industrial surfaces and infrastructure has been extensively studied [3]–[5], the application of these technologies to shared vehicles is still a developing field.

With the evolution of CNNs and Transformers based models, it has become possible to achieve high precision in identifying and classifying damages. Models like Faster R-CNN [6], SSD [7], and variants of the YOLO family [8], [9] are widely used in object detection applications due to their real-time efficiency.

More recently, models such as YOLOv5, YOLOv7, YOLOv8, and YOLOv10 have been improved to enhance rapid and accurate detection in various scenarios [10]. Simultaneously, transformers like RT-DETR [11] and Dino [12] have demonstrated excellent performance in detection and segmentation tasks, leveraging self-attention mechanisms to improve spatial feature representation. Additionally, hybrid models like YOLO-NAS [13] integrate attention mechanisms to balance accuracy and computational efficiency, making them a promising option for shared vehicle applications.

While defect detection in metal surfaces, textiles, and industrial materials has been well studied [14], applying these methodologies to inspect shared vehicle interiors still faces challenges. Vehicle interiors are composed of diverse materials such as leather, fabric, plastic, and rubber, which exhibit varying wear and degradation patterns. Studies such as [15] explore synthetic and real dataset generation to improve human pose detection in vehicle interiors, an approach that can also be applied to damage inspection.

Assessing damages in shared vehicle interiors requires consideration of adverse scenarios such as varying lighting conditions, the presence of objects obstructing the view, and subtle damages that may be confused with the natural texture of materials. Furthermore, detecting vandalism and intentional misuse adds a new level of complexity, requiring models capable of distinguishing between natural wear and damage caused by users.

The creation of databases for vehicle damage detection is still limited. Many studies use generic datasets, such as COCO [16] and TAO [17], which contain a wide variety of classes but are not specific to vehicle inspection. Some studies, such as that of [15], propose the automated generation of synthetic and real images to capture specific conditions inside vehicles. However, there is a gap in the literature regarding a detailed analysis of model performance on datasets containing real images of damages in shared vehicle interiors.

The MoLa-VI dataset [2] contains vehicle interior images with a focus on detecting defects such as damage, stains, and dirt. The study by [18] utilized this dataset to train and evaluate segmentation models based on CNN, such as DeepLabV3+ and U-Net, achieving mean accuracies of 67.60% and approximately 28%, respectively. However, these models (CNNs) have been surpassed by more recent approaches that leverage advanced architectures, including transformers and hybrid models with attention mechanisms. In this context, we aim to update the study initiated in [18] by evaluating state-of-the-art models and investigating whether the knowledge generated is transferable to more specific scenarios, such as damage detection in shared vehicles.

III. Dataset

For the evaluation of damage detection models in vehicle interiors, the public knowledge base MoLa-VI Dataset [2] was selected. This dataset contains 3477 images of distinct vehicle interiors, annotated with relevant damage categories. Its selection is justified by the fact that it is, to the best of our

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.

knowledge, the only publicly available dataset in the literature specifically focused on annotated interior vehicle damage, making it the primary and most comprehensive reference for evaluating models in this domain.

Despite an extensive search, no other datasets with comparable scope, quality, or specificity for interior damage were found. Most existing datasets focus on external damage or general in-cabin environments without damage annotations. Therefore, the use of MoLa-VI enables a rigorous and realistic evaluation while also reflecting the current limitations of publicly available resources in this field.

The damage classes present in this knowledge base are:

- Cuts;
- Wear;
- Broken/Cracked;
- Dirt;
- Stains.

1) Knowledge Base Characteristics: The MoLa-VI Dataset consists of images captured from the interior of 135 different vehicles, providing a diverse representation of scenarios commonly encountered in shared mobility services. The dataset includes images taken from 9 distinct perspectives, captured using cameras at various angles and distances, as illustrated in Fig. 1.



Fig. 1. Example of the perspectives from positions P1 to P9 captured through the camera inside each vehicle. Adapted from [2].

In order to understand the class distribution in the dataset, we can refer to Table I, which provides a brief summary of the dataset composition. It shows how many images contain each damage class, the number of objects found in the dataset per class, the average number of objects per image for each class, and finally, the average percentage of the object area found per image. We can conclude, for example, that the 'Cut' damage category has the smallest area compared to the other damage categories, which makes sense considering the nature of this damage.

In the study and evaluation of the selected AI models, it was ensured that the version and configuration of the dataset were exactly the same. Of the total available data (3477 images), 70% were allocated for training (2432 images), 20% for validation (697 images), and 10% for testing (348 images). The original resolution of the images (1920×1080) was retained for dataset description and visualization purposes, while a resized version (640×640) was used during training, as detailed in Section IV.

TABLE I
GENERAL STATISTICAL REPRESENTATION AND CLASS DISTRIBUTION.

Class	Images	Objects	Count (%)	Area (%)
Cuts	404	766	1.9%	1.92%
Wear	556	964	1.73%	6.12%
Broken/Cracked	75	90	1.2%	4.12%
Dirt	639	1679	2.63%	3.22%
Stains	1215	3725	3.07%	10.19%

Table II illustrates the distribution of classes across the training (training, validation and testing) sets in terms of quantity and percentage. It is also observed that there are 1689 images with at least one class, while 1788 images do not contain any damage class. In Fig. 2, we can see a sample of some images from the dataset, with the respective classes marked in the Bounding Box format. In Fig. 3, we can observe an image from perspective P4 (Fig. 1), which contains 4 damage objects (i.e., 2 "Dirt", 1 "Wear", and 1 "Stain").

TABLE II
INFORMATION ON THE CLASS DISTRIBUTION ACROSS THE TRAINING, TESTING, AND VALIDATION SETS.

Class	Train	Valid	Test	Total
Stains	838	256	121	1215
Dirt	438	131	70	639
Wear	373	130	53	566
Cut	290	80	34	404
Broken/Cracked	52	17	6	75
Partially Marked	1167	346	176	1689
Not Marked	1265	351	172	1788
Images	2432	697	348	3477

IV. IMPLEMENTATION

In this section, the methodology and study conducted to evaluate AI models applied to the detection of interior vehicle damage using the MoLa-VI dataset will be presented [2]. The models were trained and evaluated on two distinct systems:

- **System 1:** Equipped with an NVIDIA GeForce RTX 3070 Laptop GPU, paired with an AMD Ryzen 7 5800H processor with Radeon Graphics, 24 GB of RAM, and a 64-bit architecture.
- **System 2:** Equipped with an NVIDIA GeForce RTX 2080 Ti, paired with an AMD Ryzen 9 3900X processor, 128 GB of RAM, and a 64-bit architecture.

The dataset was split into 70:20:10 for training, validation, and testing, respectively. Images were resized to 640×640 pixels to balance detection accuracy and computational efficiency, as is common in YOLO-based models. The SGD optimizer was used for most models, except for the transformer-

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.



Fig. 2. Samples of images from the MoLa-VI Dataset [2] with ground truth.

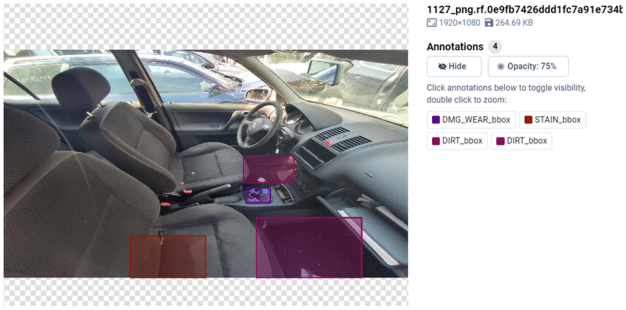


Fig. 3. Sample from the MoLa-VI Dataset [2], featuring an image captured from perspective P4, with 2 instances of 'Dirt' (pink), 1 instance of 'Wear' (purple), and 1 instance of 'Stain' (red).

based models (DINO and RT-DETR) and YOLO-NAS, where AdamW yielded better results, supported by recent studies highlighting its advantages in large-scale attention-based architectures [19].

Training was conducted over 400 epochs with a learning rate of 0.001 and a batch size of 8, based on preliminary experiments aimed at ensuring convergence and optimal performance. Alternative configurations led to instability or reduced mAP. Inference times were measured on both GPU and CPU using only System 1, ensuring consistency and fair performance comparisons across all models.

A. AI Models for Interior Damage Detection

Table III presents the categorization of the models used in this study, which represent the SoTA and encompass all major model typologies. These models were applied to the dataset MoLa-VI described in Section III.

V. Results and Discussion

The training results of the models selected in Section IV-A (Table III) using the dataset presented in Section III are shown in Table IV.

TABLE III
Model Categories

Category	Model
CNN-Based	YOLOv5
	YOLOv7
	YOLOv8
	YOLOv10
	RTMDet
Transformers	RT-DETR
	DINO
CNN with Self-Attention	YOLO-NAS

A. Overall Models Performance

- YOLO-NAS S achieved the highest recall (39.9%) and mAP50-95 (22%), demonstrating strong detection capabilities across different Intersection over Union (IoU) thresholds. However, this improved performance comes with a trade-off: significantly higher inference time, especially on the GPU, and a relatively large model size (244 MB).
- RT-DETR L obtained the second-highest recall (37.7%) and the best mAP50 (35%), indicating strong detection accuracy at the 50% IoU threshold. Additionally, it proved to be efficient in inference speed, with relatively low latency on both GPU (55.079 ms) and CPU (42.839 ms).
- YOLOv5 and YOLOv8 demonstrated a balanced trade-off between recall and precision, with larger versions (M, L, X) achieving better precision compared to smaller versions (N, S).
- Confusion matrices revealed frequent misclassifications, especially between similar classes like "Dirt" and "Stain," with many cases mistaken for background. This underscores the challenge of detecting subtle, low-contrast damage and the effect of visual ambiguity on model performance.

B. Models Efficiency

- RT-DETR L was the most efficient model in terms of inference time, achieving 55.079 ms on GPU and 42.839 ms on CPU, while maintaining high recall and the best mAP50 performance.
- YOLOv8 S offered a well-balanced trade-off between precision (48.7%), recall (22.7%), and inference time (10 ms on GPU, 116 ms on CPU), making it a strong candidate for real-time applications where both speed and accuracy are crucial.
- DINO RN50 was the slowest transformer-based model, with the highest inference time (3507 ms on GPU), a relatively large model size (199.32 MB), and only average performance in precision and recall, making it less suitable for time-sensitive applications.

C. Trade-offs between Model Size and Performance

Larger models like YOLOv5 X (165.58 MB) and YOLO-NAS L (851 MB) don't necessarily outperform smaller ones

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.

TABLE IV
PERFORMANCE OF MODELS FOR INTERIOR DAMAGE DETECTION ON THE TEST SET, COMPARING PRECISION, RECALL, AND MAP ACROSS DIFFERENT ARCHITECTURES. THE EVALUATION WAS CONDUCTED WITH AN IOU THRESHOLD OF 0.6 AND A MINIMUM CONFIDENCE OF 0.001. INFERENCE TIME WAS MEASURED AFTER A WARM-UP OF 5 IMAGES TO OPTIMIZE HARDWARE PERFORMANCE. NOTE: MODEL VARIANTS ARE DENOTED BY SIZE (N = NANO, S = SMALL, M = MEDIUM, L = LARGE, X = EXTRA-LARGE), REFLECTING INCREASING MODEL COMPLEXITY AND CAPACITY.

Models		**Recall (%)**	**Precision (%)**	**mAP50 (%)**	**mAP50-95 (%)**	**(ms)**		**MBytes**
						CPU	**GPU**	
YOLOv5	N	23.2	40.4	22.6	9.69	49.5	11.1	3.64
	S	28.5	47.5	27.1	11.4	94.2	15.4	13.71
	M	32.5	58.4	30.5	13.3	215	15.2	40.23
	L	32.5	50	31.9	15.8	347.6	20.2	88.56
	X	34.2	49.4	32.1	15.3	830.4	29	165.58
YOLOv8	N	16.8	66.9	20.4	8.24	64.8	14.21	5.97
	S	22.7	48.7	22.8	11	116	10	21.49
	M	24.3	59.8	27	13.3	309.3	16	49.62
	L	24.9	61.2	26	12.5	556.6	26.52	83.61
YOLOv7	-	24.8	26.7	15.1	6.59	772.58	12.31	71.37
	X	28.7	40.3	23.8	10.5	526.06	16.84	135.55
YOLOv10	N	22.5	38.6	21.8	10.5	120.2	16.016	5.51
	S	26	51	27.5	14.6	229.5	14.993	15.7
	M	26.2	57.2	25	10.8	430.9	18.845	31.9
	L	28.6	50.3	25.8	12.6	769.3	28.169	49.7
	X	26.8	57.3	27	12.5	952.7	30.064	61.1
	B	27.7	47.7	24.8	11.5	602.3	21.548	39.5
RTMDet	T	8.2	25	8.4	3.2	224.29	37.54	27.79
RT-DETR	L	37.7	53.6	35	14.5	42.839	55.079	63.1
DINO	RN50	12.2	49	17	8.4	3507	88.48	199.32
YOLO-NAS	S	39.9	26.2	23.9	22	358.07	242.5	244
	M	28.6	41.2	24.2	21.8	519.67	201.95	649
	L	28.9	42.1	23.4	21.9	922.35	337.9	851

like RT-DETR L (63.1 MB), likely due to high dataset variability (damage types, lighting, vision setup, focal plane, and class traits), which limits the benefits of increasing model complexity.

D. Impact of the Training and Evaluation System

The models were evaluated on two different systems; however, inference time measurements were consistently performed on System 1 (NVIDIA GeForce RTX 3070 Laptop GPU, AMD Ryzen 7 5800H processor with Radeon Graphics, 24 GB RAM, 64-bit architecture). This approach ensures fair comparisons across models, as all inference times are based on the same hardware conditions.

E. Trade-offs between Performance on CPU and GPU

Models run on the GPU showed considerably lower inference times, although the times recorded on the CPU were also considered satisfactory.

- **CPU:** Models such as RT-DETR L and YOLOv5 N are the most efficient for CPU use due to their low inference times. Models like DINO RN50 and YOLO-NAS L have very high inference times, making them impractical for efficient CPU use;
- **GPU:** Models like YOLOv8 S, YOLOv5 M, and RT-DETR L perform exceptionally well on the GPU, with very low inference times, making them ideal for real-time processing applications. On the other hand, models like YOLO-NAS S and YOLO-NAS L have high inference times, which limits their use for applications requiring fast processing.

Regardless of the hardware, compression techniques should be explored to improve the performance of the selected AI model.

This study utilized the same MoLa-VI dataset analyzed in [18], but evaluated more recent and advanced state-of-the-art detection models (Table III). Despite technological advancements and improvements in model architectures, the results obtained confirm the conclusion of [18]: the models' generalization ability remains limited due to the insufficiency of the dataset. The high variability of damage in shared vehicles, combined with the need for a more representative dataset, compromises the performance of the tested models. Therefore, to effectively address the challenge of detecting damage in shared vehicle interiors, it is essential to develop a more specific dataset that accurately reflects real-world scenarios and enables the training of models with greater robustness and precision.

Compared to the results reported in [18], which focused on CNN-based segmentation models, our evaluation using recent detection architectures—including YOLO variants, transformers, and hybrids—achieved moderate gains in mAP50 and recall. Nevertheless, both studies converge on the same key limitation: model performance is primarily constrained by the dataset's lack of diversity and representativeness. This reinforces the idea that improvements in model architecture alone are not sufficient unless accompanied by better data.

VI. CONCLUSION AND FUTURE WORK

In conclusion, the challenges posed by imbalanced datasets with insufficient variability and limited size directly impact the

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.

performance of detection models, affecting key metrics such as mAP50, mAP50-95, Recall, and Precision. Class imbalance, lack of data variability, and reduced dataset size hinder the model's ability to learn and generalize, leading to inconsistent performance, especially under varying conditions and with new data.

Among the analyzed models (Table IV), RT-DETR L stands out for its ability to balance precision, recall, and efficiency, making it suitable for environments requiring robust inference-time performance. YOLO-NAS S, while excelling in recall, struggles with inference time, making it less viable for real-time applications. Meanwhile, the smaller YOLOv5 and YOLOv8 models prove to be solid intermediate options, offering a satisfactory balance between performance and speed, making them versatile for various implementations.

In the context of damage detection in shared vehicles, selecting the optimal model is crucial, as both accuracy and speed are essential to ensure vehicle integrity and provide an efficient user experience. The MoLa-VI dataset, one of the few available in the literature, was used in this study. However, it has limitations: it does not fully capture the variety of damage seen in shared vehicles, where incidents of vandalism and certain damage types are more frequent. Additionally, the interiors of shared vehicles may exhibit lower variability compared to the dataset used, which could compromise the model's ability to generalize.

Although the MoLa-VI dataset is a valuable resource, its limitations in terms of damage variability reflect the broader challenge of dataset scarcity in this field. As it stands, more comprehensive datasets are needed to represent the true diversity of damage in shared vehicles, which would improve the robustness of the models. This limitation impacted the performance of the models, as they struggled with the high variability in damage types, resulting in limited generalization capabilities.

For future work, applying fine-tuning to transformer-based models is a promising approach, as these architectures are highly data-dependent. This strategy could enhance the models' adaptability to the specific context of damage detection in shared vehicles, contributing to more robust and effective systems. Furthermore, given the limitations of publicly available datasets, there is an urgent need for collaborations aimed at creating diverse and comprehensive datasets that accurately reflect the wide range of damage types and conditions found in shared vehicle interiors.

Such efforts could provide a stronger foundation for training models capable of generalizing across various environments, including those with frequent incidents of vandalism or atypical damage types, thereby significantly improving the performance of automatic damage detection systems.

ACKNOWLEDGMENT

This work has been supported by the European Union under the NextGenerationEU, through a grant of the Portuguese Republic's Recovery and Resilience Plan (PRR) Partnership Agreement, within the scope of the project BE.NEUTRAL – Agenda da Mobilidade para a neutralidade carbónica das cidades (Project ref. nr. 35 - C644874240-00000016).

REFERENCES

[1] Huang, Y.; Xiang, Z. RPDNet: Automatic Fabric Defect Detection Based on a Convolutional Neural Network and Repeated Pattern Analysis. Sensors 2022, 22, 6226. https://doi.org/10.3390/s22166226.
[2] Dixe, Sandra; Leite, João; Azadi, Sahar; Fonseca, Jaime; Borges, João (2021), "MoLa-VI Dataset", Mendeley Data, V1, doi: 10.17632/cfsr5kvry3.1.
[3] Guo, Y.; Kang, X.; Li, J.; Yang, Y. Automatic Fabric Defect Detection Method Using AC-YOLOv5. Electronics 2023, 12, 2950. https://doi.org/10.3390/electronics12132950.
[4] Ma, Y., Yin, J., Huang, F. et al. Surface defect inspection of industrial products with object detection deep networks: a systematic review. Artif Intell Rev 57, 333 (2024). https://doi.org/10.1007/s10462-024-10956-3.
[5] Jing, J., Zhang, H., Wang, J., Li, P., & Jia, J. (2012). Fabric defect detection using Gabor filters and defect classification based on LBP and Tamura method. The Journal of The Textile Institute, 104(1), 18–27. https://doi.org/10.1080/00405000.2012.692940.
[6] Ren, Shaoqing et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015): 1137-1149.
[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C., & Berg, A.C. (2015). SSD: Single Shot MultiBox Detector. European Conference on Computer Vision.
[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.
[9] Bochkovskiy, A., Wang, C., & Liao, H.M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv, abs/2004.10934.
[10] Bento, J.; Paixão, T.; Alvarez, A.B. Performance Evaluation of YOLOv8, YOLOv9, YOLOv10, and YOLOv11 for Stamp Detection in Scanned Documents. Appl. Sci. 2025, 15, 3154. https://doi.org/10.3390/app15063154.
[11] Song, X.; Fan, B.; Liu, H.; Wang, L.; Niu, J. HPRT-DETR: A High-Precision Real-Time Object Detection Algorithm for Intelligent Driving Vehicles. Sensors 2025, 25, 1778. https://doi.org/10.3390/s25061778.
[12] Caron, Mathilde et al. "Emerging Properties in Self-Supervised Vision Transformers." 2021 IEEE/CVF International Conference on Computer Vision (ICCV) (2021): 9630-9640.
[13] Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr. 2023, 5, 1680-1716. https://doi.org/10.3390/make5040083.
[14] Chen, Y.; Ding, Y.; Zhao, F.; Zhang, E.; Wu, Z.; Shao, L. Surface Defect Detection Methods for Industrial Products: A Review. Appl. Sci. 2021, 11, 7657. https://doi.org/10.3390/app11167657.
[15] Borges, J., Queirós, S., Oliveira, B. et al. A system for the generation of in-car human body pose datasets. Machine Vision and Applications 32, 4 (2021). https://doi.org/10.1007/s00138-020-01131-z.
[16] Lin, T. Y., et al. (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision (ECCV).
[17] Dave, R., et al. (2020). TAO Dataset for Object Detection: A Benchmark for Real-World Applications. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[18] Dixe, S., Leite, J., Azadi, S., Faria, P., Mendes, J., Fonseca, J. C., & Borges, J. (2021, February). In-car Damage Dirt and Stain Estimation with RGB Images. In ICAART (2) (pp. 672-679).
[19] Wang, X.,& Aitchison, L. (2024). How to set AdamW's weight decay as you scale model and dataset size. https://arxiv.org/abs/2405.13698.

Authorized licensed use limited to: b-on: UNIVERSIDADE DO MINHO. Downloaded on April 23,2026 at 11:50:03 UTC from IEEE Xplore. Restrictions apply.