

Automated Detection and Mapping of Pavement Cracks from Videos for Road Inspections

Giulio Perda¹, Mohamedelmustafa O. I. Eid¹, Nazanin Padkan^{1,2}, Salim Malek¹, Luca Morelli¹, Fabio Remondino¹

¹ 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: <http://3dom.fbk.eu> – Email: (gperda, meid, npadkan, smalek, lmorelli, remondino)@fbk.eu

² Dept. Mathematics, Computer Science and Physics, University of Udine, Italy

Commission II

KEY WORDS: crack detection, pavement inspection, road crack mapping, deep learning, artificial intelligence, low-cost.

ABSTRACT: Road inspections are essential for assessing pavement conditions and planning maintenance to extend infrastructure lifespan. Cracks are key indicators of distress, and their timely detection can reduce repair costs. While manual inspections are costly and subjective, automated crack detection using vehicle-mounted cameras and digital image processing has improved efficiency but still requires extensive manual review. Crack detection research typically focuses on improving the accuracy of the detection and feature extraction, neglecting the geospatial mapping part which is essential in pavement inspection. This paper presents an end-to-end pipeline that combines V-SLAM, deep learning (DL) and 2D-3D mapping to detect and map pavement cracks using vehicle-mounted cameras footage and provide georeferenced results interpretable inside a Geospatial Information System (GIS). Results significantly reduce inspection time by providing preliminary detections for rapid verification by expert operators in GIS environments. This enhances data management, efficient validation, advanced spatial analysis and time-based tracking of crack progression, ensuring informed decision-making and optimized maintenance planning, ultimately extending infrastructure lifespan and reducing costs.

1. INTRODUCTION

Urban mapping, inspection and monitoring normally involve high-cost mobile mapping vehicles - equipped with LiDAR, cameras and positioning sensors - driving through the built environment. Due to the high costs and limited vehicle fleets, single camera sequences are often used for collecting up-to-date information of semantic landmarks and change detection (Zhanabatyrova et al., 2023; Zhang et al., 2023; Lin et al., 2025). In particular, road inspection and monitoring are periodically conducted to assess pavement conditions and plan reconditioning and maintenance actions (Barbieri and Lou, 2024). Cracks are key indicators of early pavement distress and can accelerate deterioration, affecting driving comfort, user safety and reducing the lifespan of the infrastructure. The timely detection and monitoring of cracks are crucial steps in lowering maintenance costs. Automatic systems which can reliably quantify and classify cracks are desired since conventional manual field inspections are subjective, time-consuming – as they may require traffic interruption or diversion- and costly (Hsieh and Tsai, 2020). A first degree of automation in road inspections comes with the adoption of vehicle-mounted camera footage coupled with digital image processing techniques. Regardless, video-based monitoring still requires extensive manual evaluation and must be backed by an effective data management system to store cracks information. Generally, existing management systems do not fully leverage spatial information for visualization in a Geospatial Information System (GIS) environment. To further automate the process, research has focused on developing automated crack detection methods using 2D/3D data and machine (ML) or deep learning (DL) algorithms (Zhang et al., 2018; Jung et al., 2019; Jing et al., 2023) although a data management system which can allow an intuitive visualization by leveraging the spatial information is still needed.

1.1 Paper's aim

This contribution proposes an end-to-end pipeline which leverages Visual-SLAM, DL methods and a 2D-3D mapping process to detect and map pavement cracks from vehicle-mounted camera footage. A user-friendly and geospatially

accurate representation of road cracks inside a GIS is delivered for a faster, cheaper and more effective road condition assessment. Mapping of the damaged roads inside a GIS environment allows the operator to perform further visual analysis, evaluations and measurements of the damages. In contrast to available solutions that either rely on expensive pavement inspection systems, work on a single image level, or require expert data acquisition, the pipeline can be implemented with cheap off-the-shelf instrumentation, automatically extracts cracks from video frames and maps them in GIS, facilitating efficient inspection by experienced operators.

2. RELATED WORKS

Many crack detection and mapping methods in the road inspection domain start from image sequences acquired by pavement inspection systems (Li et al., 2017; Stricker et al., 2021; Opara et al., 2021; Wu et al., 2025) which provide an optimal view of the road and high-resolution images. In some cases, these systems are also equipped with LiDAR for enhanced surface analysis (Malone Geary and Tsai, 2021). Common working principle for these systems is the synchronized acquisition from vehicle-mounted sensors (camera or stereo system, GNSS receiver, IMU, LiDAR) to create a georeferenced 3D reconstruction of the road and perform analysis of its defects based on predefined metrics. These systems are equipped with instrumentation to generate a mapping of the road surface but do not focus on automating the detection of cracks. Automated crack detection has shifted from methods like laser and thermal testing to image-based methods due to their acquisition and processing efficiency (Munawar et al., 2021). Available image-based detection methods include image processing techniques, digital image correlation (DIC), ML and DL methods. Image processing techniques like thresholding and edge detection are usually used as preprocessing steps to reduce unnecessary information and enhance crack features before ML (Golding et al., 2022). DIC is less accurate than DL even with optimal thresholding, and requires multi-temporal images (Rezaie et al., 2020). ML algorithms need pre-defined features, and their performance deteriorates in images with complex backgrounds (pavements

with illumination changes), whereas DL algorithms automate feature extraction and offer higher accuracy (Alipour et al., 2019). The adoption of DL for image-based road inspection can further automate and speed-up detection, classification and severity assessment (Ha et al., 2022). The rise of highly efficient and highly accurate semantic segmentation and object detection algorithms - such as the YOLO family (Wu et al., 2024; Choi et al., 2024; Wang et al., 2023; Wu et al., 2023) - has made AI a leading approach in this field. With DL, cracks can be detected and classified at the image, patch or pixel level. Patches allow for more data generation in comparison to full images, whilst also providing approximate localization of the cracks. Pixel level classification allows the contour of the crack to be extracted, instead of its bounding box, enabling more accurate classification of the pavement area occupied by the crack. This approach requires meticulous annotation for training data generation and results can be used for crack features (width, length, occupied area) calculation (Hsieh and Tsai, 2020).

Few low-cost end-to-end solutions exist for crack detection and mapping, particularly for road damage, as most research prioritizes detection accuracy. However, some studies have proposed solutions that also address crack mapping. For instance, Chun et al. (2021) developed a highly accurate crack detection model, in which cracks are detected on a patch-level and the damage is mapped as pins in GIS. This approach limits further visual and quantitative analysis as pin-based mapping lacks the visual detail provided by orthophotos, furthermore, the end-user cannot easily extract crack features from the patches. Ranyal et al. (2024) used geotagged image sequences captured by a vehicle-mounted GoPro camera to detect and classify cracks through a combination of DL and image processing techniques. The severity of damage at each location was determined based on the maximum crack width and visualized using a heatmap. Heatmaps can be a good initial result, but similarly to (Chun et al. 2021), without the orthophotos, the end-user would need to go back to the damage location for visual analysis. Baduge et al., (2023) used a vehicle-mounted smartphone to capture geotagged image sequences and applied two deep learning models for crack detection and segmentation. The detected cracks were mapped as pins in Google Earth with a positioning accuracy of 10 meters. However, compared to GIS, Google Earth offers limited geospatial analysis capabilities, map production, and data storage for long-term change detection. GIS offers more advanced capabilities for further spatial analysis, enabling users to track changes over time, integrate multiple data sources, and generate detailed reports and visualizations.

Compared to other approaches, the proposed pipeline streamlines data acquisition while reducing costs, and significantly minimizes the manual effort required from the end users.

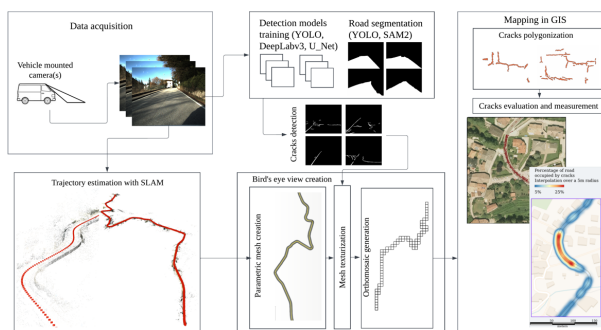


Figure 1: The proposed pipeline for road crack detection and mapping.

3. METHODOLOGY

The pipeline involves capturing images using vehicle-mounted cameras, estimating the trajectory with SLAM, segmenting the road and cracks in the input images, creating a bird's eye view of the road by generating and texturing a parametric mesh based on the estimated trajectory and the segmentation results, and producing a georeferenced orthomosaic of the cracks for use in a GIS environment. Figure 1 provides an overview of the pipeline.

Acquisition system. Sub-cm GSD images were acquired using the in-house GuPho system (Torresani et al., 2021; Padkan et al., 2023), a handheld mobile mapping stereo-vision system capable of recovering system trajectory and sparse 3D of the surveyed scene in real-time with a SLAM algorithm. The system was mounted on the front of a van, with the camera axis approximately parallel to the road axis. This configuration was chosen to ensure optimal coverage of the road surface while also providing a favorable viewing angle of the surrounding environment, which is beneficial for the automatic orientation of the images. Although stereo sequences were acquired, the presented pipeline can also be applied to monocular videos.

Trajectory estimation. The proposed workflow requires the georeferenced trajectory of the vehicle-mounted camera(s), including both position and orientation. This trajectory can be obtained through various methods, including commercial positioning solutions. In our study, camera poses are estimated with COLMAP-SLAM (Morelli et al., 2023), a tool that provides camera trajectory from a monocular or multi-camera system based on handcrafted or learning-based tie points. The keyframe selection is performed by analyzing the optical flow within the image stream and only frames exhibiting an optical flow above a predefined threshold are designated as new keyframes. The trajectory can be georeferenced by a synchronized vehicle-mounted GNSS receiver or by using ground control points.

Road segmentation. The road is segmented using the YOLOv8-world model and SAM2 (Ravi et al., 2024) to prevent the detection of cracks outside of the road and to remove moving objects (e.g. cars) that would hamper SLAM performances. The result of the segmentation process is a binary mask of the road for each keyframe.

Cracks detection. Three deep learning models are utilized to segment road cracks in one of the two streams acquired with GuPho: U-net (Ronneberger et al., 2015), DeepLabV3 (Chen et al., 2017), and YOLOv8m (Jocher, Chaurasia, & Qiu, 2023). For training, 385 annotated images (resolutions of 1024x768 px and 1280x1024 px) from vehicle-mounted camera footage recorded around Trento, Italy were supplemented with 125 images from an existing crack dataset, recorded by handheld GuPho (Padkan et al., 2023). Both sets were annotated on the Roboflow web platform (Dwyer et al., 2024). To enhance training performance, data augmentation techniques such as random rotations, exposure adjustments and random noise addition were utilized. The trained models are tested on 87 manually annotated images of roads around Trento.

Bird's eye view generation. The estimated trajectory is used to parametrically generate a mesh of the road. The motivation for using a parametric mesh instead of a traditional meshing approach - which relies on sparse point clouds or depth maps - stems from the need for a smooth surface. Traditional meshing methods tend to incorporate tie point noise, resulting in a rough surface that is unsuitable for further processing. Moreover, generating a mesh from such input data typically demands significant processing time. Generating a mesh from the dense point cloud is equally difficult due to the poor texture of the road surface. For each camera pose in the trajectory, two collinear vertices are created at a fixed width from the camera in the

perpendicular direction to the tangent of the trajectory. The camera-to-ground height is used as vertical offset to place the vertices at the correct height on ground level. By using the camera rotation parameters, the vertices are adjusted to account for the longitudinal and transverse slopes of the road. Two triangular faces are created by connecting pairs of vertices for two consecutive camera poses, resulting in a 3D surface modelling the road. Next, a UV mapping (3D to 2D) is computed by automatically unwrapping the mesh onto a plane and assigning each mesh vertex a coordinate in the UV space. The texture for the mesh is created by projecting each detected crack mask onto the corresponding faces of the mesh with a perspective transform (homography). No RGB blending between consecutive faces is utilized. The texture is applied to the mesh with the computed UV mapping and, finally, a 1cm/px orthomosaic of the cracks is created by orthorectifying the textured mesh. Additionally, a basemap of the road can be created by repeating the texture creation and orthomosaic generation with the original, undistorted frames. The generated bird's eye view offers a nadiral view of the road, enabling the mapping of the reconstructed 3D road surface onto a 2D plane and its use in GIS.

Cracks size estimation in GIS. The georeferenced orthomosaic is imported into a GIS environment. The image is split into smaller tiles to optimize memory usage. Detected cracks are vectorized into polygons and standard topology algorithms are applied to the polygons to fix invalid geometries. Cracks across different tiles are merged. In literature, various definitions of cracks sizes exist. For example, crack length may be determined by measuring the total perimeter, including branches, the longest continuous segment or the projected length along the longitudinal and transverse axes. Operators can apply their own definition and measure length directly in GIS using available tools. For the sake of following tests, length is estimated as the height of the oriented minimum bounding box around the crack.

4. RESULTS AND DISCUSSION

The pipeline is tested on a ~3km long sequence over a secondary road open to traffic in Trento, Italy. The sequence includes turns, hairpins, roundabouts and straight sections. The vehicle is kept at a constant speed of ~50km/h and at an adequate distance from preceding vehicles to have a clear view of the road. Images are acquired at 4 frames per second. A total of 8 natural points (pedestrian crossings, corners, manholes) well distributed along the sequence were acquired by means of a Real Time Kinematics (RTK) enabled GNSS receiver, with estimated accuracies in the centimetres range. These points are used as bundle adjustment constraints, e.g. Ground Control Points (GCPs) for recovering a georeferenced trajectory (UTM 32N coordinate reference system). Incoming traffic is successfully masked out by the road masks (Figure 2).



Figure 2: Road segmentation results (purple mask) to exclude incoming traffic and objects outside the road.

The use of masks prevents the mapping of cracks detected on objects outside of the road. The resulting parametric mesh of the road, computed from the estimated vehicle trajectory, is shown in Figure 3. The mesh aligns with the camera positions and rotations, ensuring that the normal vector of each corresponding

face remains perpendicular to the camera's forward axis throughout the trajectory. As a result, errors can occur during the orthorectification of the mesh. Inaccurate camera pose estimations, particularly in orientation, may lead to incorrect projections of detected cracks and potential distortions in their shapes.

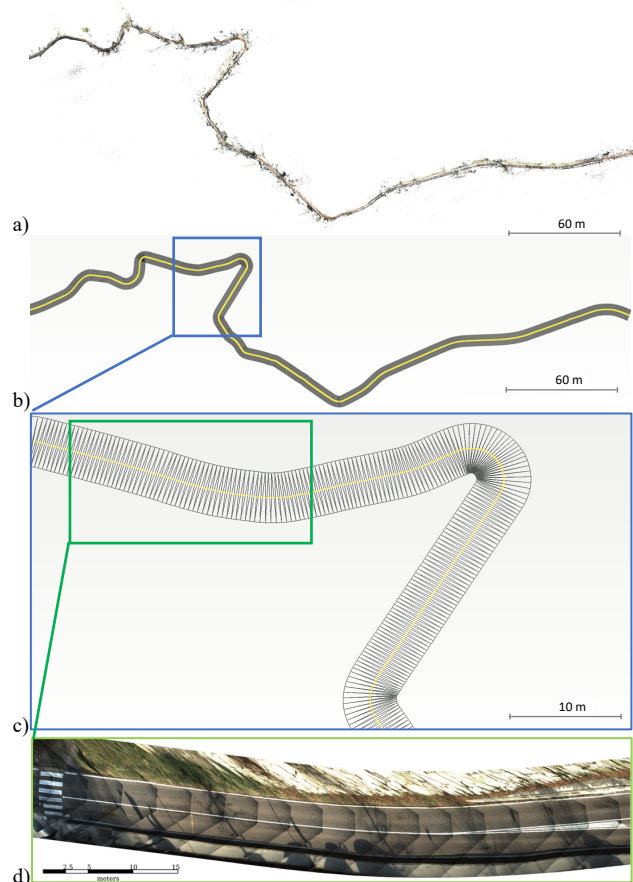


Figure 3: SLAM-based sparse point cloud and camera trajectory of the surveyed road (a). Parametric mesh (grey) and trajectory (yellow) (b) with a zoom-in to show faces with wireframe visualization (c). The mesh is shown in grey while the trajectory is in yellow. Highlight of the bird's eye view created by texturing the mesh (c). Notice that no blending is utilized between frames.

In case of sharp turns (90 degrees or U turns) the faces of the mesh may overlap, rendering the projection of the cracks discontinuous. A geometry verification and fixing algorithm which approximates the turn with the osculating circle is implemented to unravel the mesh when this happens.

To evaluate the accuracy of the proposed procedure, ground truth data were collected through on-site measurements of identifiable features such as manholes, pedestrian crossings and crack lengths. While known features are relatively straightforward to identify and measure, it is important to note that on-site measurement of crack sizes is inherently subjective due to unclear crack boundaries. These measurements are also replicated in GIS, using the orthomosaic of the road for known features and vector shapes for cracks. Some examples are reported in Figure 4. Results show that the mapping process achieves accuracies in the centimeter range, which is acceptable given the measurement uncertainties, including systematic and random errors on field and pixel localization inaccuracies on the orthomosaics. The measurement depends on the estimated

camera-to-ground height used in the mesh creation step: a slight variation from the actual value would cause the pixels of the images to be projected on a closer or further away plane, hence shortening or lengthening the size of objects.

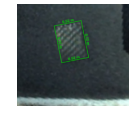
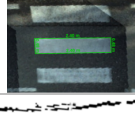
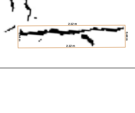
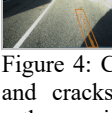
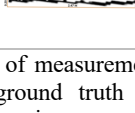
Known features and cracks			Ground truth (m)	Measurement on map (m)
			0.75	0.72
			2.50 x 0.50	2.40 x 0.48
			2.14	2.12
			2.54	2.47

Figure 4: Comparisons of measurements of known features and cracks between ground truth and on the resulting orthomosaics. The mapping process is scaled in the centimeters range.

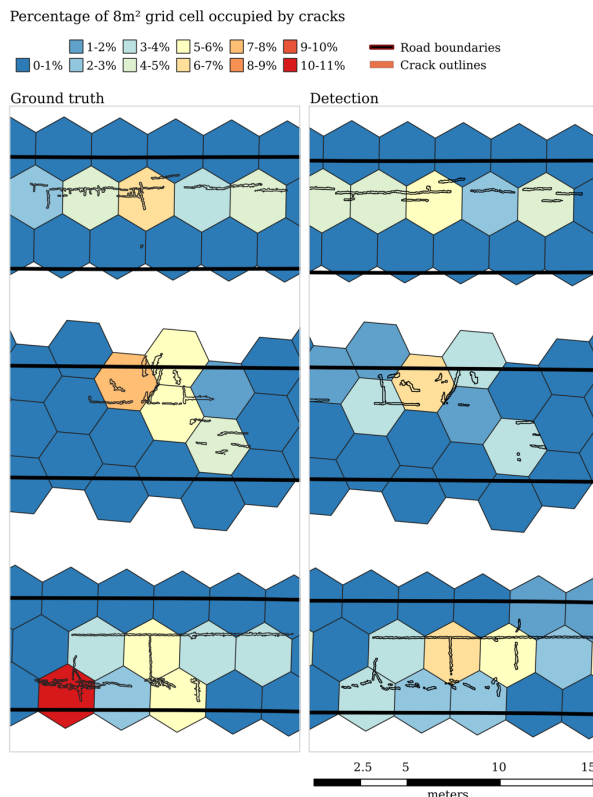


Figure 5: Percentage area of each cell of the grid occupied by cracks in three parts of the study area. Comparison between manually annotated cracks (ground truth) and YOLOv8m detected cracks. The detection achieves comparable results to the ground truth, although under detection is sometimes visible.

This explains the variation in measurement between the length and the width of the pedestrian crossing in Figure 4, as the width is not affected by the inaccuracy of the camera-to-ground height parameter. This systematic error is present in other pedestrian

crossing lengths measured on the orthomosaic, while the error on the width was consistent with the 1cm GSD of the orthomosaic. We evaluate the performance of the DL detection models by looking at the completeness of the detection over different short sub-sequences of 3 to 5 frames which are not utilized as training data. The chosen sequences are spread out along the surveyed road and exhibit different characteristics. Cracks may be distributed on the incoming lane only, over the entire surface of the road or in the centre line or may be missing at all. We manually annotate cracks in the sub-sequences and use the resulting masks to map them. Then, we subdivide the mapped road surface with a grid of 8m² hexagonal cells. For a 5 to 6m wide road as in the case of our dataset, 3 grid cells cover the whole width of the road. An overlap analysis is carried out by overlaying the grid cells and the vectors of the cracks, and by computing the percentage of the area of each cell occupied by cracks. We motivate the choice of this metric by noticing that the detection models fail to isolate single cracks when multiple cracks are present, merging the result in a single. Individual crack parameters such as length and width, which may have practical implications for maintenance, hence lose significance. The process is repeated by inputting the masks of the detected cracks. The results of the comparison for the YOLOv8 model are shown in Figure 5, while Table 1 reports statistics about the occupancy rates of grid cells for all models across the three sub-sequences. Generally, the models achieve lower occupancy percentages than the ground truth because of under-detection, especially for multi-branch, widespread cracks, as visible in the last row of Figure 5. U-net, on the other hand, tends to overdetect cracks, resulting in higher maximum and average occupancy rates. Although none of the models are fully robust in identifying all cracks, their detection results occupy a similar number of cells as the ground truth, indicating that the road surface is well represented.

A new point layer is created by assigning each cell centroid the area occupancy value. An interpolation is done on the points to create the heatmaps reported in Figure 6. These maps are useful tools to quickly evaluate which areas of the road are undergoing the highest deterioration and would require intervention.

Subsequence #1 -17 cells				
	Ground truth	YOLO v8m	DeepLab V3	U-net
Min %	0.1	0.9	0.1	1.1
Max %	6.2	5.9	5.3	11.1
Average %	2.7	3.5	3.1	5.7
Numb. of occupied cells	8	7	8	7
Subsequence #2 -18 cells				
	Ground truth	YOLO v8m	DeepLab V3	U-net
Min %	0.3	1.2	0.1	1.0
Max %	7.4	6.8	6.7	11.5
Average %	3.7	3.0	1.9	4.3
Numb. of occupied cells	7	7	6	7
Subsequence #3 -17 cells				
	Ground truth	YOLO v8m	DeepLab V3	U-net
Min %	0.3	0.1	0.1	0.1
Max %	10.2	6.3	6.0	8.8
Average %	3.6	3.0	3.1	3.3
Numb. of occupied cells	10	10	9	12

Table 1: Statistics about the occupancy rates for grid cells occupied by detected cracks in three subsequences of the surveyed area between ground truth annotations and the results of the detection models.

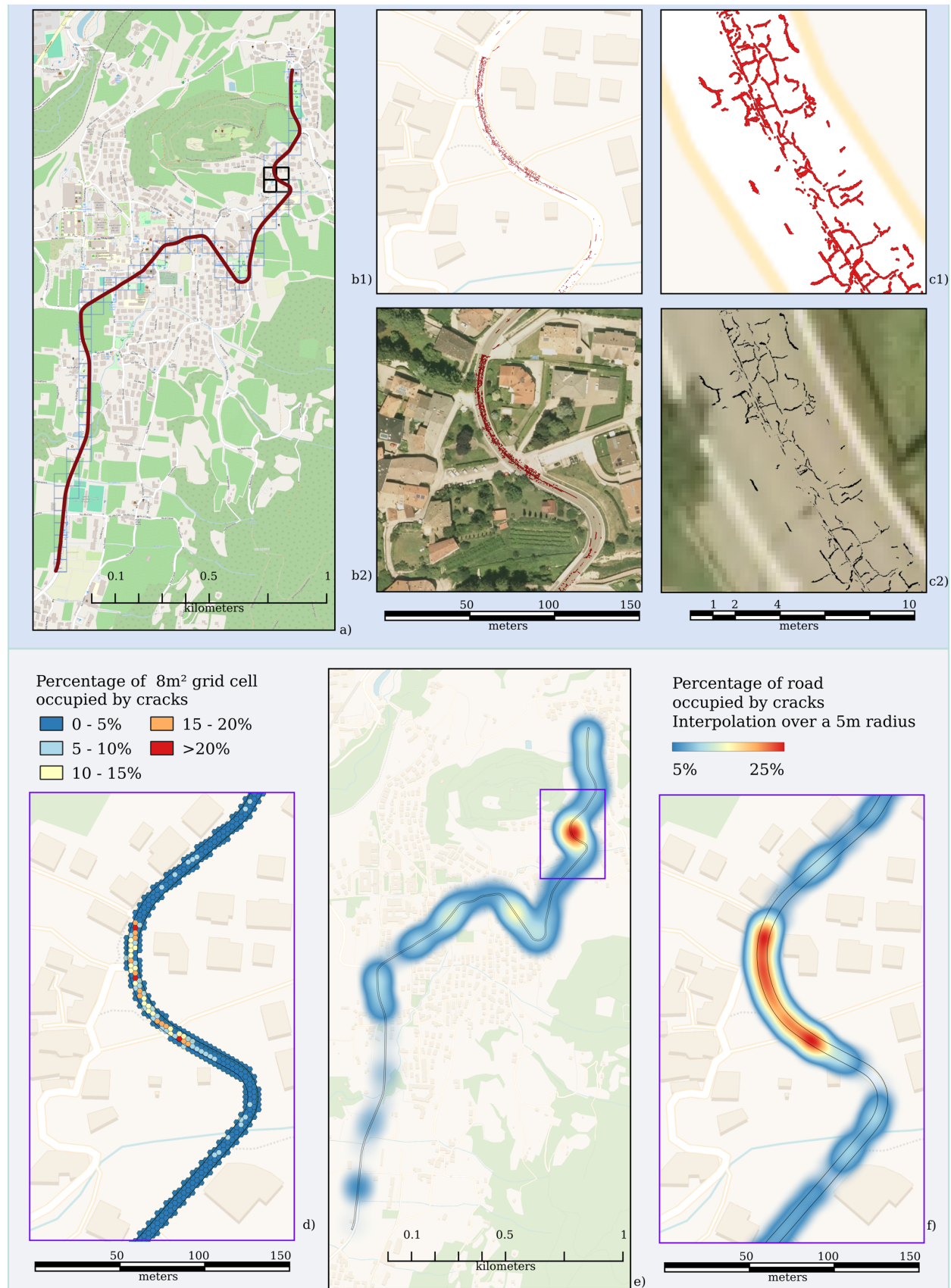


Figure 6: Survey trajectory (dark red) and orthomosaic tiles boundaries (light blue/black) overlaid onto OpenStreetMap - OSM (a). Cracks detected with YOLOv8m overlaid onto OSM as vector shapes (b1, c1) and onto a 20cm orthophoto (b2, c2). The surface of the road is divided into a grid of 8m² hexagonal cells and the percentage of the area occupied by cracks is computed for each cell (d). Interpolation of the occupancy rate in a radius of 5m at two different scales (e, f).

4.1 Model Evaluation

The performances of the three models, evaluated using Precision(P), Recall (R), F1-score (F1), and inference time, are reported in Table 2. High Recall means a high number of actual crack pixels detected, high Precision indicates higher certainty in the detections whereas the F1-score combines Precision and Recall into a single metric, balancing the trade-off between detecting all cracks and ensuring the detections are accurate. Inference time is an important metric in case of large datasets or for real time capabilities. All models were tested for 250 epochs on an NVIDIA GeForce RTX 4070 with 16GB VRAM.

The models were trained without any modifications to their original layer structures. U-net and DeepLabv3 were trained and tested on 256x256 px patches. Different confidence levels were tested to evaluate the models: 0.25 was used as minimum threshold instead of the conventional 0.5. This could increase the number of false positives but maximizes the Recall, leading to more cracks detected.

Model	P (%)	R (%)	F1 (%)	Inference time (s)
U-net	54	86	63	2.6
DeepLabV3	60	83	66	4
YOLOv8m	63	52	57	0.016

Table 2: Evaluation metrics on 87 manually annotated images of roads around Trento. Highest values are in bold.

As seen in Table 2, DeeplabV3 generally outperforms other models achieving the highest F1-score, while U-net returned the highest recall. U-net and DeeLabv3 achieved comparable metrics, while YOLOv8m showed lower recall. An explanation for this is provided in Figure 7 where a complex, multi-branch ‘alligator’ crack is only partially detected by YOLOv8m. This example is also a clear demonstration of how, as previously reported, U-net overdetects cracks, enlarging their actual dimensions. Despite its lower performance, YOLOv8m is the only model demonstrating real-time capabilities with milliseconds per image inference time.

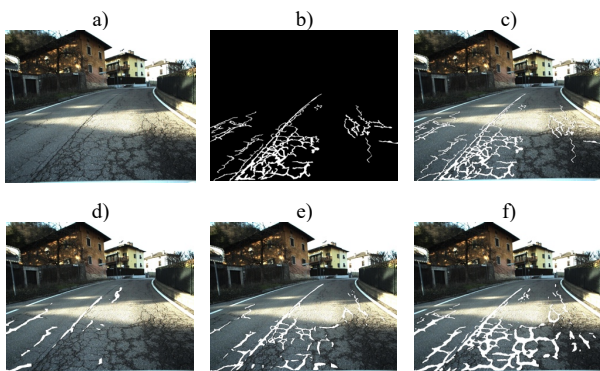


Figure 7: An image with a complex alligator crack. Original image (a), ground truth mask (b), ground truth overlay (c) and detection results using YOLOv8m (d), DeepLabV3 (e), and U-net (f).

In general, the models’ performances are imperfect, with challenges in the detection given by the presence of shadows which have a similar dark and thin appearance as cracks. The models suffer specifically in detecting far cracks that appear blurry, as shown in Figure 8. This problem is partially solved with the detection in following frames. In addition, cracks are sometimes undetected or detected partially probably because of

the limited image resolution, the acquisition perspective and the noise of the compression that greatly influence the detection. Adding more images to the dataset improved the models’ performance. However, more images increase the chance of wrong annotations, given the subjective task and the fact that it is not clear to the human eye where a crack starts and finishes. This was evident when on-site measurements of cracks were performed, since cracks features may become visible only under certain light conditions, scales or perspectives.

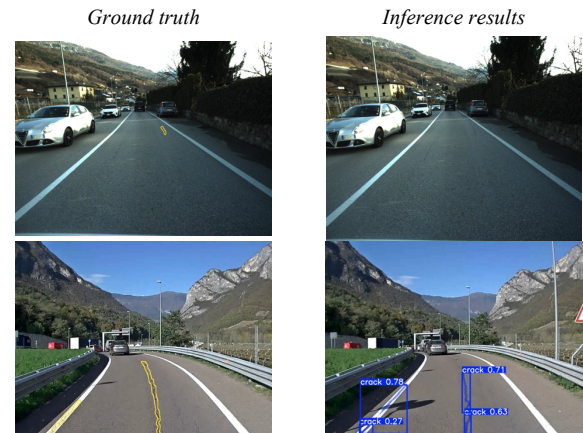


Figure 8: Examples of YOLOv8m inaccuracies: a distant crack entirely missed (top) and two vertical cracks partially detected and incorrectly split into separate cracks (bottom).

5. CONCLUSIONS

The paper introduced an end-to-end pipeline to effectively extract and map road surface cracks to assess road conditions from vehicle-mounted camera footage. The pipeline combines SLAM processing and learning detection to support georeferencing and metric results visualized in a GIS environment. In general, Precision and Recall values are greatly affected by the quality of images and available training data. Although values read lower than those of human operators, the methodology offers significant benefits in terms of efficiency. While achieving the same level of accuracy as an operator may not be feasible, the pipeline delivers rapid preliminary results, such as crack locations and distribution. These results can be reviewed by operators for validation with significantly reduced time, allowing them to focus on verification rather than the initial detection process, with considerable time savings and overall operational efficiency. A key advantage of the proposed pipeline is its integration with GIS environments, which enhances data management and provides an intuitive platform for further analysis. By mapping detected cracks, operators can validate results more efficiently but also leverage advanced spatial analysis tools to identify high-risk areas and monitor damage progression. This approach reduces the subjectivity and labor-intensive aspects of traditional inspections while enabling more strategic planning for road maintenance. By utilizing layered visualizations that incorporate satellite imagery and topographic details, operators obtain a comprehensive view of pavement distress, facilitating clear assessments of damage distribution. Additionally, by conducting repeated surveys and comparing results over time, operators can track crack progression, assess deterioration trends, and make informed predictions about future pavement conditions. The ability to generate professional, stakeholder-friendly maps further enhances communication, ensuring that repair priorities and inspection outcomes are

effectively conveyed, ultimately contributing to cost savings and a prolonged lifespan for road infrastructure. Future work includes enhancing the cracks detection with other approaches, incorporating additional training data or exploiting Multimodal Large Language Model (MLLM). Moreover, GNSS data will be acquired synchronously with the camera footage to directly georeference the results and aid the reconstruction process.

REFERENCES

- Alipour, M., Harris, D.K. and Miller, G.R., 2019. Robust pixel-level crack detection using deep fully convolutional neural networks. *J. Computing in Civil Engineering*, 33(6), p.04019040.
- Baduge, S.K., Thilakarathna, S., Perera, J.S., Ruwanpathirana, G.P., Doyle, L., Duckett, M., Lee, J., Saenda, J. and Mendis, P., 2023. Assessment of crack severity of asphalt pavements using deep learning algorithms and geospatial system. *Construction and Building Materials*, 401, p.132684.
- Barbieri, D.M., Lou, B., 2024. Instrumentation and testing for road condition monitoring – A state-of-the-art review. *NDT & E International*, Vol. 146, 103161.
- Chen, L.C., Papandreou, G., Schroff, F. and Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.
- Choi, Y., Bae, B., Han, T.H. and Ahn, J., 2024. Application of Mask R-CNN and YOLOv8 Algorithms for concrete crack detection. *IEEE Access*.
- Chun, P.J., Yamane, T. and Tsuzuki, Y., 2021. Automatic detection of cracks in asphalt pavement using deep learning to overcome weaknesses in images and GIS visualization. *Applied Sciences*, 11(3), p.892.
- Dwyer, B., Nelson, J., Hansen, T., et al., 2024. *Roboflow* (Version 1.0).
- Golding, V.P., Gharineiat, Z., Munawar, H.S. and Ullah, F., 2022. Crack detection in concrete structures using deep learning. *Sustainability*, 14(13), p.8117.
- Ha, J., Kim, D., Kim, M., 2022. Assessing severity of road cracks using deep learning-based segmentation and detection. *J. Supercomputing*, Vol. 78, 17721-17735.
- Hsieh, Y.A. and Tsai, Y.J., 2020. Machine learning for crack detection: Review and model performance comparison. *Journal of Computing in Civil Engineering*, 34(5), p.04020038.
- Jing, P., Yu, H., Hua, Z., Xie, S., Song, C., 2023. Road crack detection using deep neural network based on attention mechanism and residual structure. *IEEE Access*, Bol. 11, pp. 919-929.
- Jocher, G., Chaurasia, A. and Qiu, J., 2023. *YOLO by Ultralytics* (Version 8.0.0).
- Jung, W.M., Naveed, F., Hu, B., Wang, J. and Li, N., 2019. Exploitation of deep learning in the automatic detection of cracks on paved roads. *Geomatica*, 73(2), pp.29-44.
- Li, H., Liu, Y., Fan, L. and Chen, X., 2017, December. Towards robust and optimal image stitching for pavement crack inspection and mapping. *Proc. International Conference on Robotics and Biomimetics (ROBIO)*, pp. 2390-2394.
- Lin, C.J., Garg, S., Chin, T.J., Dayoub, F., 2025. Robust Scene Change Detection Using Visual Foundation Models and Cross-Attention Mechanisms. *Proc. ICRA*.
- Malone Geary, G. and Tsai, Y., 2021. Jointed plain concrete (JPC) pavement variability and method to complement JPC design with 3D pavement data. *Transportation Research Record*, 2675(8), pp.332-344.
- Morelli, L., Ioli, F., Beber, R., Menna, F., Remondino, F. and Vitti, A., 2023. COLMAP-SLAM: A framework for visual odometry. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, pp.317-324.
- Munawar, H.S., Hammad, A.W., Haddad, A., Soares, C.A.P. and Waller, S.T., 2021. Image-based crack detection methods: A review. *Infrastructures*, 6(8), p.115.
- Opara, J.N., Thein, A.B.B., Izumi, S., Yasuhara, H. and Chun, P.J., 2021. Defect detection on asphalt pavement by deep learning. *GEOMATE Journal*, 21(83), pp.87-94.
- Padkan, N., Battisti, R., Menna, F. and Remondino, F., 2023. Deep learning to support 3D mapping capabilities of a portable vslam-based system. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48(1), pp.363-370.
- Ranyal, E., Sadhu, A. and Jain, K., 2024. Enhancing pavement health assessment: An attention-based approach for accurate crack detection, measurement, and mapping. *Expert Systems with Applications*, 247, p.123314.
- Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L. and Mintun, E., 2024. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*.
- Rezaie, A., Achanta, R., Godio, M. and Beyer, K., 2020. Comparison of crack segmentation using digital image correlation measurements and deep learning. *Construction and Building Materials*, 261, p.120474.
- Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. *Proc. 18th MICCAI*, pp. 234-241.
- Stricker, R., Aganian, D., Sesselmann, M., Seichter, D., Engelhardt, M., Spielhofer, R., Hahn, M., Hautz, A., Debes, K. and Gross, H.M., 2021. Road surface segmentation-pixel-perfect distress and object detection for road assessment. *Proc. 17th International Conference on Automation Science and Engineering (CASE)*, pp. 1789-1796.
- Torresani, A., Menna, F., Battisti, R. and Remondino, F., 2021. A V-SLAM guided and portable system for photogrammetric applications. *Remote Sensing*, 13(12), p.2351.

Wang, X., Gao, H., Jia, Z. and Li, Z., 2023. BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors*, 23(20), p.8361

Wu, H., Kong, L. and Liu, D., 2024. Crack detection on road surfaces based on improved YOLOv8. *IEEE Access*.

Wu, Y., Han, Q., Jin, Q., Li, J. and Zhang, Y., 2023. LCA-YOLOv8-Seg: an improved lightweight YOLOv8-Seg for real-time pixel-level crack detection of dams and bridges. *Applied Sciences*, 13(19), p.10583.

Wu, Y., Li, S., Li, J., Yu, Y., Li, J., Li, Y., 2025. Deep learning in crack detection: A comprehensive scientometric review. *J. of Infrastructure Intelligence and Resilience*, Vol. 4(3), 100144.

Zhanabatyrova, A., Leite, C., Xiao, X., 2023. Automatic map update using dashcam videos. *Internet of Things*, Vol. 10(13).

Zhang, H., Venkatramani, S., Paz, D., Li, Q., Xiang, H., Christensen, H.I., 2023. Probabilistic Semantic Mapping for Autonomous Driving in Urban Environments. *Sensors*, 23, 6504.

Zhang, A., Wang, K.C., Fei, Y., Liu, Y., Tao, S., Chen, C., Li, J.Q. and Li, B., 2018. Deep learning-based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet. *Journal of Computing in Civil Engineering*, 32(5), p.04018041.