Performance Evaluation of a Visual Defects Detection System for Railways Monitoring

Saša Radosavljevic, Alain Rivero, Sergio Rodríguez Flórez, Abdelhafid Elouardi, Pauline Michel, Belkacem O Bouamama, Philippe Vanheeghe

Performance Evaluation of a Visual Defects Detection System for Railways Monitoring

Saša Radosavljevic[1,2,*], *Alain* Rivero[1,**], *Sergio* Rodríguez Flórez[2], *Abdelhafid* El Ouardi[2], *Pauline* Michel[2], *Belkacem* O. Bouamama[3], and *Philippe* Vanheeghe[3]

[1]SNCF Réseau
[2]Université Paris-Saclay, ENS Paris-Saclay, CNRS, SATIE, 91190, Gif-sur-Yvette, France
[3]Université de Lille, CNRS, UMR 9189 - CRIStAL, 59000 Lille, France

Abstract. SNCF Réseau introduces a novel multi-modal embedded monitoring system, addressing challenges in railway infrastructure maintenance. The design incorporates visual, inertial, and sound sensors, enhancing adaptability, improving overall detection precision, and could reduce operational costs. This study addresses visual defects detection that can be integrated in a multi-modal monitoring system. The paper details the system's architecture, synchronisation methods, and decision fusion process to improve the precision of limited mono-modal systems. A deep-learning visual based railway defects inspection was explored. Results show that small CNN (Yolov8 nano) can achieve similar (Yolov8 XL) high precision (mAP@0.5 $\geq$ 0.89) for a small number of objects (9) while improving implementation capability on embedded systems.

1 Introduction

Recent ecological concerns have spurred a renewed interest in railways due to their lower carbon footprint. However, disuse threatens the very existence of smaller, less profitable lines, leading to their gradual disappearance. This trend coincides with France's goal of opening its railway infrastructure to competition while facing a concerning decline in its overall condition. Effective railway infrastructure maintenance needs comprehensive and meticulous monitoring of all integrated components.

To navigate these challenges, one solution involves optimising infrastructure maintenance and operational costs [1]. In response, SNCF Réseau, the French Infrastructure Manager (IM), is developing a novel track monitoring system. This portable, cost-effective, and efficient system offers a comprehensive solution, integrating visual inspections with advanced techniques for in-rail defect detection and environmental perception and analysis. Its seamless installation on various train types further enhances its practicality.

Existing infrastructure monitoring systems exhibit limitations in their adaptability to diverse railway lines [2]. Some systems are well-suited under low-speed motion conditions, while others demand a highly controlled environment [3]. Additionally, the most effective monitoring systems often come with prohibitively high operational and acquisition costs, rendering them impractical for deployment on smaller rail lines [2].

This paper proposes a multi-modal system-based monitoring design considering the automation of visual inspection in railway infrastructure. Traditional monitoring systems relying on visual inspection often lack automation processes[2], leading to the accumulation of large datasets that pose challenges for analysis. Even with an initial processing method, a significant number of false positives are generated. Remarkably, less than 1% of the images presented to a human operator exhibit actual defects.

Although 38% of the lines carry 80% of the train traffic, small lines constitute 29% of the French Railway Network (RFN) [1]. Additionally, 20% of the smallest lines account for only 1% of the overall traffic. The total annual maintenance cost across all lines amounts to approximately €2.7 billion [1]. The proposed system integrates cutting-edge analysis techniques, such as neural networks, data fusion, and decision models. This integration is oriented to achieve a substantial economic gain through innovative monitoring processes, which will lead to a reduction in surveillance, acquisition, and operational maintenance costs. Consequently, it contributes to an overall decrease in infrastructure maintenance expenses. Using commercial trains for monitoring infrastructure status, as opposed to employing expensive dedicated systems, makes this new monitoring system particularly suitable for secondary railway lines. Moreover, the system should be independent from the train's electrical circuit or drain as little as possible as trains are not designed to provide energy to external systems. The isolation of such systems explains the frugality and compactness requirements to be met.

In summary, the main contributions of this work include:

- a comprehensive analysis of railway monitoring and limitations on detecting defects infrastructure problems,

*e-mail: sasa.radosavljevic@ens-paris-saclay.fr
**e-mail: alain.rivero@reseau.sncf.fr

- a conceptual description and study of an embedded system-based monitoring design, and
- a performance impact study of a deep-learning visual-based railway defects inspection.

The remainder of this paper is structured as follows: Section 2 discusses railway monitoring, followed by Section 3 which explores embedded system-based monitoring. Section 4 presents the visual defect detection part, and finally, Section 5 concludes the paper with insights drawn from the research.

2 Railway monitoring

Railway monitoring non-destructive methods are various but few can be exploited without meticulous configuration and on flexible support[3]. Moreover, very few studies in railways are made on combining multiple sensors to have a complete monitoring system[4] or embedding detection systems[3] on various trains contrary to dedicated monitoring non-passenger trains[2]. Naturally, combining multiple sensors would lead to a very versatile main sensor coupled with more precise detection to compensate for its flaws [4]. The most versatile sensor is a camera[5]. Studies on object detection often start with a primary focus on its accuracy and later comes its computing cost and embedded limitations [4][6][7].

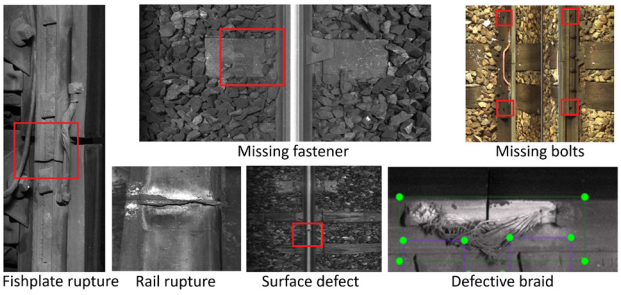

Figure 1. Defect types examples, SNCF Réseau. From left to right, defects such as ruptures, missing fasteners and bolts, surface defect examples, and a context element.

Deeper studies should be made about the reliability of these systems integrating existing inspection systems. Besides, embedding these computing-heavy systems means that good precision often comes with large hardware architectures and energy consumption [4][7]. Installing this (flexible embedded) system on commercial trains will allow us to reduce main-line maintenance costs and enable more time slots for them. Improvements to this system will, over time, enable its installation on high-speed line (HSL) trains.

Artificial intelligence (AI) facilitates the automated and autonomous analysis of defects in railway infrastructure [8], eliminating the need for pre-defined defect models. This approach offers several advantages, including scalability, robustness, and reliability, making it well-suited to the economic constraints of railway operations [9].

In France, heavily trafficked rail lines exhibit an annual increase in track defects. These defects manifest in various types, shapes, and sizes [10]. Notably, once defined, they may display recognisable patterns, as illustrated in Fig. 1, with some exceptions for surface defects that may vary in severity [11]. These patterns can be identified and adapted to with the use of Convolutionnal Neural Networks (CNN) that use filters as pattern recognition, emulating human vision[12].

3 Embedded Multi-modal Monitoring System

The concept underlying the proposed monitoring system involves employing a primary polyvalent sensor, in this case, a camera, designed to ideally detect multiple types of defects. However, the efficacy of the camera-based detection algorithm may be compromised even if the camera captures the defect due to a lack of physicality. Indeed, leaves or grease marks due to their imprint may lead the system to detect surface defects as illustrated in Fig. 1. Historically, researchers have drawn inspiration from nature to enhance the performance of their systems [12]. Consequently, the augmentation of extrasensory organs [3] in our system is proposed to enhance the functional spectrum of its defect recognition capabilities, denoted as a multi-modal system. In addition to cameras, the inclusion of inertial sensors [13] and microphones aims to augment the information available to the system. This augmentation facilitates the ability to either eliminate false detections, such as stains or incorporate physical contextual attributes into the system, for instance, sensing wheel motion.

The first functional block in the implementation concerns the **visual-based defect detection** module (Fig. 2). This block provides versatile and informative data, similar to the eyes in a biological system. The multimodality of the systems is defined by its multiple **sensors synchronised** and framed using **odometric** information to ensure **data processing** coherence and matching frames during **data fusion**.

3.1 Camera

This module aims to minimise the occurrence of false negatives (missed detection) in railway defect identification, thereby mitigating the risk of overlooking critical features on the tracks and enhancing overall data reliability.

3.1.1 Technical specification

The requirements for the cameras include their acquisition capacity and a few geometric considerations, as listed hereafter:

- Length of the smallest defect to be detected: 2 mm
- Minimum projected pixel: 1 mm (Shannon criteria)
- Minimum field-of-view: 300 mm
- Maximum scanning speed: 160 km/h, 44,445 mm/s
- Working distance from rails (y_{cam}, Fig. 3): $\approx$ 60 cm

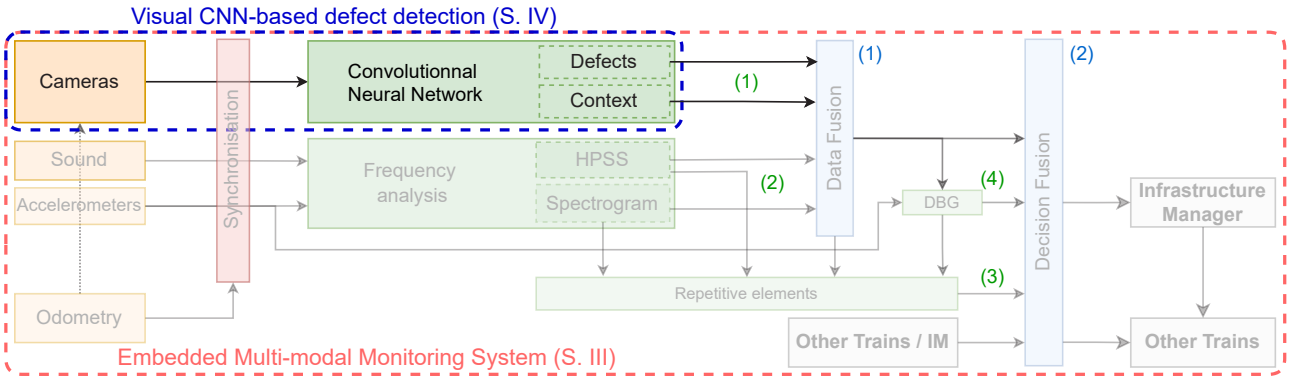

Figure 2. Embedded multi-modal monitoring system architecture. Sensors (orange), synchronization (red), data processing blocks (green), and data fusion stages (blue) aggregate information to improve precision. Final results are sent to IM and relevant trains.

The ability to use this system on commercial trains requires a scanning speed equivalent to the train's maximum speed. The working distance is both chosen according to the rail vehicle coupler's height and the minimum projected pixel size that impacts the lens. Finally, the camera is chosen according to the maximum scanning speed that can be matched by the camera's frequency. For this task, 4 cameras have been placed, one on each side of the 2 rails. It will enable us to see webs and their elements, running surfaces, and fasteners. Their positioning can be visualised in Fig. 3. To ensure optimal image quality, linear cameras were chosen [6], which exhibit less lighting complexity at high speeds and no vertical distortion, given that the images are recomposed from a single or few lines.

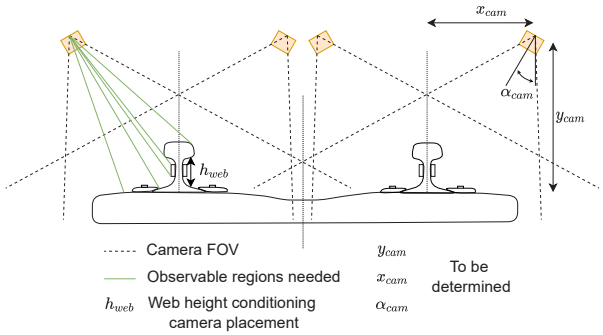

Figure 3. Future cameras positioning to capture details needed for detection. Regions of interest are identified with green lines.

3.1.2 Exposure

Such a high scanning speed needs meticulous attention to lighting. Firstly, the exposure should be as short as possible to prevent stretching. The movement during exposure time should be less than half of the projected pixel size, which is less than 500 µm, resulting in an exposure time of 10 µs. Two options are available: either optimising power consumption by utilising stroboscopic light synchronised with camera acquisition or simplifying the process with constant lighting. Illuminating the scanned surface helps mitigate the impact of varying histograms (e.g., tunnels, high exposure to sun, night...), thereby reducing complexity in the CNN architecture when aiming to enhance performance. Once again, the goal of the system is to balance every element to maximise detection while reducing computing requirements and power consumption then tuning correctly the overall system is the key.
Thus, the first experiments will use a continuous light for its convenience with a long term goal of reducing power consumption with stroboscopic light.

3.2 Synchronisation

The synchronisation component of the system (Fig. 2) serves a dual purpose. It controls the camera to trigger every millimeter and synchronises other sensors with the camera acquisition, allowing the system to associate temporal windows with the images processed during CNN detection.

3.2.1 Odometry

Measuring the train speed is crucial for both labeling data and defects with the current location and facilitating the synchronisation of camera acquisition, ensuring a line acquisition is triggered every millimeter of forward movement. However, utilising an encoder on the train's wheel presents challenges due to standard Counts Per Revolution (CPR), making it difficult to achieve one count per millimeter (see Fig. 4) requiring signal interpolation in-between. The installation on a train could also necessitate safety regulation studies.

Wheelslipage may lead to failure triggering the camera. To address this issue, a precise solution involves sensor fusion, incorporating data from sources like encoders, GPS, and train speed information. Employing a Kalman filter enhances accuracy, and the gathered information can also be utilised for repetitive pattern recognition in accelerometry.

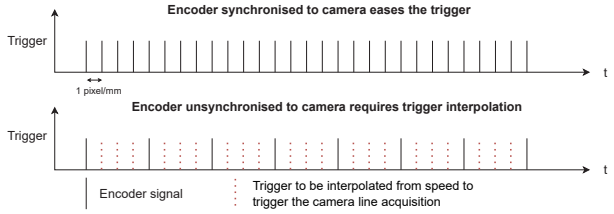

Figure 4. Encoder should fit the desired trigger distance spacing. A lesser encoder precision leads to a gap in movement information and needs to be filled with an interpolated trigger.

3.2.2 Triggers

Using the speed evolution, we can interpolate the motion and activate the camera line acquisition (Fig. 4). This signal allows us to segment frames from each sensor, facilitating the correlation of information for data fusion (Fig. 5).

Since accelerometers are motion-dependent, the system must travel $\Delta Ticks = d_{cam-acc}$. On the other hand, the microphones are travel-time-dependent, with $\Delta t = \frac{d_{acc-mic}}{v_{sound}}$.

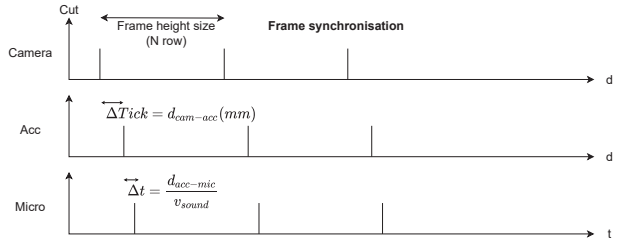

Figure 5. Data space-framing ensuring data fusion coherence. A delay appears between sensors and can be managed with encoder triggers and a small delay.

3.3 Processing

At this stage, using neural networks is beneficial since they help reduce the need for detecting specific features, allowing for a broader range of defects to be identified [6].

3.3.1 Pre-processing

Focused on the railroad track and its surroundings, our system aims to enhance efficiency. Strategies such as trimming edges and exploring techniques like image resizing, down-scaling, and histogram equalisation offer potential for more effective data processing. These approaches efficiently handle processing loads, improve accuracy, and address changes in lighting conditions.

3.3.2 Frequency analysis

As previously mentioned, some objects may be misinterpreted by image detection. To mitigate this, accelerometric and sound[14] data complement the detection, aiding in discriminating surface defects and providing additional information on defect types. These sensors also facilitate wheel defect detection by analysing repetitive elements, distinguished from standard railway joints with fixed spacing (e.g., 12, 18, or 32 meters). The system incorporates a dedicated component for repetitive object detection, contributing to decision fusion.

3.3.3 Diagnostic Bond Graph (DBG)

To add a verification process to our system, we used a model based Fault Detection and Isolation of the rail/wheel contact and the impact of the entire body on acceleration. For this task, Bond Graph theory - well suited for mechatronic systems - is used here not only for modeling but also for systematic generation of residuals (as fault indicators) for robust alarm generation and fault isolation based on fault signature matrix reasoning. The BG model is used in derivative causality while initial conditions in real systems are not known[15] (Fig. 6). When the real system encounters a specific defect and classifies it, the DBG model is activated to compare the modeled response and the real accelerometric response from the train. This comparison results in a residual signal which, when combined with a machine-learning classifying method, effectively confirms or denies the presence of the detected defect. To improve the system's recall, the DBG should also be capable of detecting an abnormal situation from sensor data alone.

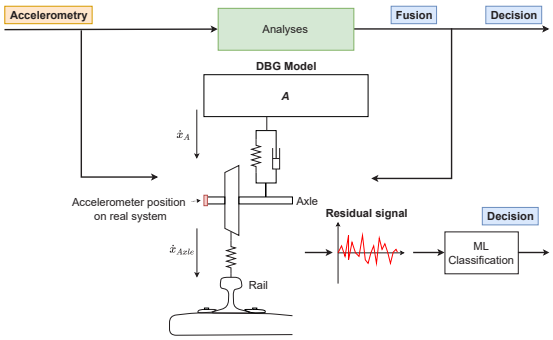

Figure 6. Decision expert - Diagnostic ML-Bond Graph. Accelerometry input is compared to modeled detected defect leading to residual signal classified by ML.

3.4 Fusion

To integrate each component of the system, it is imperative to consider parameters derived from various processing stages. For instance, image processing yields a confidence score based on Intersection over Union (IoU), while frequency analysis provides a correlation value relative to established defect signatures. The augmentation of information contributes to the enhancement of the detection system. A coherent temporal alignment of these diverse inputs allows for the effective detection of defects.

3.4.1 Data fusion

Data fusion [16] serves as a means to enhance the initial processing chain. Specifically, defects detectable solely through visual means may not be ameliorated by additional sensors. However, as illustrated in Table 3 and Fig. 7, the most challenging detection concerns surface defects. Interestingly, surface defects can be discerned at the rail/wheel contact point, influencing accelerometers and microphones. Certain scenarios may also be prone to misinterpretation, such as the conflation of joints with rail rupture.

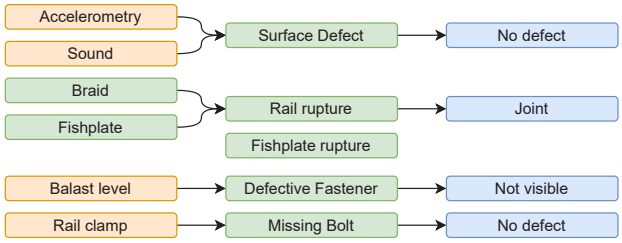

Figure 7. Data fusion discrimination by adding more information to the camera processing incorrect detection.

3.4.2 Decision fusion

The absence of a human operator in an automated detection system emphasises a significant reliance on data and algorithmic precision and integrity [17]. Prior to reporting detections to the central application, it is imperative to diversify detection methods, even if they provide minimal information such as warnings. Aggregating multiple detection sources, even with individually low confidence scores, enhances the overall precision and confidence rate. However, challenges arise due to uncertainties in processing times within the chain, potentially resulting in waiting queues and limiting frames per second (FPS). Determining

the optimal timing for activating data fusion or decision fusion in the event of missing processing is a critical consideration. One approach is to allocate a maximum waiting time, shared between data and decision fusion (Fig. 8). A more improved method involves statistical analysis, combining probabilities, and maximising waiting time based on mean and median processing times of each algorithm to optimise detection precision effectively.

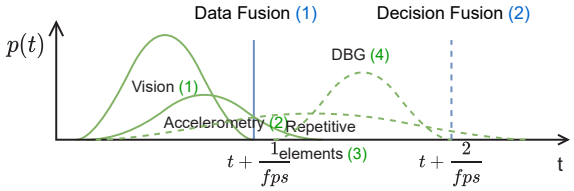

Figure 8. Processing queue example (see Fig. 2 for inputs).

4 Visual-based defect detection

In this use-case, vision is the sensor capturing the most sought-after objects. Its study and implementation offer the possibility to have a first operational system. For that, we chose the recent Yolov8 (You Only Look Once) convolutional neural network (CNN) for its prevalent use in railway track monitoring [7] to process images from 3.A. The first step is to validate its usability and quality to detect either defects and context objects (Fig. 1, 7). The goal is to use the smallest network but some under-represented defects in the dataset may lead to imbalances.

4.1 Dataset

Two datasets have been used for training. They have been manually collected, annotated, and post-training corrected from images captured by a track monitoring vehicle, which are the exclusive property of SNCF Réseau. The annotation has been made by a railway expert but the actual characterisation of a defect may vary from one expert to another, especially for the surface defects.

Our data is composed of two different views: topview (TV) that is of 1504x1500 pixels (grayscale) and sideview (SV) of 768x1508 pixels (rgb). The context dataset [**C**] is composed of 1327 TV + 9585 SV training (Train.) images, 167 TV + 1296 SV validation (Val.) images and 173 TV + 1304 SV test images. The defects dataset [**D**] is composed of 3456 TV + 5779 SV training, 282 TV + 414 SV validation images and 263 TV + 409 SV test (Test) images.

Table 1. Defects original dataset [**D**]

Class	Composition					
	Train	***%***	***Val***	***%***	***Test***	***%***
fastener	8019	64.3	593	67.3	590	68.1
surface	2949	23.6	195	22.1	183	21.1
nut	1502	12.1	93	10.6	93	10.8
Total	12470	87.7	881	6.2	866	6.1

Train. images: 9235, Val. images: 696, Test images: 672

The evaluation dataset is the combination of the two datasets. Object distribution can be found in Tables 2 & 1. This distribution has been made without any prior information on potential training optimisation proportionally to their appearance on the infrastructure. One possible improvement of the overall precision could be an equalisation of class proportion in the datasets.

Table 2. Context original dataset [**C**]

Class	Composition					
	Train	***%***	***Val***	***%***	***Test***	***%***
braid	3786	15.9	522	16.5	524	15.7
clamp	578	2.4	56	1.8	60	1.8
plate	6045	25.5	795	25.2	872	26.2
seal	6661	28.1	868	27.6	935	28.1
weld	2818	11.9	374	11.9	373	11.2
mark	3828	16.2	533	17	563	17
Total	23716	78.5	3148	10.5	3327	11

Train. images: 10912, Val. images: 1463, Test images: 1477

4.2 Results

To the best of our knowledge, the majority of the previous studies focus on a main defect and its sub-type classification. The focus has been put on surface defect classification and missing fasteners attaining a very high detection rate for fasteners. Some studies used the early development of CNN to track components via segmentation [18]. Moreover, recent studies on CNN focused on improving existing models and modifying their structures to improve detection rate while not targeting embedded systems yet [19]. Our concern is about aggregating multiple detections into one real-time embedded system, therefore we analysed the ability to reduce architecture constraints with dataset operations and multiple defect representation strategies. A hypothesis was made on splitting a larger model D+C combining defects and context into two separated models D and C to observe the impact on precision. To this end, two sizes of Yolov8 were trained, nano and extra large with 3 cases: defects only (D), context only (C), and their combination (D+C). Results of Table 3 and Fig. 9 show that decreasing the number of objects into two smaller models leads to improved detection precision while showing that lower numbers of objects do not reach extensively better results on large models. Exception to be made for complex object detection such as surface defects here. Surface defects' broad range of representation is impacted by the model's sharpness that increases with the model's size. Plus, the way of labeling surface defects misleads the model to combine multiple defects into one detection. The detection part on rail ruptures could not have been studied due to no available data. Yet, one can consider that the ruptures can be wrongly classified with rail seals. This problem can be tackled by fusing visual-based defect inference results with context elements and seals spatial repetitiveness.

Metrics used in this paper are Average Precision (eq. 1 AP in %) and timings (in ms) according to models' Floating Point operations (FLOPs) volume (in Gigas).

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}$$

With TP being the True Positives, FP the False Positives and FN the False Negatives.

$$AP(@R) = \int_{R^0}^{R^1} P(r)dr, \quad mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \qquad (1)$$

With P as precision and r as recall, AP summarizes the PR curve at a specific model selectivity (recall) while capturing its significance.

Table 3. Yolov8 mAP@0.5 for all classes, using Intel Xeon 5218 CPU and Nvidia Tesla T4 GPU. The first 3 objects are defects, the last 6 are context (D: Defect, C: Context, D+C: Combined). Input size: max 640x640 pixels. Grayscale transformation added to assess hardware simplification.

Model	AP@0.5 (%)									mAP	*Inference* (ms)	*FLOPs* (G)
	fastener	*surface*	*nut*	*braid*	*clamp*	*fishplate*	*seal*	*weld*	*mark*	*All*	preprocess + infer + post	
Y8n D	**98.2**	67.8	**94.6**	-	-	-	-	-	-	**86.8**	**1.9 + 6.3 + 0.7**	
Y8n D Gray	97.9	65.7	92.2	-	-	-	-	-	-	85.5	**1.9 + 6.3 + 0.7**	
Y8n C	-	-	-	**97.8**	98.8	99.2	86.7	95.8	93.9	95.4	2.0 + 6.3 + 0.9	**8.1**
Y8n C Gray	-	-	-	97.6	98.7	99.2	**88.2**	**95.9**	93.7	95.6	2.0 + 6.3 + 0.9	
Y8n D+C	86.9	57.8	31.3	91.7	63.0	93.2	81.5	91.9	89.6	76.3	2.0 + 6.3 + 0.8	
Y8x D	97.5	**70.5**	92.4	-	-	-	-	-	-	**86.8**	1.9 + 42.0 + 0.7	
Y8x C	-	-	-	97.7	**99.5**	**99.3**	87.6	96.6	**94.6**	**95.9**	1.9 + 39.1 + 1.0	257.4
Y8x D+C	89.1	57.2	33.6	91.2	62.3	94.4	81.8	92.7	88.5	76.8	1.9 + 39.6 + 0.9	

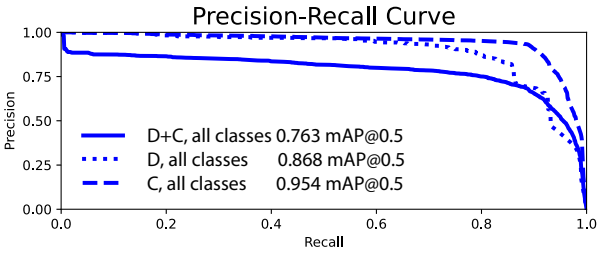

Figure 9. PR Curve for C, D and C+D Yolov8-nano models

5 Conclusions

In this paper, we have studied the architecture of a vision-based detection system for multi-modal monitoring. We proposed splitting an extra-large Yolov8 model (D+C) into two nano models, leading to a later fusion layer that combines the results of each smaller model. This demonstrates that we can reduce the requirements of the architecture while lowering inference times without sacrificing precision performance. Results involving a visual deep-learning-based railway inspection were presented and analyzed.

As a future perspective, our approach involves splitting a larger model into smaller, more manageable sub-models. This offers several advantages: it balances processing demands across multiple hardware units, improves detection precision by enabling optimal fusion techniques, and allows for post-inference fusion using specific rules. Additionally, this method reduces the impact of hardware failures by redistributing disabled processes to other functioning targets, ensuring greater system resilience.

To this end, a hardware-software codesign methodology is required to develop a dedicated embedded system for real-time railway structure monitoring. The system envisioned in our study enables direct savings on infrastructure maintenance and personnel, while also providing socio-economic gains by preventing incidents. The pricing of embedded systems is based on the costs of competing solutions and the economic benefits they offer. Additionally, by using low-cost, off-the-shelf components, we expect achieving lower selling prices compared to existing systems.

References

[1] Transport Regulatory Activities, 'Activity report', (2020).
[2] Eurailscout, 'Monitoring vehicles in France', https://eurailscout-france.fr/nos-engins-utilises-en-france/
[3] M. Mićić et al., 'Inspection of RCF rail defects – Review of NDT methods', Mechanical Systems and Signal Processing, vol. 182, p. 109568, (2023).
[4] F. Akhyar et al., 'A Real-Time Application for Rail Surface Defect Inspection Utilizing Rectangular-Shaped Labels', 2nd International Conference on Computer System, Information Technology, and Electrical Engineering, pp. 214–219, (2023).
[5] X. Chu et al., 'Defect Detection for a Vertical Shaft Surface Based on Multimodal Sensors', IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 8109–8117, (2022).
[6] C. Mandriota et al., 'Filter-based feature selection for rail defect detection', Machine Vision and Applications, vol. 15, no. 4, pp. 179–185, (2004).
[7] N. Anwar et al., 'YOLOv4 Based Deep Learning Algorithm for Defects Detection and Classification of Rail Surfaces', IEEE International Intelligent Transportation Systems Conference, pp. 1616–1620, (2021).
[8] SNCF Réseau and Eurailscout, AIDO Railway AI video monitoring, (2023).
[9] G. Kefalidou et al., 'Identifying rail asset maintenance processes: a human-centric and sensemaking approach', Cogn Tech Work, vol. 20, no. 1, pp. 73–92, (2018).
[10] Z. Popović et al., 'The Importance of Rail Inspections in the Urban Area -Aspect of Head Checking Rail Defects', Procedia Engineering, vol. 117, pp. 596–608, (2015).
[11] T. for NSW, 'Rail Defects Handbook 1.2'. Available: https://www.transport.nsw.gov.au/node/7836
[12] M. Porat and Y. Y. Zeevi, 'The generalized Gabor scheme of image representation in biological and machine vision', IEEE Trans. Pattern Anal. Machine Intell., vol. 10, no. 4, pp. 452–468, (1988).
[13] M. Molodova et al., 'Health condition monitoring of insulated joints based on axle box acceleration measurements', Engineering Structures, vol. 123, pp. 225–235, (2016).
[14] J. Lee et al., 'Fault Detection and Diagnosis of Railway Point Machines by Sound Analysis', Sensors, vol. 16, no. 4, Art. no. 4, (2016).
[15] B. M. Dash et al., 'A Comparison of Model-Based and Machine Learning Techniques for Fault Diagnosis', in 2022 23rd International Middle East Power Systems Conference, pp. 1–7, (2022).
[16] D. Lahat et al., "Challenges in multimodal data fusion," 2014 22nd European Signal Processing Conference (EUSIPCO), pp. 101-105, (2014).
[17] F. Vanderhaegen et al., 'Human factors and automation in future railway systems', Cogn Tech Work, vol. 23, no. 2, pp. 189–192, (2021).
[18] F. Guo et al., 'Real-time railroad track components inspection based on the improved YOLOv4 framework', Automation in Construction, vol. 125, p. 103596, (2021).
[19] C. Zhang et al., 'Rail Surface Defect Detection Based on Image Enhancement and Improved YOLOX', Electronics, vol. 12, no. 12, Art. no. 12, (2023).