



SMART RAIL VEHICLE MONITORING: A STUDY OF CNN ARCHITECTURES

Marko PERIĆ¹ [0000-0001-9350-0351]
Aleksandar MILTENOVIĆ² [0000-0002-1453-2548]
Emina PETROVIĆ³ [0000-0002-4230-2416]
Damjan RANGELOV⁴ [0000-0003-3071-893X]
Saša PAVLOVIĆ⁵ [0000-0001-8806-6693]

Abstract – This paper presents performance analysis of two deep learning architectures for object detection YOLOv2 and YOLOv4-Tiny aimed at identifying the model state for inspecting the braking systems of Series 441 and 444 locomotives. These models were trained and evaluated using a unified dataset comprising images of 3220 Braking system, as well as instances where these components were absent. Model evaluation was conducted using standard metrics, with a particular focus on the Precision-Recall curve and Average Precision (AP). Experimental results demonstrate that the YOLOv2 architecture achieving a high AP value of 0.97199 after 700 training iterations, accompanied by rapid loss function convergence. In contrast, YOLOv4-Tiny achieved stable convergence after 2500 iterations but yielded a lower AP value of 0.92546. These findings indicate that YOLOv2 can deliver superior performance and higher precision in object detection tasks where fine details are not the primary focus, while simultaneously significantly reducing the demand for complex computational resources.

Keywords – Monitoring, object detection, YOLOv2, YOLOv4-Tiny, locomotive braking systems.

1. INTRODUCTION

The concept of smart maintenance of rail vehicles has become a pillar of the modernization of modern railway transport. In systems where safety does not allow for compromises, such as locomotive braking systems, the transition from periodic manual inspections to automated, continuous inspection is a technological imperative [1,2]. Brake shoes and their fuses are exposed to extreme mechanical and thermal stresses, which is why their damage or falling out directly threatens the stability of the train and traffic safety. Nevertheless, the visual inspection of these components on the undercarriage still relies predominantly on the human factor, which slows down processes in the depot and increases the risk of oversights due to controller fatigue. With the advent of Industrial Revolution 4.0, computer vision and deep learning algorithms have transformed material and component inspection systems [3,4]. Convolutional neural networks made it possible to automate the process of recognizing anomalies with a precision that rivals that of a human, but with a

drastically higher speed [5,6,7]. The key challenge in practice is no longer just achieving high accuracy, but finding computationally efficient models that can be successfully implemented in the field, often on hardware with limited performance [8,9].

This paper analyzes and provides a solution to the problem through the development and evaluation of a system for automatic detection of brake system elements on electric locomotives series 441, 444 and 461. In the paper, a robust detection model was developed based on the analysis of a unique set of 3220 photos, which includes correct conditions, but also critical situations of the absence of brake shoes or fuses. The research focus is set on a rigorous comparison of two architectures of different generations: YOLOv2 and YOLOv4-Tiny. Through the analysis of convergence speed, loss functions and precision-recall curves, the paper demonstrates how older and more optimized architectures under certain conditions can outperform more complex models, offering high accuracy with minimal hardware requirements, which is crucial for economically viable

¹ Faculty of Mechanical engineering, University of Niš, Niš, Serbia, marko.peric@masfak.ni.ac.rs

² Faculty of Mechanical engineering, University of Niš, Niš, Serbia, aleksandar.miltanovic@masfak.ni.ac.rs

³ Faculty of Mechanical engineering, University of Niš, Niš, Serbia, emina.petrovic@masfak.ni.ac.rs

⁴ Faculty of Mechanical engineering, University of Niš, Niš, Serbia, damjan.rangelov@masfak.ni.ac.rs

⁵ Faculty of Mechanical engineering, University of Niš, Niš, Serbia, sasa.pavlovic@masfak.ni.ac.rs

application in railway depots.

2. CNN ARCHITECTURES

In this research, the architectures YOLOv2 and Yolov4 - tiny are taken, these CNN models, treat detection as a unique regression problem. By dividing the image into a grid, the prediction of boundary frames and classes is performed in only one pass (single-stage detection), which enables extremely high processing speed in industrial and transport systems. YOLOv2 relies on the DarkNet-19 classification network with convolutional layers (3×3 and 1×1) and mandatory mini-group normalization. The key innovations of this architecture are the introduction of Anchor boxes (predefined frames obtained by clustering) and a Passthrough layer that merges high- and low-resolution feature maps for more efficient detection at a single output scale [10,11]. YOLOv4 brings a technological leap through a more complex topology divided into three segments, the first of which is called Backbone, which uses CSPDarknet53 that shares feature maps, which drastically reduces computing operations and memory load. The second segment is the Neck, which combines SPP (Spatial Pyramid Pooling) and FPN/PAN (Path Aggregation Network) structures for top-down and bottom-up semantic pooling of information. The third segment is the Head, which performs detection on several different scales (multiple scaling), which significantly increases the accuracy of recognizing extremely small or large objects in the same scene [12,13].

3. BRAKE SYSTEM

Continuous monitoring of railway braking systems is critical for traffic safety, as brake shoes and their lock pins (keys) endure extreme mechanical and thermal stresses. Traditional visual inspections performed by drivers or inspectors are labor-intensive, prone to human error from fatigue, and require shifting locomotives to specialized inspection channels, which reduces operational availability and increases maintenance costs. To automate this process and enhance inspection reliability, deep learning models combined with automatic undercarriage image acquisition offer a highly efficient alternative. Consequently, this paper focuses on detection of mechanical brake and analysis of architectures between YOLOv2 and Yolov4 - Tiny.

Monitoring in the railway transport is a critical aspect of ensuring the safety, efficiency, and reliability of rail transport systems. It involves the continuous observation and assessment of various components, systems, and operations to detect potential issues before they escalate into failures [14].

4. DATASET AND IMAGE ANNOTATION

The images were captured using a GoPro Hero 11 camera. A total of 3220 images were used for detector training. The original resolution of images, 2774 high-resolution images measuring 5568×4872 pixels and 566 images, with a resolution of 1920×1080 pixels. During preprocessing, the final resolution was adjusted to 416×416 for YOLOv2 and 608×608 for YOLOv4-Tiny.

Each image is manually annotated by drawing a bounding box around the entire braking system. The annotations formed in this way, which contain information about the position and dimensions of the observed object in the image, were used to train the YOLOv2 and YOLOv4-Tiny detector.

The first state of Braking system “Brake shoes and key present” included 1529 images, the second state “Brake shoes and key missing” included 686 images, while the third state “Brake shoe key missing” included 1005 images.



Fig.1. Braking system on locomotive 441/444

5. DETECTION

To simplify the localization process, the task is defined as a single-class detection problem, where the model focuses exclusively on identifying and bounding the complete mechanical brake system on the vehicle's undercarriage (Figure 2).

For the development and performance assessment of the detectors, a subset of 3220 images was extracted from the main dataset. This annotated image corpus was randomly split, allocating 60% of the data for training, 10% for validation during the optimization phase, and the remaining 30% as an independent test set.

Parameters used in training CNNs is shown in Table 1.

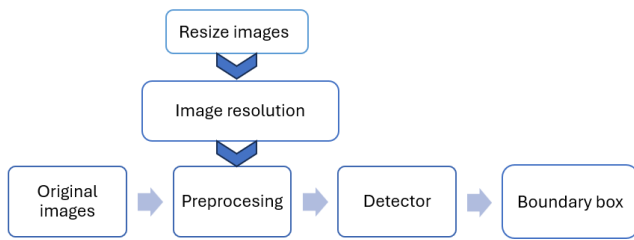


Fig.2 Flow of data processing from the input image to the output (boundary box)

Tab. 1. Parameters used in training CNNs.

Parameter	Value
Optimizer	adam
Gradient Decay Factor	0.9
Learn Rate Schedule	piecewise
Initial learning rate	0.001
Max Epochs	80
Mini-Batch Size	64
Learn Rate Drop Period	10
Learn Rate Drop Factor	0.6

6. RESULTS AND DISCUSSION

The training processes and experimental detection performances of the YOLOv2 and YOLOv4-Tiny detectors are evaluated using their respective loss

convergence trajectories and Precision - Recall characteristics. The learning behaviors of the two models exhibit distinct patterns in terms of optimization speed and convergence stability. The loss curve of the YOLOv2 detector demonstrates a highly rapid decrease and reaches early stabilization, achieving an optimal state within 1000 training iterations. Conversely, the YOLOv4-Tiny detector requires a prolonged learning process of 2500 iterations to achieve stable convergence, which is visualized using a logarithmic loss scale due to its structural complexity. For these architectures, the minimal gap between the final training and validation loss values indicates strong generalization capabilities and confirms that the models avoided overfitting on the undercarriage dataset.

The Precision-Recall curves further illustrate the definitive detection capabilities of the trained models. Evaluating the detection quality via Average Precision (AP) reveals that YOLOv2 significantly performs YOLOv4-Tiny, achieving an AP value of 0.97199 (97.199%) compared to 0.92546 (92.55%) for the latter. This performance gap provides a crucial architectural insight for locomotive undercarriage inspection. Consequently, YOLOv2 delivers higher localization accuracy and stable detection performance under varying confidence thresholds.

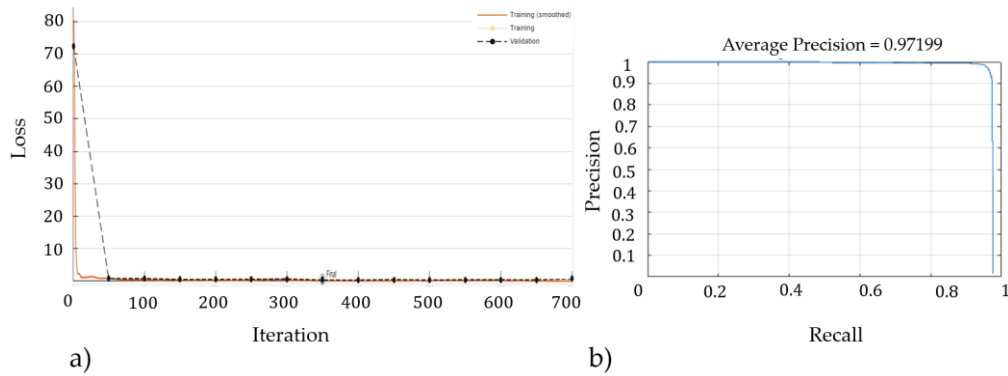


Fig. 3. Training loss curve (a) and precision-recall curve (b) for the proposed YOLOv2 object detector.

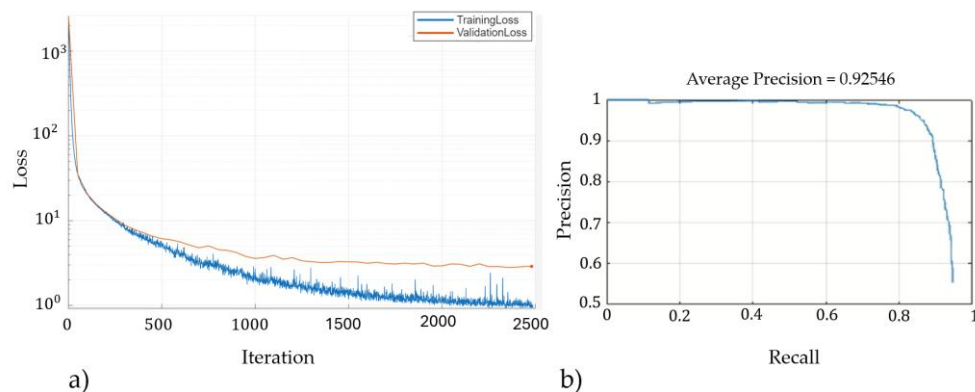


Fig. 4. Training loss curve (a) and precision-recall curve (b) for the proposed YOLOv4 - Tiny object detector.

CONCLUSION

Through this architectural evaluation, the paper demonstrates that under reduced input resolutions, a direct single-scale architecture like YOLOv2 also can give good results, achieving a high average precision of 0.97199 for less complex target objects. In contrast, newer, models like YOLOv4-Tiny struggle to leverage their structural advantages when fine spatial details are minimized. In the railway sector, these results confirm that traditional, manual visual inspections of locomotive undercarriages can be effectively enhanced using artificial intelligence. By expanding the dataset, ensuring precise annotations, and balancing model capacity with input scaling, the efficiency and safety monitoring of railway maintenance can be significantly improved.

ACKNOWLEDGEMENT

This research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (Contract No. 415-03-137/2025-03/200109) and Innovation Fund of the Republic of Serbia in the frame of the Collaborative Grant Scheme Program (Contract No. IF50348).

REFERENCES

- [1] D. Stamenković, A. Miltenović, M. Banić, M. Simonović, M. Perić and D. Rangelov, *Application of the Device for Inspecting the Undercarriage of Railway Vehicles in the Framework of Inspection*, *Železnice* Vol. 69, 95-101, 2024.
- [2] R. Lagneback, *Evaluation of wayside condition monitoring technologies for condition-based maintenance of railway vehicles*, Luleå University of Technology, Sweden, 2007.
- [3] S. Mittal and D. Rao, *Vision Based Railway Track Monitoring using Deep Learning*, Computer Science, <https://doi.org/10.48550/arXiv.1711.06423>, 2018.
- [4] R. L. Rocha, A. Carolina Q. Siravenha, A. Cláudia S. Gomes, and other, *A Deep-Learning-Based Approach for Automated Wagon Component Inspection*, DOI: 10.1145/3167132.3167157, April 2018.
- [5] S. Takaya and M. Rehmsmeier, *The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets*, *PLOS ONE* Vol. 10, Issue 3, e0118432, 2015.
- [6] S. K. Bussa and N. K. Boppana, *Enhanced ResNet50 deep learning algorithm for classification of crack images in RCC structures*, *Asian J. Civ. Eng.* Vol. 26, 3773–3784, <https://doi.org/10.1007/s42107-025-01396-7>, 2025.
- [7] K. Krishnamohan, Ch. Raghava Prasad and P. V. V. Kishore, *Train rolling stock video segmentation and classification for bogie part inspection automation: a deep learning approach*, *Journal of Engineering and Applied Science*, <https://doi.org/10.1186/s44147-022-00128-x>, 2022.
- [8] J. Sánchez-Cubillo, J. Del Ser and J. L. Martín, *Toward Fully Automated Inspection of Critical Assets Supported by Autonomous Mobile Robots*, *Vision Sensors, and Artificial Intelligence*, MDPI, *Sensors* Vol. 24, Issue 12, 3721, <https://doi.org/10.3390/s24123721>, 2024.
- [9] Marko Perić, Aleksandar Miltenović, Milan Banić, Szabolcs Fischer, *Visual Inspection in Transport Using Autonomous Robots*, *Cognitive Mobility*, 2024.
- [10] J. Redmon and A. Farhadi, *YOLO9000: Better, Faster, Stronger*, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517-6525, 2017.
- [11] P. Jiang, D. Ergu, F. Liu, Y. Cai and B. Ma, *A Review of Yolo Algorithm Developments*, *Procedia Computer Science* Vol. 199, 1066-1073, 2022.
- [12] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, *YOLOv4: Optimal Speed and Accuracy of Object Detection*, arXiv preprint, arXiv:2004.10934, 2020.
- [13] T. Diwan, G. Anirudh and J. V. Tembhurne, *Object detection using YOLO: Challenges, architectural successors, datasets and frameworks*, Elsevier, *Image and Vision Computing* Vol. 131, 104639, 2023.
- [14] V. Aleksandrov, *Železnička vučena vozila*, 1st ed.; Želnid: Belgrade, Serbia, 2000; pp. 352–371.