An Obstacle Intrusion Detection Method for Complex Environment Tracks Based on Manifold Regularization

WEN-QIANG LIU, SU-MEI WANG, XIN-YUE XU
and YI-QING NI

ABSTRACT

Tracks are essential infrastructure for high-speed railroads and are the key to the safe passage of trains. Once obstacles intrude into the track area, it will threaten the safety of train operations. Therefore, there is an urgent need to propose a fast and efficient obstacle intrusion detection method. However, it greatly challenges obstacle intrusion detection in the complex natural environment (rain, fog, light changes, etc.). To this end, this paper proposes an obstacle intrusion detection method for complex environment tracks based on manifold regularization. This method uses the single-stage object detection framework YOLO as the basic structure, integrates the multi-head attention mechanism to improve the object detection performance, and uses the dilated convolution structure to reduce the model parameters and improve the detection efficiency; in the feature extraction space, it introduces the manifold regularization constraint to realize the alignment constraint of various categories of features under different natural environment images and improves the generalization performance of the detection model. The results show that the proposed method can adapt to various complex natural environment detection conditions and effectively improve the accuracy and generalization of the obstacle intrusion detection method.

1. INTRODUCTION

With the rapid development of rail transit in the world, the comprehensive construction and operation of high-speed railways, subways, urban rails, and intercity railways have become increasingly prominent, especially the intrusion of obstacles in the track area has brought serious safety hazards to driving safety [1]. For example, on February 6, 2020, when a train on the Italian high-speed railway passed near the northern Italian city of Lodi, the train collided with a freight train on the track, causing the train to derail and killing two drivers. On April 2, 2021, a train No. 408 from Shulin Station to Taitung Station in Taiwan collided with an engineering vehicle that had landslide-invaded the route between Heren Station and Chongde Station in Xiu-lin Township, Hualien County, derailed and rushed into the tunnel and rubbed against the tunnel wall, killing 49 people and injuring 213 people. On June 7, a train collision occurred in Pakistan, killing 65 people and hurting 150 people. To this end, in response to on-site needs, it is urgent to propose a set of track foreign body intrusion detection algorithms with fast response speed, high detection accuracy, and strong system robustness.

Wen-Qiang Liu (Research Fellow), Su-Mei Wang (Research Assistant Professor), Xin-Yue Xu (PhD candidate) and Yi-Qing Ni (Chair Professor) are with the National Rail Transit Electrification and Automation Engineering Technology Research Center (Hong Kong Branch), Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

In recent years, some researchers have invested a lot of scientific research funds and workforce in researching obstacle intrusion detection and have achieved a series of results [2-7]. For example, Kapoor *et al.* [3] used transfer learning methods to fine-tune the Faster RCNN [8] model pre-trained on a public dataset of track images collected by infrared cameras to improve its detection performance. He *et al*. [4] proposed an improved YOLOv4 [9] track foreign object detection network to overcome the problems of low accuracy and poor real-time performance of traditional detection methods. However, labeling data is a massive project, and covering data in all environments is difficult. The above methods only use limited labeled data for training, and it isn't easy to ensure the generalization of the detection method. Therefore, this paper proposes an obstacle intrusion detection method for complex environment tracks based on manifold regularization.

2. METHODOLOGY

This proposed method adopts the single-stage object detection framework YOLOv11 [10] as the basic structure, integrates the multi-head attention mechanism to improve the object detection performance, and uses the dilated convolution structure to reduce the model parameters and improve the detection efficiency; in the feature extraction space, it introduces the manifold regularization [11] constraint to realize the alignment constraint of various categories of features under different natural environment images and improves the generalization performance of the detection model. The details of the proposed method will be introduced in detail below.

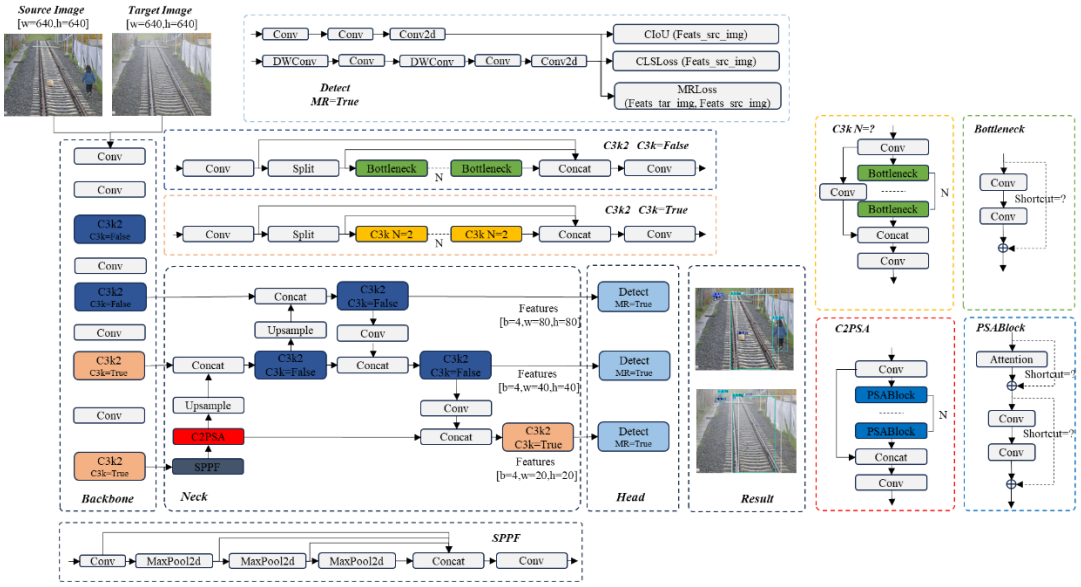


Figure 1. Schematic diagram of the network structure of the proposed method.

2.1 SINGLE STAGE OBJECT DETECTION MODEL

The detection model consists of three modules: backbone, neck, and head. The specific network functions and core modules will be introduced in detail below.

(a) Backbone with C3k2 Blocks

The backbone module is mainly used to extract object features of different scales. In this paper, the backbone leverages C3k2 blocks (Cross-Stage Partial Kernel-2), an enhanced version of CSP (Cross-Stage Partial Networks). These blocks split feature maps into two parts: one undergoes a series of Bottleneck layers with 3×3 convolutions, while the other bypasses these layers. The outputs are concatenated, improving feature fusion efficiency while reducing computational costs. The Bottleneck layer consists of two 3×3 convolutions with SiLU activation and batch normalization. The detailed structure is shown in Figure 1.

(b) Neck with SPPF (Spatial Pyramid Pooling Fast) and C2PSA (Cross-Stage Partial with Spatial Attention)

The neck module is used to combine features of different scales to improve object detection accuracy. In this paper, the neck integrates SPPF, a lightweight variant of spatial pyramid pooling. It applies max-pooling with kernel sizes 5×5, 9×9, and 13×13 to capture multi-scale features, enhancing small-object detection. In addition, it introduces attention mechanisms C2PSA. It refines spatial focus by splitting features into parallel branches. Each branch applies position-sensitive attention to highlight critical regions, improving the detection of occluded or small objects. The detailed structure is shown in Figure 1.

(c) Multi-Scale Detection Head

The head module is used to detect and output objects. In this paper, the head outputs detection boxes for three different scales (low, medium, and high) using the feature maps generated by the backbone and neck. It generates predictions at three scales (P3, P4, P5) using feature maps from the backbone and neck. Each scale predicts bounding boxes and class probabilities. In this module, it uses the dilated convolution structure to reduce the model parameters and improve the detection efficiency. The detailed structure is shown in Figure 1.

Three types of losses are mainly constructed for training the object location bounding box and category. As shown in Formula (1), CIOU (Complete IOU) and DFL (Distribution Focal Loss) are combined for location regression; BCE (Binary Cross Entropy) is used for object recognition.

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}}\mathcal{L}_{\text{CIoU}} + \lambda_{\text{cls}}\mathcal{L}_{\text{BCE}} + \lambda_{\text{obj}}\mathcal{L}_{\text{DFL}} \tag{1}$$

where $\lambda_{\text{box}}, \lambda_{\text{cls}}, \lambda_{\text{obj}}$ are the weight coefficients of different losses.

2.2 The Proposed Method with Manifold Regularization

Manifold regularization is a semi-supervised learning framework introduced by Belkin et al. [11] that leverages the geometric structure of data (assumed to lie on a low-dimensional manifold) to improve generalization. It extends traditional regularization by incorporating both labeled and unlabeled data, enforcing the smoothness of the learned function over the underlying manifold.

This approach is particularly effective when labeled data is scarce but unlabeled data is abundant, as it exploits the intrinsic data geometry to constrain the hypothesis space. Therefore, this paper introduces Manifold regularization into the obstacle intrusion detection network to improve the detection accuracy of unlabeled samples in the target domain. The detailed structure is shown in Figure 1.

As shown in Formula (2), The objective function of manifold regularization combines (a) supervised loss V for labeled data, (b) traditional regularizer $\| f\|^2_{\mathcal{H}_K}$ to control model complexity, (c) manifold regularizer $\| f\|^2_I$ to enforce smoothness on the data manifold, and typically defined using the graph Laplacian L shown in Formula (3). So, by introducing Manifold Regularization, the final loss of the proposed method is shown in Formula (4).

$$\min_{f\in\mathcal{H}_K} \quad \frac{1}{l}\sum_{i=1}^{l} V(f(x_i), y_i) + \gamma_A \| f\|^2_{\mathcal{H}_K} + \gamma_I \| f\|^2_I \tag{2}$$

$$\| f\|^2_I = \frac{1}{(l+u)^2}\sum_{i,j=1}^{l+u} W_{ij}\left(f(\mathbf{x}_i) - f(\mathbf{x}_j)\right)^2 = \mathbf{f}^\top L\mathbf{f} \tag{3}$$

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}}\mathcal{L}_{\text{CIoU}} + \lambda_{\text{cls}}\mathcal{L}_{\text{BCE}} + \lambda_{\text{obj}}\mathcal{L}_{\text{DFL}} + \gamma_A \| f\|^2_{\mathcal{H}_K} + \gamma_I \| f\|^2_I \tag{4}$$

where $\mathcal{H}_k$ represents Reproducing Kernel Hilbert Space (RKHS) for the function class, γ_A and γ_I are the trade-off parameters controlling regularization strength, W_{ij} is the similarity matrix, and $L = D - W$ (degree matrix D, edge weights W).

3. EXPERIMENTS AND ANALYSIS

3.1 EXPERIMENTAL ENVIRONMENT

The program development environment is as follows: L40S graphics card, 128GB RAM, 300GB hard disk, AMD Epyc-milan processor with 12-core CPU. All programs are developed on the Ubuntu 22.04 system using PyTorch and the Ultralytics framework. The data comes from the obstacle intrusion research system test field, which can simulate various complex environments and working conditions. The image resolution is resized to 640×640 as the input. There are 2133 labeled images in the source domain dataset with the normal weather, which are used as training datasets. There are 1280 unlabeled images in the target domain with bad weather, of which 640 are used as training datasets and 640 are used as testing datasets. There are five classes of obstacles, including brown box, people, rail, rail area, and white box.

3.2 MODEL PARAMETER SETTING AND EVALUATION INDEX

The Adam optimizer is adopted to train the model parameters, with the weight decay rate set to 0.0005, the momentum at 0.937, and the basic learning rate set to 0.001. A total of 100 epochs are trained, and the batch size is set to 4. The learning rate update way selects the Warmup strategy and its setting parameters: epoch is 3, momentum is 0.8, and bias learning rate is 0.1. The model parameters are initialized with the pre-trained model parameters by the public dataset COCO dataset.

The average precision (*AP*), the mean average precision (*mAP*), the *P*-*R* curve, and the confusion matrix are selected as the evaluation indicators of the proposed detection method. The calculation principle of *AP* and *mAP* is as shown in Formulas (6-9), where

TP is the true positive sample, *FP* is the false positive sample, and *FN* is the false negative sample. *P* is the precision rate, *R* is the recall rate, and *Q* is the number of component categories. The confusion matrix is an error matrix used to compare the prediction results with the true detection values.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$AP = \int_0^1 P(r)dr \tag{7}$$

$$mAP = \frac{1}{Q}\sum_{i=1}^{Q} AP_i \tag{8}$$

3.3 EXPERIMENT RESULTS AND ANALYSIS

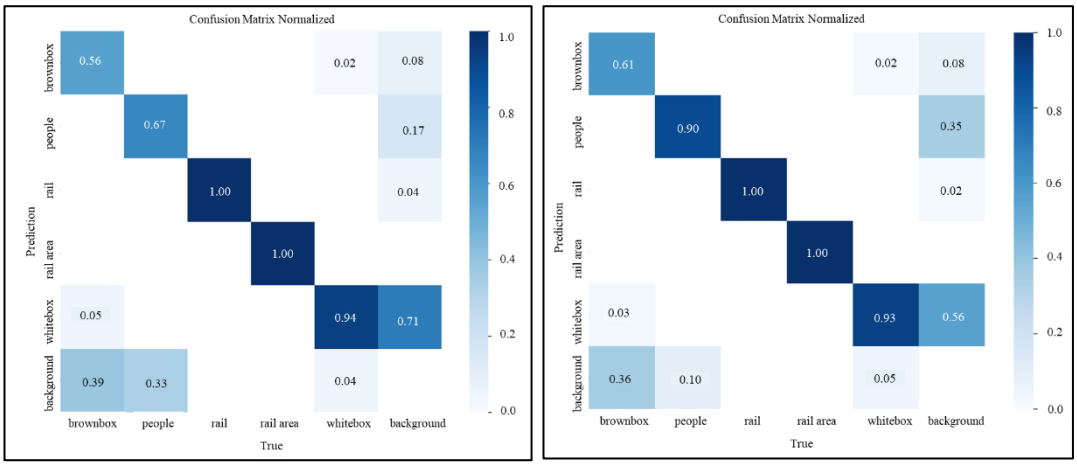
In order to demonstrate the effectiveness of the proposed method, we compared the detection effects of different methods on track obstacles on the testing dataset. The results are statistically analyzed and shown in Figures 2, 3, 4, and Table I.

Figure 2 shows the detection results of the confusion matrix. By analyzing the modified figure, it can be seen that the original YOLOv11 is not as effective as the method proposed in this paper for detecting brown boxes and people, especially for detecting people, with a detection accuracy difference of 23 percentage points. Analysis of the reasons shows that people's clothing in the source and target domain data is entirely different, as shown in Figure 4 (a) and (b). When training YOLOv11, only the labeled data in the source domain is used, so the detection effect is poor. When training the method proposed in this paper, the unlabeled data in the target domain is also used in addition to the labeled data in the source domain. Through manifold regularization, the distribution of the target domain data is made close to the source domain data, thereby effectively improving the detection accuracy of the unlabeled data in the target domain.

Figure 3 and Table 1 show the P-R curve and the *AP* and *mAP* index results, respectively, where *AP*@50 means that when the IOU overlap between the predicted target and the true target is 50%, it is regarded as *TP*, and *AP*@50:95 means the average value of *AP* at different IOU thresholds with an interval of 0.05. The analysis of the results shows that the overall *mAP* index accuracy of the method proposed in this paper is better than that of YOLOv11. In a single category, the detection accuracy of people is still 6.9 percentage points and 8.0 percentage points higher than that of YOLOv11. In addition, by comparing with the results of the confusion matrix, it can be seen that although both methods will have some targets that are not detected and mistakenly identified as background, they are rarely mistakenly identified as positive sample *FP*, which effectively reduces the false detection rate and avoids the hidden dangers of false alarm warning.

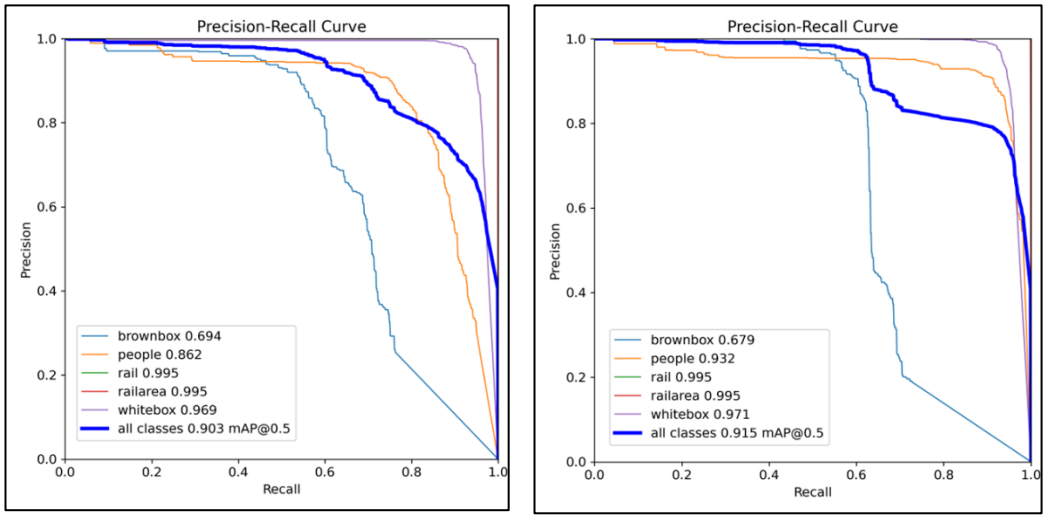
Figure 4 shows the detection effects of different methods. By analyzing this figure, we can see that YOLOv11 and the method proposed in this paper can effectively detect obstacles for source domain data. However, YOLOv11 failed to detect people and brown boxes effectively for target domain data, while the method proposed in this paper

can effectively identify objects. Therefore, the experimental results prove that the method proposed in this paper can effectively detect target domain data while ensuring the detection effect of source domain data.



(a) (b)

Figure 2. Confusion matrix results. (a) Results of YOLOv11; (b) Results of the proposed method.



(a) (b)

Figure 3. *P-R* curve results. (a) Results of YOLOv11; (b) Results of the proposed method.

TABLE I. DETECTION RESULTS ON TESTING DATASET

Classes	**YOLOv11**		**Our Proposed Method**	
	***AP*@50**	***AP*@50:95**	***AP*@50**	***AP*@50:95**
Brown box	0.692	0.392	0.679	0.404
People	0.863	0.492	**0.932**	0.572
Rail	0.995	0.985	0.995	0.986
Rail area	0.995	0.995	0.995	0.995
White box	0.969	0.672	0.971	0.677
mAP	**0.903**	**0.707**	**0.915**	**0.727**

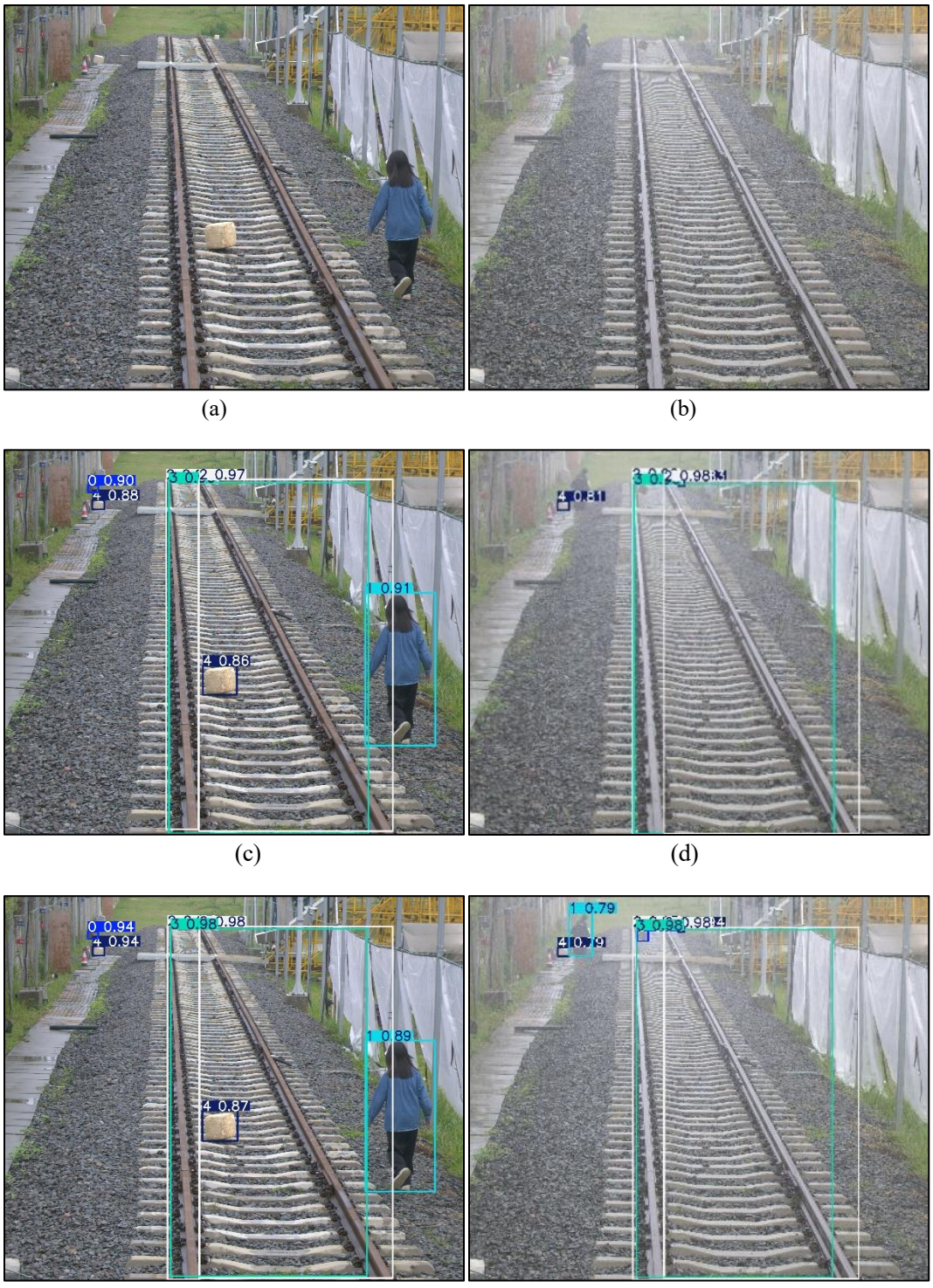


Figure 4. Image detection effect with different methods. (a) and (b) are images from two image domains; the former is the source domain image, and the latter is the target domain image; (c) and (d) are the detection results of YOLOv11; (e) and (f) are the detection results of the proposed method.

4. CONCLUSION

This paper proposed an obstacle intrusion detection method for complex environment tracks based on manifold regularization. By using manifold regularization, the unlabeled sample features of the target domain are adaptively aligned to the labeled sample features of the source domain. It can effectively solve the problem of detecting unlabeled samples in the target domain under complex environments. Without increasing the detection time, the model can improve the detection accuracy of obstacles in the target domain while reducing the false detection rate and ensuring the robustness of the detection system.

ACKNOWLEDGEMENT

The research is supported by the Innovation and Technology Commission of the Hong Kong SAR Government (Grant No. K-BBY1), in part by Wuyi University's Hong Kong and Macao Joint Research and Development Fund (Grants No. 2019WGALH15 and 2019WGALH17), in part by the university research facility in big data analytics (UBDA) of The Hong Kong Polytechnic University.

REFERENCES

1. Alexandrescu, A. R., Manole, A., and Diosan, L. (2023). "Railway Switch Classification Using Deep Neural Networks," In *VISIGRAPP (4: VISAPP)* (pp. 769-776)..
2. Selver, A. M., Ataç, E., Belenlioglu, B., Dogan, S., and Zoral, Y. E. (2018). "Visual and LIDAR data processing and fusion as an element of real time big data analysis for rail vehicle driver support systems," *Innovative Applications of Big Data in the Railway Industry*, 40-66.
3. Kapoor, R., Goel, R., and Sharma, A. (2020). "Deep learning based object and railway track recognition using train mounted thermal imaging system," *Journal of Computational and Theoretical Nanoscience*, *17*(11), 5062-5071.
4. HHe, D., Zou, Z., Chen, Y., Liu, B., Yao, X., and Shan, S. (2021). "Obstacle detection of rail transit based on deep learning," *Measurement*, 176: 109241.
5. Ristić-Durrant, D., Franke, M., and Michels, K. (2021). "A review of vision-based on-board obstacle detection and distance estimation in railways," *Sensors*, 21(10), 3452.
6. Ristić-Durrant, D., Haseeb, M. A., Banić, M., Stamenković, D., Simonović, M., and Nikolić, D. (2022). "SMART on-board multi-sensor obstacle detection system for improvement of rail transport safety," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 236(6), 623-636.
7. Tagiew, R., Leinhos, D., von der Haar, et al. (2023). "Sensor system for development of perception systems for ATO," *Discover Artificial Intelligence*, 3(1), 22.
8. Ren, S., He, K., Girshick, R., and Sun, J. (2016). "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
9. Bochkovskiy, A., Wang, C. Y., and Liao, H. Y. M. (2020). "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv*:2004.10934.
10. Khanam, R., and Hussain, M. (2024). "Yolov11: An overview of the key architectural enhancements," *arXiv preprint arXiv*:2410.17725.
11. Belkin, M., Niyogi, P., and Sindhwani, V. (2006). "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, 7(11).