Automatic Surveillance of People and Objects on Railway Tracks

Domingo Martínez Núñez[1], Fernando Carlos López Hernández[2], J. Javier Rainer Granados[3]

[1] Central Control Station, Metro de Madrid 28029, Madrid (Spain)
[2] Applied Mathematics Department, Universidad Complutense de Madrid (UCM), 28040, Madrid (Spain)
[3] Universidad Internacional de La Rioja (UNIR) 26006, Logroño (Spain)

* Corresponding author: domingo.martinez@metromadrid.es (D. Martínez Núñez), fclh@ucm.es (F. C. López Hernández), javier.rainer@unir.net (J. J. Rainer Granados).

Received 30 October 2023 | Accepted 19 July 2024 | Early Access 20 August 2024

Abstract

This paper describes the development and evaluation of a surveillance system for the detection of people and objects on railroad tracks in real time. Firstly, the paper evaluates several background subtraction techniques including CNNs and the object detection library called YOLO. Then we describe a novel strategy to mitigate the occlusion caused by the perspective of the camera and the integration of an alarms and pre-alarms policy. To evaluate its performance, we have implemented and automated the control and notification aspects of the surveillance system using computer vision techniques. This setup, running on a standard PC, achieves an average frame rate of 15 FPS and a latency of 0.54 seconds per frame, meeting real-time expectations in terms of both false alarms and precision in operational mode. The results from experiments conducted with a publicly available recorded video dataset from Metro de Madrid facilities demonstrate significant improvements over current state-of-the-art solutions. These improvements include better accident anticipation and enhanced information provided to the operator using a standard low-cost camera. Consequently, we conclude that the approach described in this paper is both effective and a more practical, cost-efficient alternative to the other solutions reviewed.

Keywords

Computer Vision,
Machine Learning,
Neural Networks,
Railway Safety,
Surveillance.

DOI: 10.9781/ijimai.2024.08.004

I. Introduction

CURRENTLY modern railroad facilities use video cameras that transmit signals to a limited set of monitors, usually called CCTV (Closed Circuit Television), for the detection of people and objects in dangerous situations. They include manual surveillance and alarms that can interrupt the circulation of trains and warn the security services. For this kind of automatic surveillance problems, well-known computer vision (CV) techniques, including object detection and image background subtraction, as well as neural network classification are frequently implemented in other problem domains.

This research contributes to improving railway safety. In addition to the integration of current computer vision techniques and deep learning algorithms, we describe a system capable of detecting in real time the presence of people and objects on the tracks, in a way that overcomes the traditional methods' limitations. The relevance of this work lies in its ability to offer an economical and effective alarm system, which relies on the use of a single low-cost camera per monitored area, thus boosting its technical and economic convenience over other more costly or complex approaches.

Our research contributes to railroad security, as it combines two innovative techniques to achieve a more practical and economically efficient surveillance, using only a low-cost single camera per monitored area (480x640), and using an alarm policy effectively.

This system contributes to the minimization of accidents that disturb the smooth running of railroad lines, due to people or objects falling on the tracks, as well as violations of railroad rules during service hours (suicides, graffiti, vandalism, crossing of tracks, assaults, etc.).

The automatic processing of the camera images lets us monitor in a more continuous and systematic way than humans could without interruptions, avoiding visual fatigue [1], distinctions, and thus letting the personnel to attend to other duties. Our system is semi-automatic because it reports potentially dangerous situations alerting the security station staff or traffic controllers, who will be able to stop the trains.

The use of affordable equipment is practical and beneficial for successful railroad surveillance. While researchers have reported successful approaches using expensive equipment such as LiDAR, laser, or multiple cameras for object detection on railroads [2], these

Please cite this article as:
D. Martínez Núñez, F. C. López Hernández, J. J. Rainer Granados. Automatic Surveillance of People and Objects on Railway Tracks, International Journal of Interactive Multimedia and Artificial Intelligence, (2024), http://dx.doi.org/10.9781/ijimai.2024.08.004

methods are often cost-prohibitive. Few studies have focused on achieving affordable railroad object detection using only a low-cost single camera with CV techniques. Existing affordable single-camera CV techniques [3][4] are primarily designed for tracking rails rather than detecting objects and people for video surveillance purposes on railroads. Our research aims to address this gap by demonstrating how a low-cost single camera can be effectively used for real-time detection of both objects and people on the tracks.

We address solutions for challenges inherent to railway surveillance, such as the variability of lighting conditions and the complexity of the scenarios on the tracks. In addition, the challenge of efficiently processing the large volume of data generated by cameras is significant. Our approach provides solutions for integrating highly efficient image processing algorithms capable of discriminating between false alarms and real risk situations. This integration not only improves the accuracy in detecting dangerous situations, but also optimizes the use of surveillance resources, resulting in a safer and more efficient operation of railway facilities.

Additionally, this study conducts a review of current approaches in existing rail surveillance installations, identifying their limitations in terms of operational costs and effectiveness. Recognizing these limitations, we have designed a system that improves the detection and classification of objects and people and effectively integrates into existing rail surveillance operational requirements. Our methodology integrates existing deep learning technology with computer vision techniques, into the constraints and operational needs of a real rail facility. The proposed solution maintains a balance between efficiency and affordability, making it a viable and attractive solution for a wide range of railway applications.

II. State of the Art

In the literature there are different proposed techniques to detect people or objects on the railroad track platforms, being relatively expensive and difficult to maintain technology (in comparison to our solution). S. Oh et al. [5] propose using multiple cameras perpendicular along the track to monitor almost the entire length of the track line of the platform. The author divides the line monitoring process into two parts: detection of train status and detection of objects and people on the track.

Other approaches [6][7] that incorporate different sensing modalities, such as LiDAR, laser or remote sensing data, have been developed to monitor pathways.

T. Xiao et al. [8] have developed a non-contact multisensory technology detection technique. In particular, they have deployed an On-Board obstacle detection device based on a camera and a LiDAR to detect obstacles in the monitored area in real-time, and to determine whether the train should brake automatically, or the train pilot should brake manually.

S. Taori et al. [9] have gone one step farther by suggesting the implementation of a multisensory barrier composed of infrared (IR) and ultrasonic (US) sensors, in addition to a CV system, in order to alert the surveillance system about the presence of obstacles on the train track and thus prevent possible accidents.

In recent years, various not as expensive methods have been developed to track the rails. Note that this problem does not coincide exactly with our surveillance problem of detecting falling objects and people on the tracks. For example, F. Kaleli and Y. S. Akgul [10] proposed an algorithm based on dynamic programming to track the trajectory of the front part of the train on the railroad. This method consists of three steps: first, a Sobel operator is used to identify the borders of the input image, then a Hough transform is applied to the binary image to detect the railroad line, and finally dynamic programming is used to efficiently track the rail lines.

Y. Wang et al. [11] developed a neural network approach called RailNet, which includes a segmentation network to track the rails, trained with their own railroad segmentation dataset, that is, a collection of annotated images specifically designed for identifying and delineating various elements of railroad infrastructure. This approach also includes a feature extraction network and a segmentation network.

M. Ghorbanalivakili et al. [12] proposed a rail path extraction process in which the pixels of the left-right rails of each path are extracted and associated using a convolutional architecture called TPE-Net. This net has two different regression branches to get the locations of the center points of each rail and generate the possible train routes (called “ego-routes”). The experimental results indicate that this technique has a high accuracy and recovery of true positives.

Recently authors such as M. Qasim Gandapur and E. Verdú [13], M. Adimoolam et al. [14] and A. D. Petrović et al. [15] have proposed a combination of Convolutional Neural Network (CNN)s based on YOLO-v5 [16] implementation, for object detection and tracking with CV. The later also includes a Canny edge detection and the Hough transform.

Another technique named Mask RCNN [17] employs a pyramidal structure to obtain high performance in the task of instance segmentation on the COCO dataset [18].

This last approach is similar to ours, but it differs in that our research does not focus on detecting traffic signals and track bifurcation, but on detecting falling objects and people on the tracks.

H. Pan et al. [19] propose a multi-task learning network that segment, detects, and classifies the rail lines. This approach makes improvements over other multitask networks such as paying more attention to accuracy than to recall. Segmentation aids in improving the classification results. The railroad detection algorithm in the multitask network can effectively deal with the blocked track line problem, and it avoids the disadvantage of segmentation network in detecting tracking lines with thinner and smaller pixel ratio. Its anchorless designs avoid the problem that exits when the size of the anchor frame is not suitable for small targets.

Recently at London Underground’s Docklands Light Railway (DLR), experimentation is underway with the development of the CCTV AI Trial project [20]. The alert system has been deployed to prevent accidents on the network lines and strives to minimize false alarms. This pilot project is in the testing phase.

In the field of low-light image processing in railway driving, the paper presented by Z. Chen et al. [21] describes an innovative network designed to optimize visual clarity in low-light conditions. They describe a progressive enhancement strategy and a lightweight network overcoming the limitations of conventional methods, delivering remarkable results that outperform previous techniques such as Zero-DCE++, SCI, RetinexDIP, and RUAS. With its efficient structure combining advanced feature extraction operators and accurate encoder-decoder architecture, these authors enhance image analysis for railway applications, contributing to the safe and reliable operation of trains in low-light conditions.

The work of C. Meng et al. [22] proposes SDRC-YOLO as an enhancement of the YOLOv5s algorithm, specifically designed to identify infiltrations in railway scenarios. It incorporates a Hybrid Attention SSA mechanism that combines a Spatial Attention Module (SAM) with Squeeze and Excitation Network (SENet) channel attention. The structure features a DW-Decoupled Head for efficiency and employs Large Convolutional Kernels from RepLKNet with strong

parameterization to create wider perceptual fields. Additionally, the lightweight universal resampling operator CARAFE is used to select more suitable sizes and proportions for infiltration features.

Recent advances in railway track segmentation algorithms have shown remarkable improvements in intrusion detection and overall security. A prominent proposal is ERTNet, an efficient railway track region segmentation algorithm based on a lightweight neural network and cross-fusion decode as described by Z. Chen et al. [23]. This network incorporates encoder-decoder architecture, using depth convolutions and cross-fusion to efficiently integrate shallow and deep features, achieving high accuracy with a lightweight model. They archive a MIoU (Mean Intersection over Union) of 92.4% with minimal computational resource requirements. ERTNet represents an advance in railway surveillance technology, and aligns with our system objectives to improve real-time and ensure the integrity of railway infrastructure.

Another significant development in railway region segmentation is the LRseg [24] model, designed to optimize the detection of foreign objects on tracks. This model incorporates a lightweight coding approach and a self-correcting decoder together with a segmentation head, enabling efficient, real-time processing, crucial for applications in on-board devices. With its low parameter requirements and its ability to operate efficiently on both embedded systems and powerful personal computers, LRseg represents a new contribution to railway image segmentation, delivering accurate and fast results, indispensable for railway safety and maintenance.

Another important advance in object detection on railway lines is the development of the RailFOD23 [25] dataset, specifically designed to improve automated detection of foreign objects such as plastic bags, flying objects, bird nests and balloons. This dataset includes 14,615 detailed annotated images generated using artificial intelligence techniques. Therefore, it represents a valuable resource for training object detection models, with direct applications in railway safety. This dataset promises to facilitate significant advances in railway surveillance technology, optimizing detection and response to potential threats in the power transmission infrastructure.

Despite advancements in railway surveillance, current systems exhibit significant limitations that justify the development of our proposed system. First, most systems rely on multiple cameras or expensive technologies like LiDAR, increasing complexity and implementation costs. Second, many of these systems are not optimized to operate in real-time on standard computing equipment, limiting their applicability in conventional railway environments. Additionally, the accuracy in detecting objects and people on tracks is often limited, leading to a high number of false positives and negatives. These limitations highlight the need for a more efficient and economical system, such as the one we propose, which uses a single low-cost camera and advanced algorithms for effective and real-time surveillance.

III. Materials and Methods

Our real-time railway surveillance system employs a two-step detection process involving background subtraction and object classification using a convolutional neural network (CNN), specifically designed to detect people and objects on the tracks. By means of background subtraction it identifies changes in the scene to flag potential intruders or left objects, while the CNN classifies the detected objects to minimize false positives.

The MOG2 algorithm is used for background detection due to its capacity to extract dynamic backgrounds and its ability to recover the background state efficiently, which is crucial for the fast-moving environment of a railway. It distinguishes between the background and the moving objects effectively, even for small or slow-moving items. MOG2 has a high processing speed, thanks to the CUDA[1] implementation.

The YOLO-v5 neural network model was selected for object classification because of its high accuracy and speed, vital for real-time processing. It identifies and classifies objects as either trains or people on the tracks, contributing to the alarm system's accuracy.

The system is implemented on standard PC hardware, achieving an average frame rate of 15 FPS with a latency of 0.54 seconds per frame, ensuring it meets real-time operation criteria. The alarm system distinguishes between actual threats and non-threats, issuing alerts for the security personnel to act accordingly.

Our system is divided into two detection steps, the background subtraction and the classification of objects in images through a CNN, integrated to form a cohesive system. This methodological approach is crucial to understanding the innovations and results of the study.

To evaluate which background subtraction technique best suits the needs of the system, several tests and iterations have been performed with the latest techniques. In particular, we have compared in terms of accuracy and image processing speed several algorithms, namely: MOG2 [26] (Gaussian Mixture Based Background/First Plane Segmentation Algorithm), GMG [27] (Combination of statistical estimation of the background image and Bayesian segmentation per pixel), KNN [28] (Nearest Neighbors), and CNT [29] (Count Based). Regarding image classification and based on the study of the existing literature and our preliminary tests of different neural networks that perform this task, YOLO-v5 has been chosen. YOLO-v5 has also been compared with its previous versions regarding accuracy and speed for image processing.

Regarding the evaluation of the background subtraction algorithm, we have measured its accuracy, i.e., how many errors the algorithm makes when identifying the background in an image or video. This has been done by comparing the results of the algorithm with a ground truth, i.e., reference images/videos that have already been manually labeled as the correct background.

Another aspect to evaluate the background subtraction algorithm is to measure its processing speed, i.e., how long it takes to process an image or video frame. This is important because real-time performance is required.

We conducted a series of tests to validate the efficiency of the system. These tests included varied scenarios, from low-light conditions to the presence of moving objects at different speeds. We continuously adjusted the algorithm's parameters to optimize its performance, paying particular attention to minimizing false positives and improving real-time detection.

This iterative process of testing and fine-tuning was key to adjusting the system to the specific conditions of the train tracks, ensuring effective and reliable surveillance.

To ensure a thorough understanding of our algorithm, below we provide a more detailed description of its components. Initially, the background subtraction stage employs the MOG2 algorithm, selected for its ability to effectively adapt to dynamic changes in the environment. This is a fundamental feature in the fast-moving railway context. This process facilitates the initial identification of potentially dangerous objects and people on the tracks. Subsequently, object classification is performed using the pre-trained YOLO-v5 model, which allows for a clear distinction between different types of objects

[1] CUDA (Compute Unified Device Architecture) Provides a development environment for creating high-performance GPU-accelerated applications. It includes GPU-accelerated libraries, debugging and optimization tools, a C/C++ compiler and a runtime library.

and checks that alerts are only issued for true dangerous situations. This workflow is illustrated in Fig. 1, which provides a simplified visual representation of the process from image capture to alert generation.

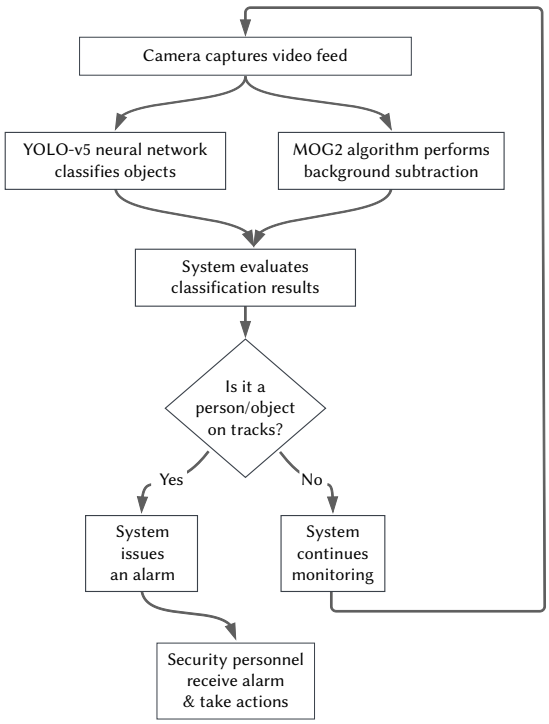


Fig. 1. Functional Diagram.

For the YOLO-v5 algorithm there exists comparisons and evaluations with different metrics already published[2]. As can be seen in the AUC-ROC[3] curve plot in Fig. 2, the YOLOv5x6 model with the YOLOv5x6.pt weights previously trained from pytorch.org/hub, is the one that shows the best performance in speed and mAP[4] [30] on the COCO [18] dataset of 5000 images at various inference sizes from 256 to 1536.

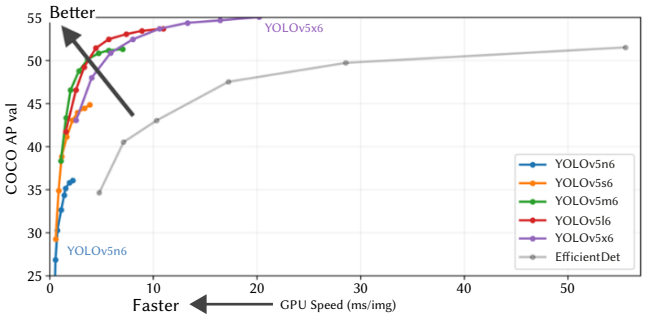


Fig. 2. Metrics of different weights in the COCO val2017 dataset.

During implementation, we faced significant technical challenges. One of the main challenges was the variability of environmental conditions affecting detection accuracy. To overcome this, we implemented adaptive algorithms that adjusted detection parameters based on lighting and weather conditions.

Another challenge was differentiating between irrelevant objects and real threats. To address this, we enhanced the algorithm's classification capability, teaching it to recognize a wider range of objects and situations. This improvement resulted in a significant reduction in false positives.

These adjustments and enhancements were crucial in ensuring the system's effectiveness and reliability in a real operational environment.

For the evaluation of the classification, the selection criteria for the videos took into account the limited availability of material, as these videos are from the actual operation of Metro de Madrid S.A. (the company in charge of Madrid's underground railway). We prioritized videos with a variety of environmental conditions, such as different lighting levels, to test the robustness of the system. In addition, videos were selected that included various types of objects and people on the tracks, ensuring that they did not reflect serious accident situations and various configurations of the camera installation, height, angle, etc. This selection allowed us to evaluate the effectiveness of the system in a wide range of realistic scenarios.

To ensure the reproducibility of our study for other researchers, we have made available the source code, scripts and the video dataset used in GitHub[5] and Zenodo[6] repositories respectively.

In this study, special emphasis was placed on the selection and preparation of the data set, which is crucial for the evaluation of our system.

The dataset used in our research comes from various video hosting platforms. This approach ensures a varied and realistic representation of the situations that could be encountered on railroad tracks. During the preprocessing and cleaning process, we focused on removing redundant elements, such as repetitions and labels, to optimize the data for analysis. This includes the removal of audio, thus improving the quality of the dataset. In addition, we have taken rigorous measures to comply with copyright regulations, ensuring respect for intellectual property and proper citation of all sources used.

As for the labeling of the data, a manual classification of the videos was carried out, from those presenting accidents of greater to lesser simplicity, highlighting the effort and meticulousness in the labeling process. However, we faced several challenges and limitations with this dataset. Critical aspects such as inherent biases, imbalance in class distribution, the presence of noise in the data, and the scarcity of videos related to the specific subject matter of the study were discussed. In addition, special emphasis was placed on ethical considerations, such as privacy, consent to data use, and anonymization, to ensure a responsible and ethical approach to our research.

A. Metrics Used for Background Subtraction

To evaluate background subtraction, the number of non-zero pixels per frame has been taken as a metric, because it indicates the effectiveness of the algorithm in detecting moving objects. A high number of non-zero pixels indicates that the algorithm has been able to effectively detect moving objects and separate them from the static background. On the other hand, a low number of pixels different from zero, indicates that the algorithm has had difficulty separating the moving objects from the static background.

To evaluate the background subtraction algorithm, it is necessary to measure its processing speed, since processing time efficiency is vital, especially in applications that require fast response, such as in railway surveillance systems. Different algorithms can have significant variations in their execution times, which directly affects the practicality of their application in real-time environments.

[2] Source code available at https://github.com/ultralytics/yolov5

[3] AUC-ROC (Area Under the Curve - Receiver Operating Characteristic) is a metric used to evaluate the ability of a model to discriminate between positive and negative classes.

[4] mAP: Mean precision (mAP) is the average of all classes of the maximum precision for each object at each recall value.

[5] https://github.com/Domy5/Rail_Surveillance.

[6] https://doi.org/10.5281/zenodo.8357129

Therefore, a balance between detection accuracy and processing speed is essential in choosing the most suitable algorithm for our system.

Our system leverages the processing capabilities of GPUs, where available, to significantly accelerate real-time video analysis. By relying on YOLOv5, the object detection model, we are guaranteeing current state of the art efficiency and speed. This model, together with the MOG2 background subtraction technique, allows the system to operate efficiently and with low energy consumption even on standard PC hardware. In addition, the choice of inexpensive cameras already installed for data collection not only makes our solution cost-effective, but it also contributes to the reduction of energy consumption, a key factor in continuous surveillance systems.

B. Metrics Used for Object Classification

In the evaluation of our object classifier model, we have employed the confusion matrix, a standard tool in the analysis of classification models. This matrix allows us a detailed understanding of the effectiveness of the model, breaking down the results into TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative). To provide a more complete view of model performance, we have supplemented these metrics with precision, sensitivity (or true positive rate) and specificity (or true negative rate). These additional measures help us evaluate the model's ability to correctly identify threats while minimizing false alarms, a crucial aspect of effective rail surveillance systems.

C. Global Metrics

The metric we used to evaluate the computational performance of the complete system (both the object classifier and the background subtraction technique) is the processing time latency of each image, which determines if it is appropriate for use in a real-time system.

We also analyzed the overall effectiveness and viability of the system through direct tests with the dataset videos. In particular, we reviewed the overall system performance and verified the result of the alarms and pre-alarms complied with the expected result.

IV. Alarm Policy

In this section, we evaluate the system while it is operating on a regular basis, without incidents that affect the tracks, or that are not triggered by false negatives. These tests were performed to ensure that the system does not generate false alarms or unnecessary pre-alarms. This is an important measure to ensure the reliability of the system and to avoid saturation or desensitization of operators to real alarms.

Our system's alarm policy has been evaluated to maximize accuracy in identifying real risk situations while minimizing disruptions caused by false positives. We have developed an innovative approach that distinguishes between alarms and pre-alarms based on the severity and probability of the detected threat. Alarms are triggered when a person or object on the tracks is identified with high certainty, while pre-alarms are issued in situations of lower certainty, allowing for additional verification prior to raising the alarm. This strategy ensures a rapid and effective response to real threats and significantly reduces unnecessary operational disruptions. Below we describe the alarm policy and the logical implementation decisions for activating or deactivating alarms.

A. Alarm per Person on Track Platform

One of the challenges of our solution, using the existing CCTV cameras, is to determine if a person is on the platform or on the road, that is, if a person is within what corresponds to the ROI.

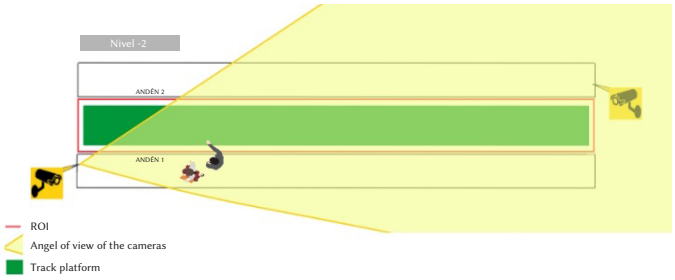


Fig. 3. Perspective of the camera with respect to the track platform.

As shown in Fig. 3, the location of the cameras is at the end of each platform, with an angle that focuses on an oblique angle to the end of the opposite platform. For this reason, the cameras generate images with a perspective that causes the occlusion of a certain part of the track platform by people near the edge of the platform where the camera is located.

If a person is located at the edge of the platform, the object detection system will generate a BBox o bounding box that will cross the area of interest or ROI as shown in the example in Fig. 4, being a false alarm, as the person is not on the track platform, but on the platform.

Fig. 4. BBox generated by object detection system.

To prevent the occlusion caused by people at the edge of the platform, with respect to the ROI, we decided to generate a point as far away as possible, this being at the bottom right of each BBox. This point corresponds to the green point in left foot of the person facing the camera in the example in Fig. 5. This point will serve as our reference to activate the alarm per object inside the ROI, and so emitting an alert sound. This means that when the reference point is inside the polygon that represents the ROI, the person on the track alarm will be triggered.

Fig. 5. Point generated at the bottom right of each BBox (left foot of each person).

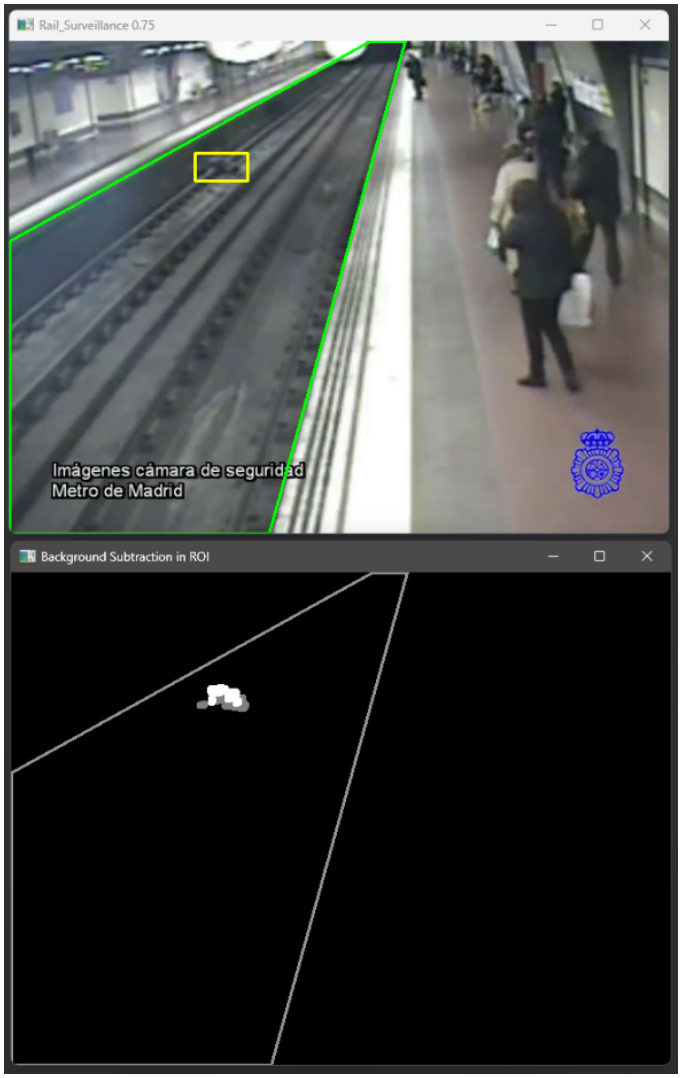


Fig. 6. Detection by background subtraction within the ROI.

B. Pre-alarms Due to Detecting Movements on the Railroad

The MOG2 background subtraction algorithm generates pre-alarms when a consecutive series of frames has a significant number of non-zero pixels in the contour mask in Fig. 6. The pre-alarm emits an alert sound different and softer than the alarms. Background subtraction is performed only within the ROI of the mask in Fig. 6, which means that the movements on the platform are ignored. This method detects events such as falling people, bulky objects, flash flooding of the track basin, vault detachment, etc.

C. Deactivating Alarms on Train Arrival

Once the arrival of the train is detected, the pre-alarms are deactivated, but not the alarms generated by person on the track platform (i.e., people within ROI). This type of alarm, in addition to emitting the alert sound, will display the operator the message: "*Arrival of the train with person on the train track*", as shown in Fig. 7.

V. Results

In this section we describe and analyze the results of the two parts of the system separately: background subtraction and object classification. Subsequently the whole system is evaluated.

To evaluate the energy efficiency and fast processing of our system, we have performed benchmark tests using a standard PC configuration

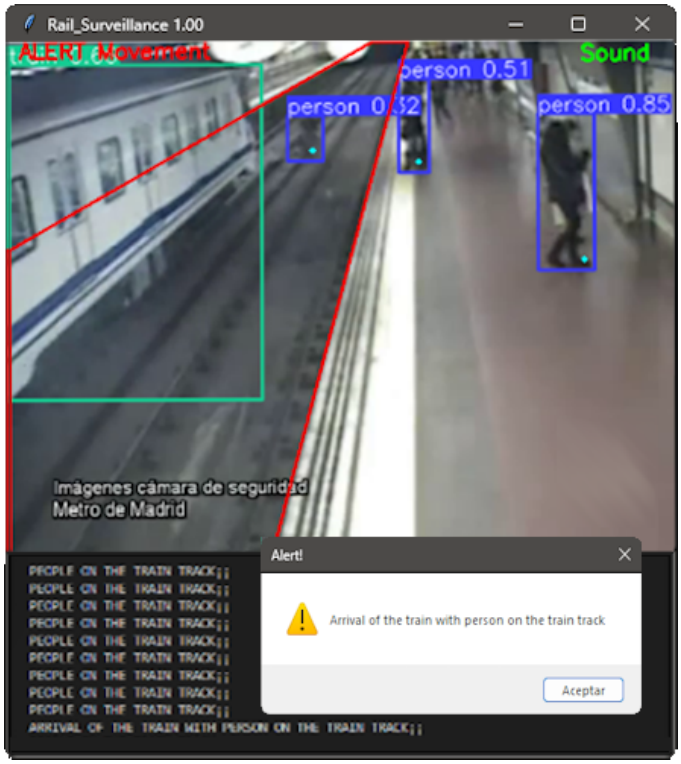


Fig. 7. Person on track message.

without additional specialized hardware. The results showed that the system maintains an average of 15 FPS and a latency per frame of 0.54 seconds, meeting real-time operation expectations and excelling in processing efficiency. This performance is due to the optimized system design and strategic use of GPU technology to accelerate computer vision tasks.

A. Evaluation of the Background Subtraction Algorithms

To evaluate the background subtraction algorithms, we compared the accuracy of the CNT, KNN, GMG and MOG2 algorithms. One of the main differences between them is that MOG2 uses an adaptive Gaussian blending model, while GMG uses a global Gaussian blending model. This means that MOG2 can adapt dynamically to changes in the scene background, allowing for greater accuracy in background identification and better performance in situations with a dynamic background. In comparison, the GMG algorithm uses the first frames of a video to build a model of the background, making it less adaptable to changes in the scene. Another important difference is that MOG2 is able to handle foreground object overlaps, so it is able to identify slowly moving objects in the scene. We have found that the GMG algorithm produces more artifacts and noise than MOG2. Therefore, there are significant differences between them that make MOG2 better than GMG to detect people moving slowly in dangerous situations.

CNT is based simply on pixel count, KNN only takes into account the distance between pixels and may work less accurately in situations with a dynamic background. MOG2, compared to KNN, is more flexible, as it allows easier adjustment of its parameters. Therefore, MOG2 achieves higher accuracy in identifying small objects.

We have tested several configurations in the parameters that let us modify the behavior of each subtraction algorithm and, in those subtraction algorithms that permit it, the application of the morphological aperture kernel to find the best result with respect to noise.

Fig. 8 shows the number of non-zero pixels per frame, which indicates that the algorithm has been able to effectively detect moving objects and separate it from the background. The train arrival occurs

at frame 740, each element or line shows the values of the number of pixels different from zero detected by each subtraction algorithm. As can be seen in the figure, the results are very similar in terms of number of pixels with the CNT and KNN algorithms (dark blue and yellow line, respectively). In both algorithms it takes many frames to recover the background state, which can lead to false positives events. GMG (gray line) has strong constant fluctuations that generate a lot of noise in the mask. The higher stability and speed in image changes can be observed in the MOG and MOG2 algorithms (light blue and orange line).

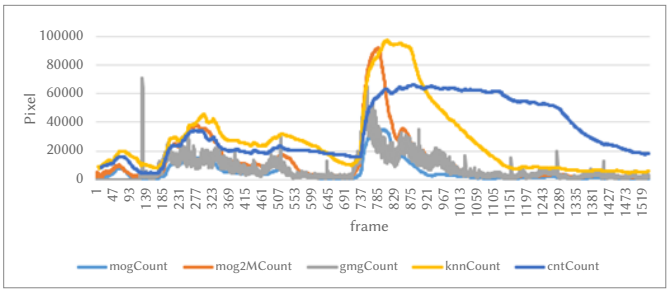


Fig. 8. Number of non-zero pixels per frame.

Fig. 9 shows the frame processing time of each algorithm. The best performance is achieved by the CNT algorithm with 0.0010 seconds average time, followed by MOG2, with 0.0032 seconds average time, being MOG2 more stable than KNN (0.005 seconds), GMG (0.015 seconds), and MOG (0.007 seconds), whose execution times double the value of MOG2.

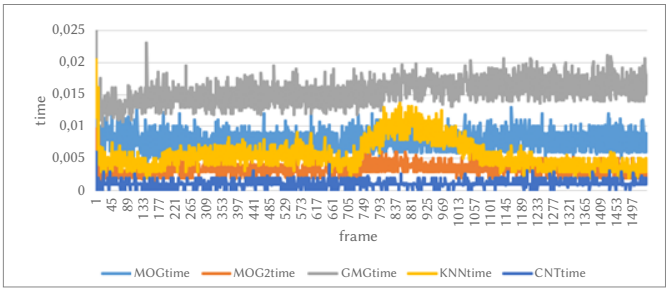


Fig. 9. Processing time per frame.

B. Object Classifier

To evaluate the object classifier, several of the videos available in the test dataset were chosen. The frames were labeled manually, then we executed the classifier to get the results for the "Train object" and "Person ontrack" event used to generate the confusion matrix in Table I and Table II.

TABLE I. Confusion Matrix of the "Train Object" Event

Object: TRAIN		Prediction	
		Positives	Negatives
Real	Positives	792	4 (Type II error)
	Negatives	8 (Type I error)	732

Next, several metrics are extracted from the confusion matrix for the "Train object" event in Table I. In particular, the precision is 99%, the accuracy is 99.21%, the specificity is 98.91%, the Recall or sensitivity is 99.49%, and the F1 score is 99.24%.

Overall, these values indicate that the classifier model performs well in classifying the "Train object" classification dataset.

TABLE II. Confusion Matrix of the Object "Person on Track" Event

Object: Person on via		Prediction	
		Positives	Negatives
Real	Positives	29	504 (Type II error)
	Negatives	0 (Type I error)	1003

Table II shows the confusion matrix for the "Person on track" dataset. In particular, the precision is 100%, the accuracy is 67.18%, the specificity is 100%, the Recall or sensitivity is 0.05%, and the F1 score is 0.10%.

These results indicate that the model has limited performance in classifying objects of the "Person on track" class. Usually this happens because the model tends to be over-fitted to the "Person on track" class, resulting in high accuracy, but limited generalization performance.

C. Overall Evaluation of the System

The first evaluation reported in Fig. 10 shows the latency inference per frame using eGPU with the test dataset.

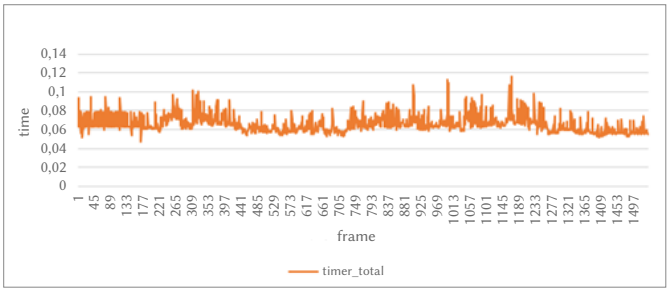


Fig. 10. Per-frame inference latency with eGPU.

The Fig. 11 shows the FPS data achieved by the system.

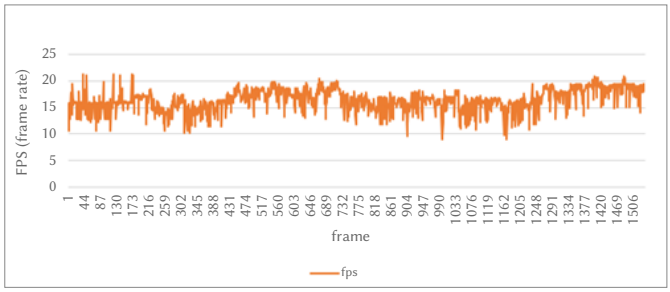


Fig. 11. FPS achieved with eGPU.

The results of the total frame time latency evaluations by inference indicate that the video surveillance system operates correctly for most of the video fragments analyzed, maintaining acceptable performance when processing incoming data with latency not exceeding 100 milliseconds. This confirms that the system meets the established standards to operate efficiently within the time limits required for real-time applications.

Finally, the system was evaluated in several video fragments in normal exploitation that is without events. In these fragments, false alarms do not occur. This means that the system is working properly and meets the operational expectations.

Our system has been rigorously evaluated to ensure its effectiveness and efficiency in a real operating environment. The accuracy of our background subtraction algorithm was validated by comparing the number of non-zero pixels per frame, revealing accurate detection of moving objects against the static background. This analysis

demonstrated that the MOG2 algorithm significantly outperforms its competitors, adapting efficiently to variations in the environment and maintaining a low number of false positives.

In terms of object classification, the YOLO-v5 model has demonstrated exceptional accuracy and speed, critical for real-time detection. Metrics derived from the confusion matrix, such as accuracy (the ratio of actual positive identifications to total predicted positive identifications), sensitivity (the model's ability to correctly identify true positives), and specificity (the ability to avoid false positives), indicated that our system effectively minimizes false alarms while maintaining a high rate of correct detection.

The overall evaluation of the system, combining both background subtraction and object classification, was performed through per-frame processing latency and alarm policy compliance. The results indicated that our system manages to maintain an average of 15 FPS and a per-frame latency of 0.54 seconds, thus ensuring real-time operation. The alarm policy, designed to differentiate between real and potential threats, was validated in diversified test scenarios, demonstrating the ability to reduce unneeded interruptions and focus the attention of security personnel on critical events.

D. Practical Applications and Use Cases

Our results have not only demonstrated the technical feasibility of an advanced surveillance system for real-time detection of people and objects on train tracks, but also highlighted its practical importance through experiments in real-life situations. Below, we present two scenarios that illustrate the system's ability to significantly improve railway safety:

Early Detection of Objects on the Track: In a hypothetical incident, our system detected a piece of heavy machinery inadvertently left on a curved section of track, where direct visibility is nil. The immediate detection of the object by our system allowed the control center to be alerted well in advance. This preventive action facilitated the halting of the approaching trains, avoiding a potential serious accident. This case highlights the importance of the system's ability to detect unexpected objects, especially in areas of difficult visibility.

Vandalism Prevention: In another scenario, a group of individuals was identified accessing the tracks at night with the intention of vandalizing parked cars. Early detection of the anomalous movement and accurate classification of 'persons' in the track area triggered security alarms. The rapid response of the security team, guided by the system's alerts, prevented the act of vandalism, thus ensuring both the safety and the security of the railcars.

VI. Discussion and Comparison

CV surveillance effectively automates and improves efficiency in the monitoring of the railroads.

The fusion of the two types of algorithms within our solution plays a vital role in maximizing real-time processing and energy efficiency. By optimizing the MOG2 background subtraction algorithm and the YOLOv5 detection technique to operate together seamlessly, the proposed system achieved a significant reduction in computational resource consumption. This optimization is reflected in the system's ability to process images in real time without compromising accuracy, even on equipment with limited capabilities. The complete source code and project documentation are publicly available on GitHub, allowing other researchers and developers to explore and contribute to the evolution of this railway surveillance system.

Our system uses an inexpensive regular camera per monitored area and object recognition to detect and analyze patterns in images and videos, enabling monitoring tasks to be performed with high precision and a reduced number of false alarms. This is an operational innovation with respect to the current process performed by human operators. Furthermore, automated video surveillance is less error prone than human supervision, as it can operate continuously without any fatigue.

In this research, we have described a system for real-time detection of people and objects on the railroad using modern image processing, object classification and background subtraction techniques, with an architecture based on YOLO-v5 and the MOG2 algorithms, respectively.

Compared to other object detection algorithms, YOLOv5 has proven to have high accuracy in object detection in different types of images and videos. It also has a high processing speed, thanks to CUDA implementation, which makes it suitable for real-time applications, flexible and easy to use.

In our experiments, the MOG2 algorithm has generated enough mean time-per-frame processing speed to process the video continuously in real time in a regular PC, showing little artifacts or noise and a higher recovery by dynamically adapting to changes in the scene background, due to the adaptive Gaussian blending technique used. Furthermore, we have found that the MOG2 algorithm achieves the higher accuracy in identifying small objects and, according to the number of non-zero pixels in the experiments, more stability and speed in image changes than the other algorithms evaluated.

Regarding image classification, the accuracy metric for the "Person on track" event detection and the accuracy metric for the "Train object" event detection are more than acceptable for the correct operation of the system on a daily basis, as verified in the experiments, which means that the model is predicting correctly every time it detects those objects.

Our research contribution focuses on combining the most advanced current techniques, such as MOG2 background subtraction and YOLO-v5 neural networks. By leveraging the latest advancements in these algorithms, we aim to significantly enhance the effectiveness and efficiency compared to other similar systems.

S. Oh et al. [5] employ multiple cameras, combining frame difference information, thresholding, labeling, and fusion with laser sensor detection. In contrast, our approach differs by not requiring expensive components, utilizing only a single standard camera per platform.

On the other hand, the work of T. Xiao et al. [8] and A. D. Petrović et al. [15] describes various approaches to detecting obstacles on the rails. They propose a non-contact multisensory method using cameras and a LiDAR device, as well as a combination of deep neural networks and edge detection methods with cameras. Both proposals aim to detect obstacles on the tracks using on-board devices. In contrast, our research focuses on detecting objects and people on the platforms.

Finally, the work of H. Pan et al. [19] focuses on railroad signal detection and issues related to blocked track lines. Their approach differs from ours as it emphasizes track line detection and utilizes a combination of multi-task learning networks.

In general, our proposal is affordable and more flexible in generating alarms and pre-alarms. This flexibility allows for the easy implementation of new alarms within the surveillance logic. It is scalable and can launch multiple instances for each camera in a single PC.

In short, our solution prioritizes railroad safety, particularly the safety of people. Unlike other approaches, our system is specifically designed for object and person detection on the railroad. Additionally, our research shows that our solution can be executed in real time on a standard PC.

VII. Conclusions and Future Research Directions

The performance of these experiments, as presented in this study, includes several limitations that are crucial to contextualize the results obtained.

The effectiveness of artificial intelligence models is intrinsically linked to the diversity and quantity of data used for training. In this case, the data may be limited in terms of scenario variability, weather conditions and types of obstacles, which could affect the generalization of the model in real situations. These surveillance systems operate in a dynamic environment where conditions can change drastically, as in the case of lighting conditions. Such variations can affect detection accuracy, as the system may not have been exposed to all possible conditions during the training phase.

Effective implementation of surveillance systems in real-world environments requires seamless integration with existing railway control infrastructure. This aspect can present significant technical challenges, especially in older systems or those that are not designed to integrate with AI-based solutions. Furthermore, our ongoing research into the use of specialized hardware technologies, such as low-consumption FPGAs and ASICs, for future iterations of the system aims to substantially reduce electrical consumption. These efforts are not only in line with enhancing the operational efficiency of rail surveillance systems but also contribute significantly to the sustainability and environmental impact of rail transportation. The development of energy-efficient surveillance solutions will be a cornerstone of our future research efforts, ensuring that technological advancements in railway surveillance align with ecological responsibility.

In addition, it is crucial to maintain a careful balance between ensuring security and respecting individual privacy. The deployment of advanced surveillance technologies must be managed with careful consideration of ethical implications and the protection of personal freedoms.

Another major challenge is the ability of the system to efficiently monitor a large number of cameras simultaneously. Railway surveillance often requires an extensive network of cameras to cover all relevant areas. The increase in the number of cameras poses challenges in terms of real-time data processing and analysis, which can impact the speed and accuracy of detecting objects and people on the tracks.

The described research highlights the feasibility of implementing a real-time railroad track surveillance system using affordable technology and advanced image processing algorithms. The challenges faced range from variations in environmental conditions to the ability to differentiate between relevant and non-relevant objects. These limitations have been overcome by iterative adjustments to the algorithms, demonstrating the adaptability and robustness of our approach. However, we found certain limitations inherent to our study, particularly in terms of the availability of railway surveillance datasets for training, which may influence the generalization of the model to all possible practical situations. Future research will benefit from a larger and more diverse dataset, allowing a more robust generalization of the model. In addition, integration with existing railway control systems presents a fertile field for further exploration, which seeks a more holistic and automated implementation of real-time monitoring.

In the context of advances in rail surveillance, it is imperative to look to the future and identify key areas where further research can lead to significant improvements in the safety, efficiency, and sustainability of rail transportation. Despite progress in applying computer vision techniques and neural networks, there are substantial opportunities to expand and enrich this field. Future research could explore the implementation of an intuitive and efficient user interface for railway surveillance systems. This includes customizable dashboards, real-time alerts, and advanced data visualizations that enhance monitoring and decision-making. Utilizing user-centered design techniques and incorporating operator feedback will be crucial in developing solutions that are not only technically advanced but also accessible and easy to use for surveillance personnel.

Another important area for future research is the integration of rail surveillance with urban transportation systems. This approach would allow for more efficient traffic management and improved coordination during emergencies or service disruptions. Research should focus on developing standardized and secure communication protocols that facilitate real-time information exchange between different modes of operation, such as passenger, maintenance, and freight. This data exchange could significantly improve emergency response, optimize resource allocation, and minimize service interruptions.

References

[1] S. Glimne, R. Brautaset and C. Österman, "Visual Fatigue During Control Room Work in Process Industries," vol. 65, no. 4, pp. 903-914, doi: 10.3233/WOR-203141. PMID: 32310219; PMCID: PMC7242839. 2020.

[2] Y. Lei, T. Tian, B. Jiang, F. Qi, F. Jia, Q. Qu, "Research and Application of the Obstacle Avoidance System for High-Speed Railway Tunnel Lining Inspection Train Based on Integrated 3D LiDAR and 2D Camera Machine Vision Technology," *Applied Sciences*, vol. 13, no. 13, p. 7689, 2023.

[3] M. Li, B. Peng, J. Liu and D. Zhai, "RBNet: An Ultrafast Rendering-Based Architecture for Railway Defect Segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-8, 2023.

[4] R. Goel, A. Sharma and R. Kapoor, "An Efficient Object and Railway Track Recognition in Thermal Images Using Deep Learning," *Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS* 2021 (pp. 241-253). Springer Singapore. 2002.

[5] S. Oh, S. Park and C. Lee, "A platform surveillance monitoring system using image processing for passenger safety in railway station," ICCAS 2007 - *International Conference on Control, Automation and Systems*, pp. 394-398, January 2007.

[6] M. Arastounia, "Automated recognition of railroad infrastructure in rural areas from LIDAR data," *Remote Sensing*, vol. 7, no. 11, 2015.

[7] B. Le Saux, A. Beaupère, A. Boulch, J. Brossard, A. Manier, and G. Villemin, "Railway detection: From Filtering to segmentation networks" *Proc. IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, Spain, Jul. 2018.

[8] T. Xiao, Y. Xu, and H. Yu, "Research on Obstacle Detection Method of Urban Rail Transit Based on Multisensor Technology," *Journal of Artificial Intelligence and Technology*, vol. 1, no. 1, pp. 61-67, 2021.

[9] S. Taori, V. Kakade, S. Gandhi, and D. Jadhav, "Multi Sensor Obstacle Detection on Railway Tracks," *International Research Journal of Engineering and Technology*, vol. 8, no. 3, pp. 826-829, 2021.

[10] F. Kaleli and Y. S. Akgul, "Vision-based railroad track extraction using dynamic programming," *Proceedings of 12th International IEEE Conference on Intelligent Transportation Systems*, St. Louis, MO, USA, oct. 2009, pp.1-6.

[11] Y. Wang, L. Wang, Y. H. Hu, and J. Qiu, "RailNet: A Segmentation Network for Railroad Detection," *IEEE Access*, vol. 7, pp. 143772–143779, 2019.

[12] M. Ghorbanalivakili, J. Kang, G. Sohn, D. Beach and V. Marin, "TPE-Net: Track Point Extraction and Association Network for Rail Path Proposal Generation," 2023 IEEE 19th International Conference on Automation Science and Engineering (CASE), Auckland, New Zealand, 2023, pp. 1-7, doi: 10.1109/CASE56687.2023.10260541.

[13] M. Qasim Gandapur, and E. Verdú, "ConvGRU-CNN: Spatiotemporal Deep Learning for Real-World Anomaly Detection in Video Surveillance System," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 4, pp. 88-95, 2023.

[14] M. Adimoolam, S. Mohan, and G. Srivastava, "A novel technique to detect and track multiple objects in dynamic video surveillance systems," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 4, pp. 112-120, 2022.

[15] A. D. Petrović, M. Banić, M. Simonović, D. Stamenković, A. Miltenović, G. Adamović, and D. Rangelov, "Integration of Computer Vision and Convolutional Neural Networks in the System for Detection of Rail Track and Signals on the Railway", *Applied Sciences*, vol. 12, no. 12, p. 6045, 2022.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Jun. 2015, doi: 10.48550/arXiv.1506.02640.

[17] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. 2017 IEEE International Conference on Computer Vision (ICCV),* Venice, Italy, 2017, pp. 2980-2988.

[18] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context", in: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) *Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*, vol 8693, Springer, Cham, 2014, pp. 740-755.

[19] H. Pan, Y. Li, H. Wang, and X. Tian, "Railway Obstacle Intrusion Detection Based on Convolution Neural Network Multitask Learning," *Electronics*, vol. 11, no. 17, p. 2697, 2022.

[20] K. Matt Nolan, "L'intelligence artificielle au service de la sécurité sur les voies de la ligne Docklands Light Railway du métro londonien," *Mobilité Urbaine, France, Intelligence Artificielle, Sécurité. Mar. 2022.*

[21] Z. Chen, J. Yang, C. Yang, "BrightsightNet: A lightweight progressive low-light image enhancement network and its application in "Rainbow" maglev train," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 10, p. 101814, 2023.

[22] C. Meng, Z. Wang, L. Shi, Y. Gao, Y. Tao, and L. Wei, "SDRC-YOLO: A Novel Foreign Object Intrusion Detection Algorithm in Railway Scenarios," *Electronics,* vol. 12, no. 5, p. 1256, 2023.

[23] Z. Chen, J. Yang, L. Chen, Z. Feng, L. Jia, "Efficient railway track region segmentation algorithm based on lightweight neural network and cross-fusion decoder," *Automation in Construction*, vol. 155, p. 105069, 2023.

[24] Z. Feng, J. Yang, Z. Chen, and Z. Kang, "LRseg: An efficient railway region extraction method based on lightweight encoder and self-correcting decoder", *Expert Systems with Applications*, vol. 238, part F., p. 122386, 2024.

[25] Z. Chen, J. Yang, Z. Feng, and H. Zhu, "RailFOD23: A dataset for foreign object detection on railroad transmission lines," *Scientific Data*, vol. 11, article no. 72, 2024.

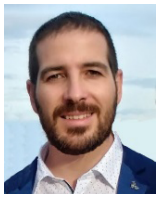
[26] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004*, vol. 2, Cambridge, UK, 2004, pp. 28-31, doi: 10.1109/ICPR.2004.1333992.

[27] F. Porikli and O. Tuzel, "Bayesian background modeling for foreground detection," in *Proceeding of the ACM International Workshop on Video Surveillance and Sensor Networks (VSSN 2005)*, November 2005, pp. 55-58.

[28] T. Cover and P. Hart, "Nearest neighbor pattern classification", *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967, doi: 10.1109/TIT.1967.1053964.

[29] R. Kalsotra and S. Arora, "Background subtraction for moving object detection: explorations of recent developments and challenges," *The Visual Computer,* vol. 38, pp. 4151–4178, 2022. https://doi.org/10.1007/s00371-021-02286-0

[30] B. Farnham, S. Tokyo, B. Boston, F. Sebastopol, and T. Beijing, "Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems," O'Reilly Second Edition, 2017.



D. Domingo Martínez Núñez

Domingo Martínez Núñez is an Inspector at the Central Control Station of Metro de Madrid, S.A, currently serving as an Engineer at the Ticketing Development & Compliance Center. He has more than 10 years of professional experience. Domingo received professional training in electronics from the Army Polytechnic Institute in Zaragoza (I.P.E.) and holds a Degree in Computer Engineering with a mention in Software Engineering at the International University of La Rioja (UNIR).

Dr. Fernando Carlos López Hernández

Fernando López is a full-time associate professor in the Applied Mathematics Department at Universidad Complutense de Madrid (UCM). His current research interests lie in image processing, computer vision, neural networks, dynamic systems, statistical machine learning and data-driven science. Before joining UCM, he was a full-time associate professor at Universidad Internacional de La Rioja (UNIR). He served as an education manager of the Doctorate Program in Computer Science, and in a Course of Robotics for Education. In addition, he lectured in the Computer Science degree covering subjects such as Statistics, Algebra, Discrete Mathematics, Algorithms Complexity, Image Processing, Signal Processing, Computer Graphics, Compilers.

Dr. J. Javier Rainer Granados

J. Javier Rainer. Director of the OTRI (Office for the Transfer of Research Results), and currently Director of the AENOR-UNIR Chair. Director of the Master in Quality Assessment and Certification Processes in Higher Education. PhD in Industrial Engineering from the Polytechnic University of Madrid (Spain). Master in Project Management, and University Expert in Management and Audit of Quality Systems. Personal Development Program in Artificial Intelligence. He has participated as a researcher and head of several R&D projects, has national and international publications, is a member of several international technical committees and has extensive experience in private companies in the telecommunications sector, where he has performed management and project management functions. In the field of university management, he has been Director of the Industrial Organization and Electronics Area of the School of Engineering and Technology of UNIR, and Deputy Director of Research of the same, and also the Director of the Degree in Industrial Organization. Research lines related to cognitive systems, decision making and learning.