



Article

A Dynamic Detection Method for Railway Slope Falling Rocks Based on the Gaussian Mixture Model Segmentation Algorithm

Xiulin Zhang, Qiang Fu, Yange Li, Zheng Han *, Nan Jiang and Changli Li 

School of Civil Engineering, Central South University, Changsha 410075, China; 15295781379@163.com (X.Z.); 214811181@csu.edu.cn (Q.F.); liyange@csu.edu.cn (Y.L.); 204801056@csu.edu.cn (N.J.); 204807061@csu.edu.cn (C.L.)

* Correspondence: zheng_han@csu.edu.cn; Tel.: +86-188-7416-3071

Abstract: Rockfall intrusion detection is crucial for the safety management of railway operations, and video detection methods help reduce deployment costs and improve detection efficiency. Mainstream neural network-based video detection methods have rapidly evolved in recent years but struggle to adapt to complex scenarios such as existing railway slope constructions due to weak generalization ability, low accuracy, and limited information acquisition. Therefore, this paper introduces a dynamic neural network detection model and establishes a dataset for rockfall intrusions in existing railway slope scenarios. The model initially relies on the YOLOv5 neural network, adopting an activation function suitable for the target scenario and addressing overfitting to achieve precise target recognition. Based on the neural network, the model dynamically detects rolling rockfalls by integrating a background subtraction algorithm based on the Gaussian Mixture Model and captures target dimensions using monocular vision technology, thus broadening the dimensions of detection information. Trials conducted on a railway in Shandong, China, demonstrate that the model accurately identifies moving rockfalls along the railway slopes and acquires the dimensions of moving rockfalls, successfully filtering out low-risk targets in the scene.

Keywords: railway slopes; object detection; YOLOv5; dynamic detection; background subtraction algorithm; monocular vision



Citation: Zhang, X.; Fu, Q.; Li, Y.; Han, Z.; Jiang, N.; Li, C. A Dynamic Detection Method for Railway Slope Falling Rocks Based on the Gaussian Mixture Model Segmentation Algorithm. *Appl. Sci.* **2024**, *14*, 4454. <https://doi.org/10.3390/app14114454>

Academic Editor: Tiago Miranda

Received: 10 April 2024

Revised: 6 May 2024

Accepted: 8 May 2024

Published: 23 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rockfalls, among the most prevalent geological hazards globally, occur abruptly, boast significant destructive potential, and are broadly distributed, thereby posing substantial threats to human life and infrastructure. On 14 August 2019, a substantial rockfall involving tens of thousands of cubic meters of rock occurred along the Chengkun Railway between Lianghong and Edai stations, damaging the railway infrastructure [1]; similarly, on 30 March 2020, a train derailment was triggered by a landslide along the Beijing–Guangzhou railway [2]. As the railway network continues to expand, particularly through complex and hazardous mountainous terrains, the reliance on sophisticated foreign object detection systems becomes essential [3]. Traditional railway foreign object intrusion detection relies on manual inspections, which are time-consuming, labor-intensive, and often lack high accuracy. In recent years, an array of automated foreign object detection technologies has been implemented in railway operation and maintenance for safety management, including grid, fiber Bragg grating, video, ultrasonic, and infrared detection methods [4]. These automated methods offer significant improvements over traditional manual detection in terms of efficiency, cost reduction, and enhanced stability and accuracy.

Detection methods are broadly classified into two categories by researchers: contact and non-contact detection [4]. Contact detection methods, which include grid detection and fiber Bragg grating detection, involve acquiring information through direct or indirect interaction between the sensor and the foreign object. Conversely, non-contact detection

methods, such as ultrasonic, infrared, and video detection, obtain parameter information about the intrusive object without any physical contact with its surface.

Researchers have also conducted extensive work in the area of contact-based detection. Liu [5] designed a railway foreign object intrusion detection system utilizing fiber Bragg grating stress sensors instead of traditional electrical sensors for the sensing module in railway foreign object intrusion monitoring. Xu et al. [6] devised a system that utilizes broadband Boolean chaotic signals for detection, employing a pair of leaky coaxial cables (LCX) to transmit the detection signal and receive the echo signal. Through comparison of the chaotic echo signal with its delayed repetition signal and analysis of the correlation trajectories before and after an intrusion, they introduced a foreign object intrusion detection system relying on leaky coaxial cables. Liu [7] enhanced traditional approaches by incorporating a layer of netting above and below the protective net, examining the distinct fracture waveforms induced by external forces, thereby proposing a novel foreign object intrusion monitoring scheme grounded in smart dual-electric grid technology. The contact detection method is challenging to deploy due to its reliance on sensor triggers and secondary signal transmission, resulting in a limited detection range. Additionally, manual reinspection is required for daily use. In contrast to contact detection methods, Berg et al. [8] employed computer software to acquire thermal images, calculate scene geometries, detect railways, and identify potential obstacles, subsequently issuing alerts to drivers based on predefined criteria. Garcia et al. [9] proposed a railway foreign object detection system that utilizes signal encoding based on complementary sequence pairs and employs optical transmitters. The non-contact detection method offers a longer monitoring distance, wider range, and higher reliability in comparison to the contact detection method. Compared to contact-based detection methods, non-contact approaches offer longer monitoring distances, broader coverage, and higher reliability. However, the effectiveness of detection is still subject to instability due to factors such as sensor sensitivity, infrared thermal radiation, and background radiation. Moreover, non-contact detection methods are unable to determine the precise location, shape, and size of foreign objects, resulting in a limited scope of detection information.

With the rapid advancement of neural networks and deep learning technologies in recent years, railway professionals have integrated these with image and video analysis to detect and alarm against foreign object intrusions in railway slope scenarios, conduct hazard assessments, and perform real-time monitoring of the operational environment. Unlike traditional non-contact video detection methods, neural networks not only enhance the learning capability of the entire detection system but also offer faster detection speeds and stronger adaptability, significantly improving the efficiency of railway operation and maintenance safety management. Currently, target detection algorithms based on convolutional neural networks (CNNs) have been widely applied in railway foreign object intrusion detection. As an image processing technique enabled by deep learning algorithms, target detection technology can identify and locate specific targets within images, constituting a crucial technology in the field of computer vision. Notable architectures include Faster R-CNN [10], YOLO [11], and SSD [12], among which the YOLO algorithm is particularly noted for its speed, accuracy, and end-to-end characteristics. Over the past few years, the field of target detection has seen rapid advancements. Bojarski et al. [13] have utilized convolutional neural networks for image recognition in applications such as vehicle navigation, obstacle avoidance, and decision-making. Taigman et al. [14] have combined machine vision facial detection technology with deep neural networks to achieve high-precision facial recognition under various environments and conditions. Unlu et al. [15] have employed multi-camera setups and a modified lightweight YOLOv3 architecture for detecting and tracking drones, demonstrating the application of deep learning algorithms in real-time object detection and classification within the realm of machine vision. Today, railway slope foreign object intrusion detection represents one of the most widely applied scenarios in railway operation and maintenance safety management [16]. Neural network-based target detection methods for railway slope foreign object intrusion can enhance detection

accuracy, reduce the frequency of false negatives and positives, accelerate inference speed, and facilitate deployment.

Neural network-based detection methodologies are highly reliant on the structure of the model itself and the datasets used for training. In specific target scenarios, prevalent mainstream models face challenges with limited generalization and low detection accuracy. Due to the complexity of slope environments, conventional neural networks often struggle to distinguish the mobility of rockfalls on the target slope and frequently misidentify high-speed moving rockfalls and stationary rock blocks as similar threats. Given that these systems can only recognize targets, such singular detection results fail to provide sufficient and effective information for comprehensive early warning systems. Furthermore, the lack of open-source rockfall datasets hampers effective model training, significantly diminishing the practical effectiveness of detections.

To address these challenges, this paper utilizes imagery data from the target construction section to develop a proprietary dataset for existing railway line slope scenarios. Concurrently, by conducting a comparative analysis of mainstream activation functions within the YOLOv5 model framework, function configurations suitable for specific scenarios have been identified, and issues related to model overfitting have been addressed. Building upon the improved neural network model, by integrating a background subtraction algorithm predicated on Gaussian mixture models with monocular vision technology, a dynamic neural network detection methodology capable of apprehending dimensional information regarding rockfalls has been proposed. This approach boasts high detection accuracy, strong generalization capabilities, and the ability to gather comprehensive information, addressing the limitations of current systems and enhancing the effectiveness of railway slope monitoring and safety management.

2. Enhanced YOLO Algorithm

2.1. Overview of the Enhanced YOLO Algorithm

YOLOv5 (You Only Look Once version 5) represents the fifth iteration in the YOLO series of real-time object detection network models. Since the inception of YOLOv1, the YOLO series has undergone several enhancements, with YOLOv5 showcasing significant improvements in performance and speed over its predecessors like YOLOv4. It is capable of real-time object detection with minimal latency, making it exceptionally well-suited for applications that demand rapid responses, such as autonomous driving and security surveillance. The model employs an end-to-end training approach, allowing it to accomplish the task of object detection in a single pass without the need for multi-stage processing. Meanwhile, complex engineering scenarios such as rockfall containment require ensuring the stability and maturity of the model. The YOLOv5 model, with its sufficient development time and history of applications, boasts a more robust user base and community support, offering developers a more stable performance compared to newer models like YOLOv8.

As depicted in Figure 1, the architectural framework of YOLOv5 is intricately composed of convolutional layers, pooling layers, residual blocks, and upsampling layers. Diverging from other object detection algorithms, YOLOv5 harnesses the technology of cross-stage feature aggregation (PANet) [17], amalgamating feature maps from various hierarchical levels to enhance the precision and recall of detections. Concurrently, YOLOv5 incorporates residual connections and convolutional modules akin to those found in ResNet [18] alongside attention mechanisms, endowing the model with robust feature representation capabilities. Moreover, YOLOv5 integrates methodologies such as data augmentation and label smoothing, significantly augmenting the model's generalization capacity and robustness. In this section, we compare the performance of mainstream activation functions to ensure the generalization ability of the training model and select the most suitable activation function and data processing configuration for specific scenarios.

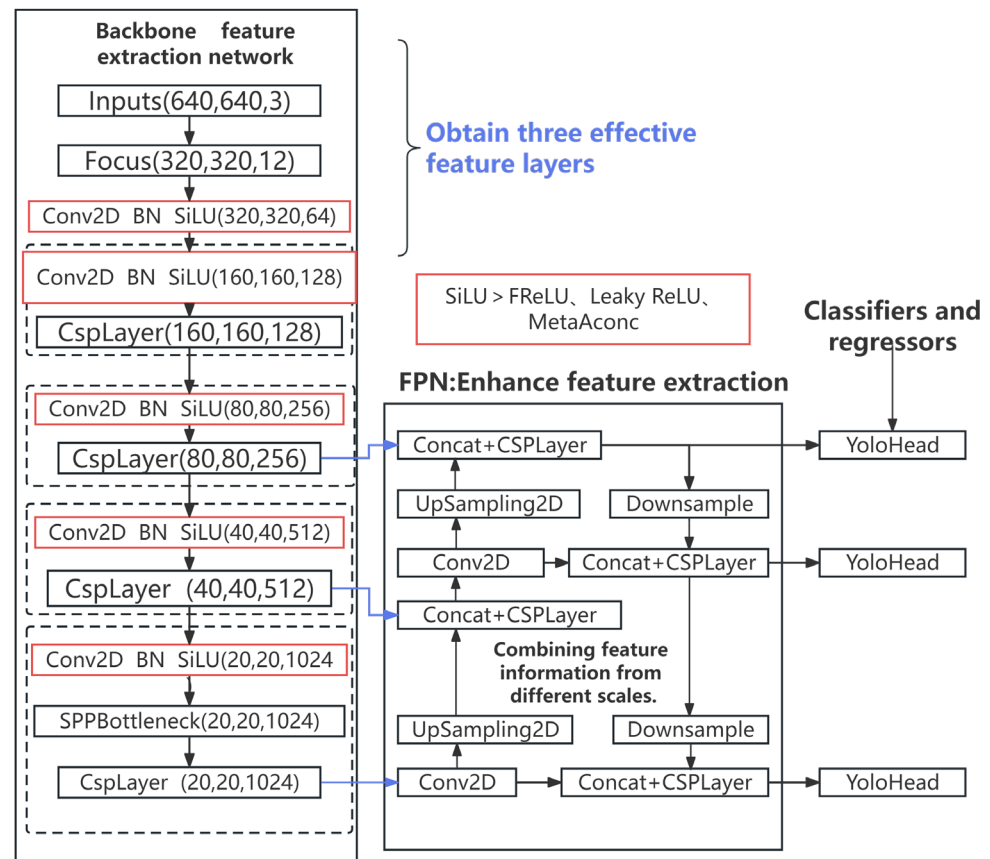


Figure 1. The YOLOv5 neural network detection workflow.

2.2. Dataset Collection and Processing

Currently, the mainstream public datasets in the field of object detection include the COCO dataset and the PASCAL VOC dataset [19]. However, it is challenging to find a sufficient quantity of data related to foreign object intrusion in railway slope scenarios within these public datasets, underscoring the importance of constructing a proprietary dataset for the detection of rockfall intrusions on railway slopes. This paper, through the processing and analysis of aerial survey data from an existing railway project in Shandong, China, examines the installation positions, heights, and angles of monitoring cameras within the railway context, thereby collecting a dataset for foreign object intrusion detection against the backdrop of railway slopes. The dataset primarily encompasses hazardous rocks and falling stones against the backdrop of existing railway slopes, including rocks of various shapes, sizes, and orientations. To accurately reflect engineering scenarios, the dataset also accounts for weather conditions, varying lighting conditions at different times, and the impact of camera focal lengths. Pre-collection of data for the automated detection model of existing line slope rockfall disasters primarily involves the gathering and processing of image data about experimental slope rockfalls. Based on a high-precision three-dimensional visualization model of the target area, image collection equipment is deployed along the experimental target section, and preliminary coordinated testing is conducted to pre-sample the slope structure data, strategically selecting surveillance locations with a focus on monitoring high-risk rockfall areas. This involves repeatedly collecting and filtering data that meet the criteria of image collection resolution, scene representativeness, and substantial image data volume. The data are then processed, including frame-by-frame extraction from videos and data augmentation, to amass a dataset that meets the requirements for later algorithm testing and training, providing a reliable material basis for the automated recognition and interpretation of high slope rockfall disasters. In this study, slope rockfalls of various sizes and orientations were

primarily selected. During the collection process, different camera heights, angles, and positions were chosen to ensure the diversity and coverage of the dataset, with some of the data illustrated in Figure 2.



Figure 2. Dataset excerpt: the three columns, from left to right, sequentially depict natural slope scenes, road surface scenes, and excavated cut slope scenes.

The collected data are annotated to facilitate the training and testing of models. Specifically, the initial phase involves using labeling to annotate rocks in the field-collected images. Upon completion, an XML file containing information about the target location and category is automatically generated. The XML annotation files created through manual annotation are then configured to the dataset path under the Annotations directory. The organized dataset comprises data from the Annotations and ImageSets directories, forming the improved dataset, including the training, validation, and test sets.

2.3. Selection of Activation Functions

To enhance the expressive power of the model while considering the speed of computation, selecting an appropriate activation function is crucial. Currently, the activation functions widely used in the YOLOv5 model include SiLU [20], FReLU [21], Leaky ReLU [22], and MetaAConc, among others. The SiLU function's primary advantage is its ability to produce non-zero activation values for negative inputs, thereby addressing the issue of traditional ReLU functions yielding zero activation for negative inputs. It also facilitates better gradient propagation while retaining certain non-linear characteristics. The FReLU func-

tion addresses the spatial insensitivity issue within activation functions, enabling regular convolutions to capture complex visual layouts and endowing the model with pixel-level modeling capabilities. Leaky ReLU, a variant of the rectified linear unit (ReLU) activation function, ensures neuron activation, enhancing the neural network model's non-linear fitting ability. The MetaAConc function, starting from the ReLU function, uses the Smooth Maximum approximation to prove that Swish is an approximate smooth representation of the ReLU function [23].

2.4. Handling Overfitting in the Model

Regularization is a technique employed in machine learning and statistical modeling to mitigate the risk of model overfitting. Overfitting occurs when a model exhibits exceptional performance on training data but fails to generalize well to unseen test data. This phenomenon typically arises in overly complex models that learn the random noise in the training data rather than the underlying data distribution. Regularization addresses this issue by introducing additional information or constraints to reduce the complexity of the model. This research focuses on the detection of rockfall intrusions on existing railway slopes, relying on a bespoke dataset. The effectiveness of neural network recognition is strongly associated with the relevance of the dataset to similar scenarios, and only a dataset that fully represents the characteristics of the scene can ensure optimal detection outcomes. Consequently, extensive data collection from existing railway slope scenes is essential to ensure effective neural network training. However, data from specific scenes inherently possess limitations, including restrictions on the shape of rocks and the angles of light and shadow. Thus, when a dataset tailored to a specific scene is established for model training, the process is highly susceptible to overfitting. This study proposes to employ weight decay, Dropout, and data augmentation strategies to address the overfitting issue within the bespoke dataset model, thereby enhancing the model's generalization capabilities while ensuring precision and efficiency in specific scenarios.

In this research, weight decay is implemented by adding the L2 norm of the weights as a penalty term to the loss function, which constrains the model's complexity and prevents it from learning the noise in the training data, thereby reducing the error on test data. Data augmentation, a common method to prevent overfitting, involves applying random transformations (such as rotation, scaling, translation, flipping, etc.) to the training data to generate new and varied training samples, thereby expanding the training set. This approach aids in better model generalization and improves performance on new data. Dropout [24] is a regularization method used during neural network training, functioning by randomly "dropping" (i.e., temporarily removing) a subset of neurons and their connections in the network during each training iteration. Consequently, the network encounters a slightly different version of itself at each iteration, introducing randomness that enhances the robustness of the model since it cannot rely on any specific set of weights. This technique helps prevent the model from overfitting the training data.

3. The Neural Network Dynamic Detection Model

To enhance the differentiation of hazardous rockfalls in a scene and filter out stationary targets, a background subtraction algorithm has been introduced to assist the neural network in identifying rockfalls within moving target areas. This approach not only reduces the detection area and accelerates the detection speed but also delineates the degree of risk posed by rockfalls in the target scene. Concurrently, the inclusion of a monocular vision algorithm for measuring the three-dimensional dimensions of the rockfall targets facilitates the acquisition of information for the comprehensive operation and maintenance safety management system. This establishes a neural network dynamic detection model tailored for rockfall detection in existing railway slope scenarios.

3.1. Background Subtraction Algorithm

The background subtraction algorithm [25] is a prevalent method in computer vision utilized to segregate foreground objects (such as moving entities) from the background within video frames. This approach finds utility across numerous applications, including surveillance cameras, pedestrian detection, and vehicle counting.

In the context of video surveillance or dynamic scenes, the background typically refers to the invariant or slowly changing components, whereas the foreground encompasses objects moving at a relatively rapid pace against the background. The principal objective of background subtraction is to identify and isolate these foreground objects. Table 1 lists the common background subtraction algorithms.

Table 1. An Overview of Common Background Subtraction Algorithms.

Method	Characteristics	Advantages
Frame Differencing	Simple subtraction between consecutive frames	Easy to implement, low computational cost
Gaussian Average	Modeling each pixel with a Gaussian distribution	Handles gradual illumination changes well
Gaussian Mixture Model	Modeling each pixel with a mixture of Gaussian distributions	Better at dealing with dynamic backgrounds than the simple Gaussian model, more precise
Median Filtering	Modeling the background using the median of a series of frames	Robust against noise

Median filtering models the background by computing the median pixel value at each position, exhibiting robustness in scenes with high noise levels and effectively mitigating some noise. However, it tends to generate false detections in scenarios with abundant moving objects due to its insensitivity to motion.

Frame differencing is highly susceptible to lighting changes and suffers from motion residue issues, making images processed with motion masks challenging for iterative computation by YOLO models.

The Gaussian averaging method demands substantial computational resources and struggles to swiftly adapt to scene changes, especially in environments with frequent or rapid object movement, resulting in slow background model updates.

In contrast, the BackgroundSubtractorMOG2 [26] algorithm employs a Gaussian mixture model with a more advanced background modeling strategy, including adaptive learning and updating of Gaussian model parameters for each pixel. This approach demonstrates robustness against factors like lighting changes and shadows, effectively extracting the foreground.

Unlike allocating a fixed number of Gaussian distributions for each pixel, BackgroundSubtractorMOG2 [26] dynamically assigns Gaussian distributions to each pixel as needed. This means that while some pixels may have only a few Gaussian distributions, others might have more. Such dynamism enables the algorithm to better adapt to diverse scenes and background changes. With the arrival of new frames, the algorithm updates the Gaussian Mixture Model (GMM) for each pixel. Specifically, it examines the color or brightness value of each pixel in the new frame and matches it to the corresponding Gaussian distribution. The parameters and weights of the matched Gaussian distribution are then updated accordingly. Moreover, the weights of those Gaussian distributions that do not match the current observation gradually decrease until they are eliminated. Beyond basic background modeling, BackgroundSubtractorMOG2 also features shadow handling capabilities, addressing a common issue in background subtraction where the shadows of moving objects may be mistakenly identified as foreground.

In summary, the BackgroundSubtractorMOG2 method exhibits stronger robustness and generalization capabilities in complex scenes compared to other mainstream algorithms, particularly in resisting interference from outdoor factors like lighting changes and shadows. Therefore, this study proposes integrating BackgroundSubtractorMOG2 as

an auxiliary detection algorithm, enabling YOLOv5 to intelligently recognize and identify intrusion events like falling rocks through dynamic monitoring results obtained from background subtraction.

3.2. Combining Background Subtraction Algorithms with YOLOv5 Neural Network

Given the complexity of the target railway slope sections and the significant differences between operational and non-operational periods, the task presents unique challenges. Normally operational railway slope areas tend to have a clean background, where any foreign object or stone becomes a hazard; however, construction sections along the main railway lines often feature complex backgrounds, such as scenes with numerous rocks or industrial materials, making it difficult to identify risk objects directly through neural network object detection technologies. It is challenging to achieve precise identification of hazardous rocks solely through accumulating data volume, annotating comprehensive datasets, and training mature deep learning models. Often, it necessitates extensive human intervention for accurate risk assessment across various scenes, which is not only time-consuming and labor-intensive but also results in models lacking in generalizability and interpretability, requiring retraining for new scenes.

To achieve high recognition efficiency and precision, this study analyzes the characteristics of high-risk targets on railway slopes and finds that target foreign objects often exhibit movement traits. In real scenarios, as opposed to low-risk stationary objects, hazardous rocks display significant kinetic energy, moving down slopes toward the railway. Even in non-operational periods with clean backgrounds, the intrusion of foreign objects into slope areas is due to their kinetic energy. Therefore, this research incorporates Background-SubtractorMOG2 as an auxiliary detection algorithm, combining YOLOv5's identification capabilities with background subtraction to obtain dynamic monitoring results for comprehensively determining the location of falling rocks, as illustrated in Figure 3.

In this study, the background subtraction object is initialized using the `createBackgroundSubtractorMOG2` function from the OpenCV library. This function implements a background subtraction algorithm based on a Gaussian mixture model, capable of adapting to environmental changes and effectively isolating foreground targets in images. When processing video sequences or continuous image frames, the algorithm distinguishes between background and foreground by analyzing pixel changes over time, generating a foreground mask. This mask contains only the pixels considered as foreground, i.e., potential target objects, significantly suppressing background information. After the background subtraction step, the preprocessed image, with the foreground mask applied, is fed into a pre-trained YOLO model for object detection. The YOLO model conducts a one-time analysis of the image through its convolutional network architecture, predicting the category and location of each object in the image. Since the input image has been preprocessed with the background subtraction algorithm, the model can more focusedly recognize and locate target objects in the foreground, thus reducing the negative impact of background complexity on detection performance.

To further enhance the accuracy of the detection results, the non-maximum suppression (NMS) algorithm is applied to the YOLO model's output to eliminate overlapping detection boxes and retain the most representative results. Moreover, by further analyzing the detected target locations within the foreground mask, false positives located in static backgrounds can be filtered out, retaining only the true positive detections in the foreground. This method also encompasses post-processing steps for the detection results, including resizing the detection boxes to fit the original image dimensions, processing specific category targets, and performing statistical analysis of the detection performance. Through a custom display function, the processed images and detection results can be displayed in real time, with the option to adjust the image display size as needed. Additionally, the detection results can be saved to a text file or directly drawn on the image with bounding boxes and labels, depending on user configuration.

Specifically, the background subtractor is applied to each frame to obtain a binary foreground mask. In this mask, regions with a value of 1 (or white) represent moving foreground targets, while regions with a value of 0 represent the background. A bitwise operation between the foreground mask and the original image retains only the moving foreground targets. Consequently, subsequent object detection steps are conducted solely on these moving targets, ignoring static backgrounds. Based on the results of the background subtraction algorithm, the YOLO model is used to detect objects in the preprocessed image. Since only moving foreground targets are present in the image, detection efficiency is enhanced. To ensure that each detected target is in motion, it is verified whether the center point of each target bounding box lies within the white area of the foreground mask. If the center point is outside the foreground area, the target is considered stationary and filtered out from the detection results. In summary, by preprocessing images with the background subtraction method and then combining it with YOLO object detection, stationary targets are effectively ignored, focusing detection solely on moving targets. This not only improves the efficiency of object detection but also prevents false positives on stationary targets.

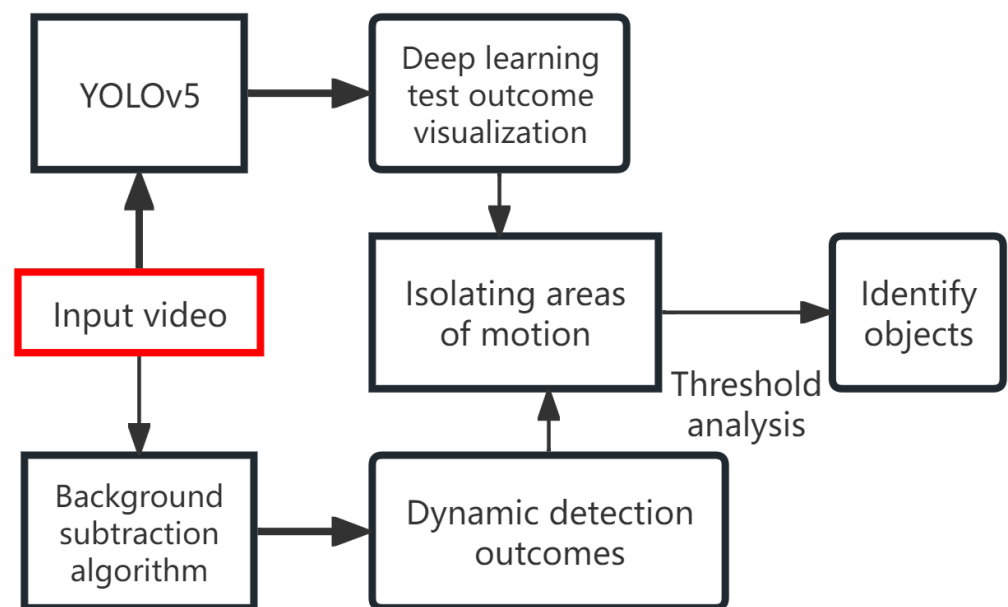


Figure 3. Methods of integrating background subtraction algorithms with neural networks.

Figure 4 demonstrates the basic detection workflow with simple images, from filtering moving targets to YOLO-based detection, represented from a to b.

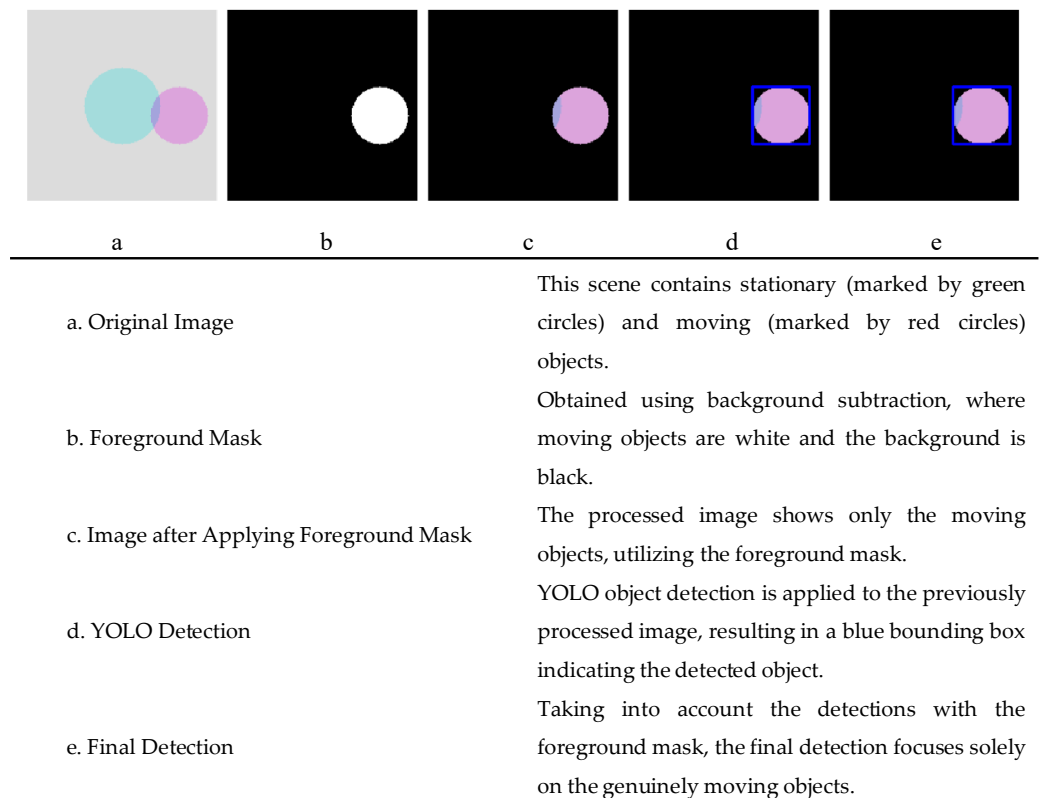


Figure 4. Workflow of integrating background subtraction algorithms with the YOLO model.

3.3. Monocular Vision Distance Measurement

Distance measurement methods based on machine vision primarily include binocular vision, monocular vision, and structured light vision techniques. The structured light vision method has numerous constraints regarding matching scenes, heavily relies on objective environments, and is significantly affected by lighting conditions. The binocular vision method requires calibration of internal and external parameters for multiple sensors and conversion of spatial coordinate systems. Despite its high precision and ability to obtain depth information, it is greatly influenced by the feature-matching process. Variations in parameters and the matching of synchronously captured images can lead to decreased accuracy [27].

On the other hand, the monocular vision measurement method can avoid these issues, offering a simpler structure and faster computational speed. Moreover, the cost of the widespread installation of monocular cameras is lower compared to the first two methods. Current monocular vision distance measurement techniques [28] primarily include the following methods: one involves using camera model parameters for coordinate system conversion, with accuracy highly dependent on camera parameter calibration. This method primarily utilizes internal camera parameters and principles of perspective projection geometry to deduce the relationship between the image coordinate system and the world coordinate system. Another method is mathematical regression, which establishes a regression model based on the relationship between the position of the target in the image under different states and the actual distance, with the need for extensive data increasing the difficulty of estimating distance information. Additionally, there is a method based on the imaging principle of the camera and the relationship of similar triangles, using actual distance information, distance information in the image, and the camera's focal length to establish a proportional relationship.

Compared to the first two methods, the monocular ranging model based on camera imaging principles features a simple structure and high accuracy, facilitating easy deploy-

ment and widespread application. This study will employ the method depicted in Figure 5 for capturing the dimensional information of the target.

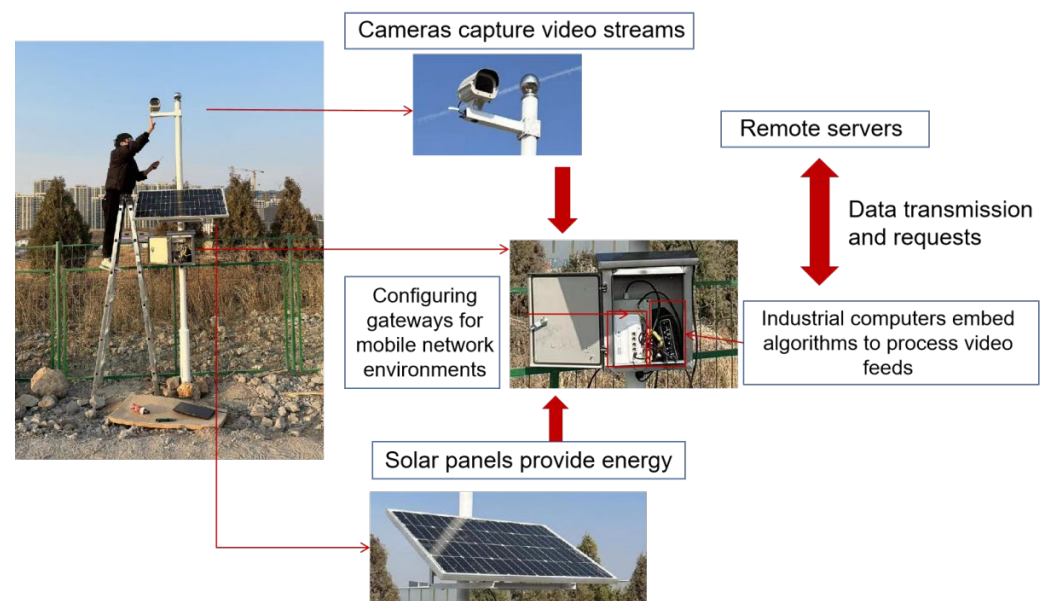


Figure 5. Automated detection equipment calibration.

Measuring the width of an object using monocular vision involves converting the pixel dimensions in the image to actual physical dimensions. This conversion depends on multiple factors, including the camera's internal parameters (such as focal length), the distance between the camera and the object, and the resolution of the image itself. In the absence of depth information (provided by stereo vision systems or depth cameras), monocular vision systems often require additional information or assumptions to estimate dimensions.

Here is a commonly used method for estimating the dimensions of objects based on monocular vision, assuming knowledge of the camera's internal parameters and the distance from the camera to the object through related calibration work:

1. Camera Model:

Focal Length (f): The focal length of the camera lens, is typically expressed in millimeters, but it needs to be converted into units consistent with the image resolution for calculations.

Sensor Size (S_x, S_y): The actual dimensions of the camera sensor in both horizontal and vertical directions, usually measured in millimeters.

Image Resolution (W, H): The pixel dimensions of the image, i.e., the number of pixels in both horizontal and vertical directions.

2. Object Size Estimation:

Assume you know the actual distance (D) from the camera to the object and the pixel width (w_{pixels}) of the object in the image.

3. Dimension Conversion:

First, calculate the actual size corresponding to each pixel based on the camera's sensor size and the image resolution:

$$(P_x) : P_x = \frac{S_x}{W} \quad (1)$$

Then, by employing the principle of similar triangles, the actual width of the object can be estimated:

$$W_{object} = w_{pixels} \cdot P_x \cdot \frac{D}{f} \quad (2)$$

W_{object} is the actual width of the object.

w_{pixels} is the width of the object in the image in pixels.

P_x is the actual size of each pixel in the image, expressed in the same units as the distance (D) from the camera to the object and the focal length (f).

D is the distance from the camera to the object.

f is the focal length of the camera expressed in the same units.

4. Experiment and Analysis

This paper introduces a railway slope foreign object intrusion detection algorithm that integrates neural networks with background subtraction algorithms, aimed at detecting foreign object intrusions on a certain railway slope in Shandong, China.

4.1. Equipment Coordination and Joint Testing

Researchers set up automated monitoring devices at the site as depicted in Figure 5, to furnish the requisite hardware environment for subsequent testing and eventual application of the algorithm, and also to facilitate pilot experiments for extensive deployment. The core components of the automated detection system include a high-definition camera, an industrial computer, and a gateway. The high-definition camera monitors and records the entire target area, transmitting the data to the industrial computer for processing; the industrial computer then detects slope rockfalls using pre-programmed model code; the gateway provides a robust network environment for overall computation and data transmission.

Initially, experimenters identified areas at risk of rockfall activity on the target railway based on preliminary point cloud data and aerial survey results. Cameras were used to collect data on moving rockfalls in the target area, ensuring diversity of background and breadth of targets. After preliminary data collection, an on-site survey was conducted in the target area to determine the best locations for camera installation, ensuring the detection range would adequately cover the target area. After this, equipment installation involved excavating sites for the cameras, laying reinforced concrete foundations, and mounting the cameras after the concrete had been set. Due to the unavailability of a local power supply, solar panels were used to power the devices. After connecting the power supply, the industrial computer and gateway were configured, activating the camera equipment. The cameras then monitored the target area to assess coverage detection accuracy and orientation. The industrial computer was connected to a remote server, preparing the system for algorithm integration and application.

4.2. The Experiment with the Enhanced YOLO Model

4.2.1. Comparison of Activation Functions in Slope Scenarios

Initially, based on the original model, this section employed four types of activation functions for adjustments: FReLU, SiLU, LeakyReLU, and MetaAconC. The training epochs were consistently maintained at 300 rounds.

This article primarily employs mAP_{0.5}: mean average precision (IoU = 0.5), training/validation set object confidence loss (obj_loss), and detection speed (FPS) as the performance evaluation metrics for the model. mAP_{0.5} indicates the average detection precision at an IoU threshold of 0.5; obj_loss assesses the disparity between the model's predicted bounding boxes (rectangles enclosing objects) and the actual boxes; FPS measures how many images the target detection network can process per second, evaluating the model's detection speed.

The results revealed that adjustments to the activation function caused noticeable differences in the model's accuracy and detection speed, with these differences primarily

manifesting in terms of detection speed. Moreover, when observing the entire detection process, the variance in accuracy altered as the training progressed. Throughout the training phase, as visualized in Figure 6, even though the distinctions between models narrowed after 250 rounds, the performance of SiLU and LeakyReLU was significantly superior to that of FReLU and MetaAconC functions in terms of accuracy. Simultaneously, as shown in Table 2, regarding detection speed, the model utilizing the SiLU function achieved a detection speed of 81.45 fps, and the model employing the LeakyReLU function reached a speed of 79.10 fps, both significantly outpacing the model with the FReLU function at 44.63 fps and the model with the MetaAconC function at 27.49 fps.

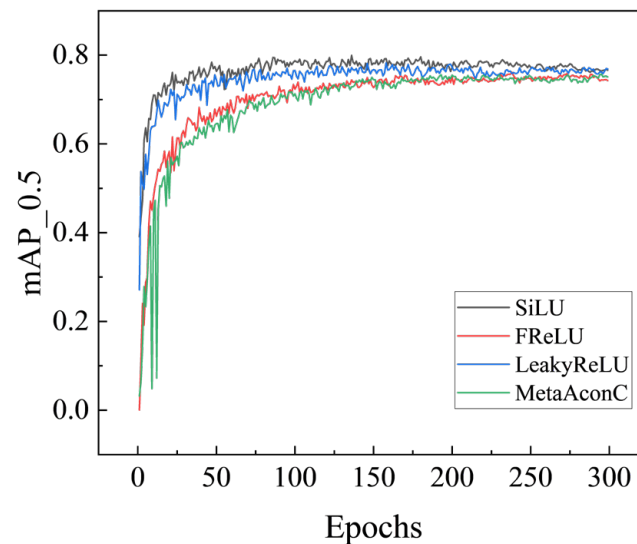


Figure 6. Visualization of the training process.

Table 2. Comparison of four activation functions.

	Metrics/mAP_0.5	Frame per Second
FReLU-based	0.74	44.63
SiLU-based (proposed model)	0.77	81.45
LeakyReLU	0.77	79.10
MetaAconC	0.75	27.49

4.2.2. Mitigation of Model Overfitting

In light of the results from the previous section, models employing the SiLU function demonstrated a marked superiority over those utilizing the other three functions, particularly in terms of detection speed. However, unlike the other functions, models trained with the SiLU function experienced a degree of overfitting, which compromised their generalizability. To ensure optimal usage of the function without overfitting within a certain number of training cycles, this section focuses on addressing the overfitting issue associated with the SiLU function.

In the experiments, weight decay—a regularization technique—was initially employed to tackle the overfitting problem in models using the SiLU function. This method operates by incorporating a term into the model's loss function, which constitutes a portion of the sum of squares of all model weights. Minimizing this new loss function compels the model to reduce the weights towards zero, though not exactly to zero, thereby preventing the model from overly relying on the training data and enhancing its performance on new data. The hyperparameter for weight decay was adjusted from the original 0.0005 to 0.001, multiplied by the sum of the squares of the weights. Despite the increased regularization strength through the adjustment of the weight decay coefficient, the overfitting issue

remained unresolved, as evident from the significant overfitting observed in the models, as illustrated in Figure 7.

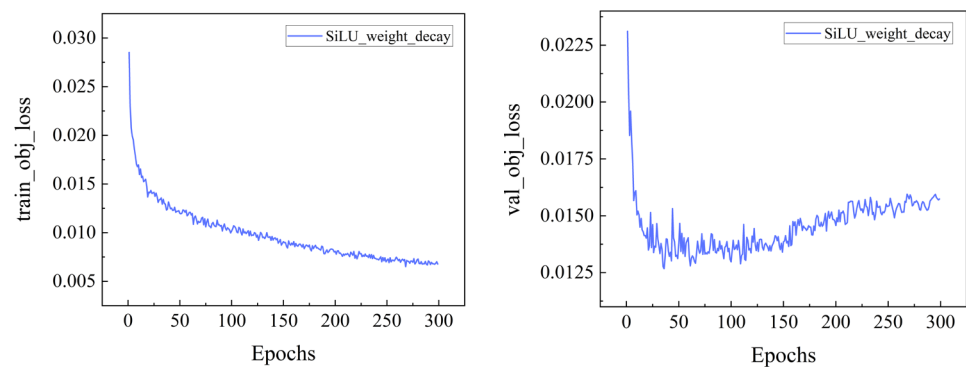


Figure 7. Confidence function curves for training and validation sets after weight decay treatment.

Subsequently, this section employed another regularization method—adding dropout layers—to address the overfitting issue in the model. The core concept of dropout involves randomly dropping (i.e., temporarily removing) a subset of neurons within the network during training, along with their corresponding connections. This prevents the model from overfitting the training data excessively, thereby enhancing the model’s generalization ability. In practice, dropout layers randomly select a portion of neurons (determined by the dropout rate) and set their outputs to zero during the current training step. The neurons to be dropped are chosen randomly at each training step, allowing the network to learn more robust features.

However, due to the random dropping of neurons, this approach may lead to the loss of useful features in convolutional neural networks, potentially resulting in decreased model performance. As shown in Figure 8, the introduction of dropout layers into the model’s training process led to a decline in model performance. Although the overfitting issue was addressed, the accuracy of the model decreased by approximately 8%.

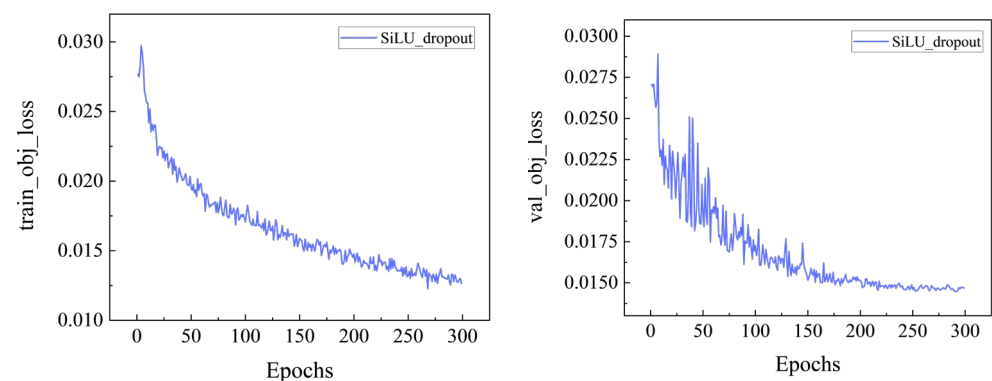


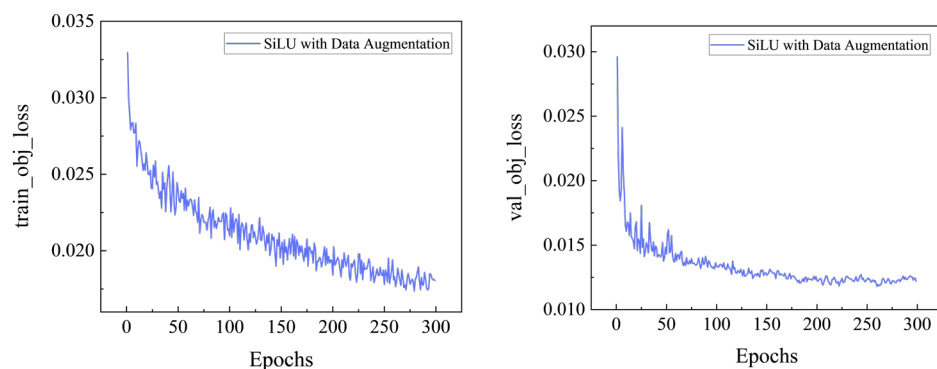
Figure 8. Confidence function curves for training and validation sets after dropout regularization treatment.

Ultimately, this section addresses the issue of model overfitting by modulating the intensity of data augmentation. Adjustments were made to the parameters governing the built-in data augmentation techniques of YOLOv5, thereby varying their strength. The fundamental parameters are delineated in Table 3.

Table 3. Comparison of four data augmentation parameters.

Parameter	Initial	Weak	Medium	Strong
Degrees: The angle of rotation	0.0	0.1	0.1	0.1
Translate: The extent of translation	0.1	0.1	0.1	0.1
Scale: The extent of scaling	0.5	0.5	0.9	0.9
Shear: The extent of shearing	0.0	0.1	0.1	0.1
Perspective: The degree of perspective transformation	0.0	0.0	0.01	0.01
FlipUD: The probability of flipping upside down	0.0	0.1	0.1	0.4
FlipLR: The probability of flipping left to right	0.5	0.5	0.5	0.5
Mosaic: The probability of using mosaic data augmentation	1.0	1.0	1.0	1.0
Mixup: The probability of using mixup data augmentation	0.0	0.1	0.1	0.4

The outcomes indicate that by escalating the level of data augmentation, the problem of overfitting was efficaciously mitigated. With robust data augmentation, the overfitting issue was essentially resolved over 300 training cycles, and a noticeable enhancement in the average accuracy of the training was also observed, as depicted in Figure 9. Furthermore, as Table 4 elucidates, in comparison to the pre-improvement SiLU function models, there was a significant augmentation in both detection precision and speed.

**Figure 9.** Confidence function curves for training and validation sets after intensive data augmentation treatment.**Table 4.** Comparison before and after data augmentation.

Model Parameters	mAP_0.5	mAP_0.5:0.95	fps
Before Improvement	0.766	0.766	81.45
After Improvement	0.806	0.806	83.49

In summary, the model utilizing the SiLU activation function achieves optimal accuracy and detection speed under the condition of strong data augmentation, effectively addressing the issue of overfitting when applied to proprietary datasets.

4.2.3. Analysis of Experimental Results

This section delineates the experimental environment and results analysis for the improvement of the YOLOv5 model. The experiments were conducted on a Windows 11 system, within a Python 3.8 environment, and using the PyTorch framework version 1.13.1, with CUDA version 11.7. To ensure consistency in hardware for each experiment, all algorithms were trained on a proprietary dataset as mentioned in Section 3.2, and all experiments were performed on a server equipped with an NVIDIA RTX 3060 (NAVID, Santa Clara, CA, USA) with 8 GB of video memory and 16 GB of GPU memory. The model weights were initialized based on pre-trained models from the VOC2007 dataset.

With the training cycles maintained at 300 to prevent overfitting and to achieve the optimal effect of activation functions, Table 5 compares the following four activation function configurations.

Table 5. Comparison of four schemes.

Strategies	Metrics/mAP_0.5	fps
FReLU	0.743	44.63
LeakyReLU	0.769	79.10
MetaAconC	0.751	27.49
SiLU with Data Augmentation	0.806	83.49

Therefore, the model configuration that employs the SiLU activation function and intensive data augmentation is most appropriate for detecting encroaching fallen rocks on railway slopes.

4.3. The Trial of the Neural Network Dynamic Detection Model

In summary, this paper integrates the background subtraction algorithm into the enhanced YOLOv5 model for the selection of moving targets and employs a monocular vision algorithm for the measurement of the targets' dimensions. This approach results in a dynamic detection model that incorporates machine vision with neural networks. This section will apply the model to test its efficacy in real-world railway slope scenarios.

In practical engineering scenarios, the mobility of an object is crucial in assessing its level of danger. In existing railway slopes, instances of singular rolling rockfall invasions are rare; most scenes are complex, featuring stationary objects that can interfere with risk detection. These are non-moving objects in the neural network that share origins with the detection targets, which a single neural network cannot avoid recognizing. Therefore, filtering out these less risky stationary objects is critical. As illustrated in Figure 10, interference targets exist in both scenes 1 and 2. A solitary neural network indiscriminately identifies targets, leading to potential misguidance, whereas models employing background subtraction algorithms can effectively filter out stationary objects, yielding more accurate detection results. The aforementioned images represent the detection outcomes of our model for specific slope scenarios.

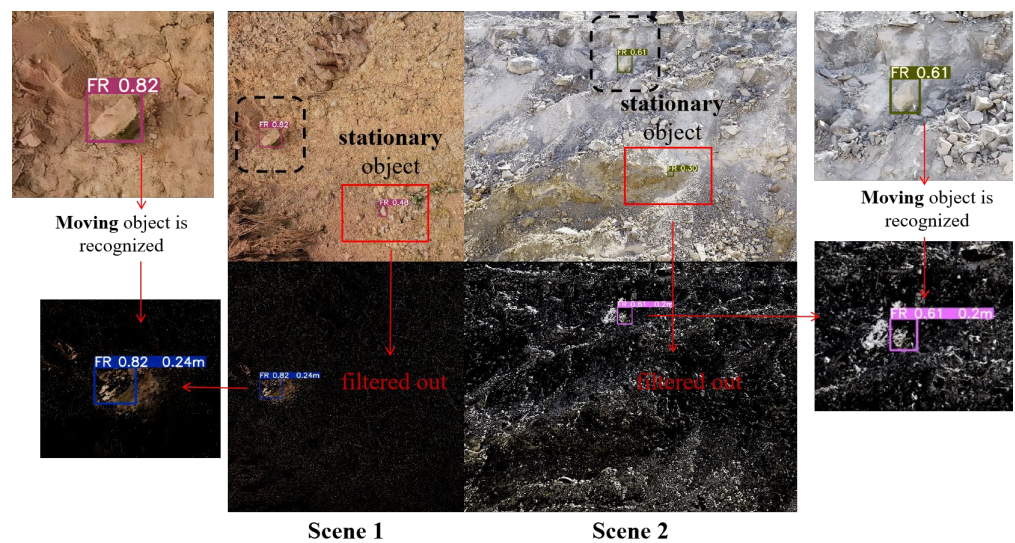


Figure 10. The detection performance of combining background subtraction algorithms with neural networks.

On the other hand, slope foreign object intrusion detection is dedicated to rapidly alerting of risks during the construction process along existing railway slopes. The scale

information obtained through monocular vision technology will facilitate a comprehensive understanding of the intrusion target information by the warning system. Hence, based on the model described earlier, a machine vision algorithm based on monocular imaging has been added to estimate the size information of moving foreign objects in the target scene.

In the same two scenarios mentioned above, as the detection model filters out moving targets and identifies them, the monocular vision-based detection simultaneously calculates the corresponding size information based on the basic data of the target in the scene.

As shown in Figure 11, by incorporating terrain data acquired earlier and leveraging neural networks and background subtraction algorithms, the final detection results not only effectively identified moving rocks but also calculated the estimated size of the target rocks using monocular vision technology. This has further enriched the existing railway slope rockfall intrusion detection system and essentially achieved a scientific and effective recognition of potential slope disaster risks.

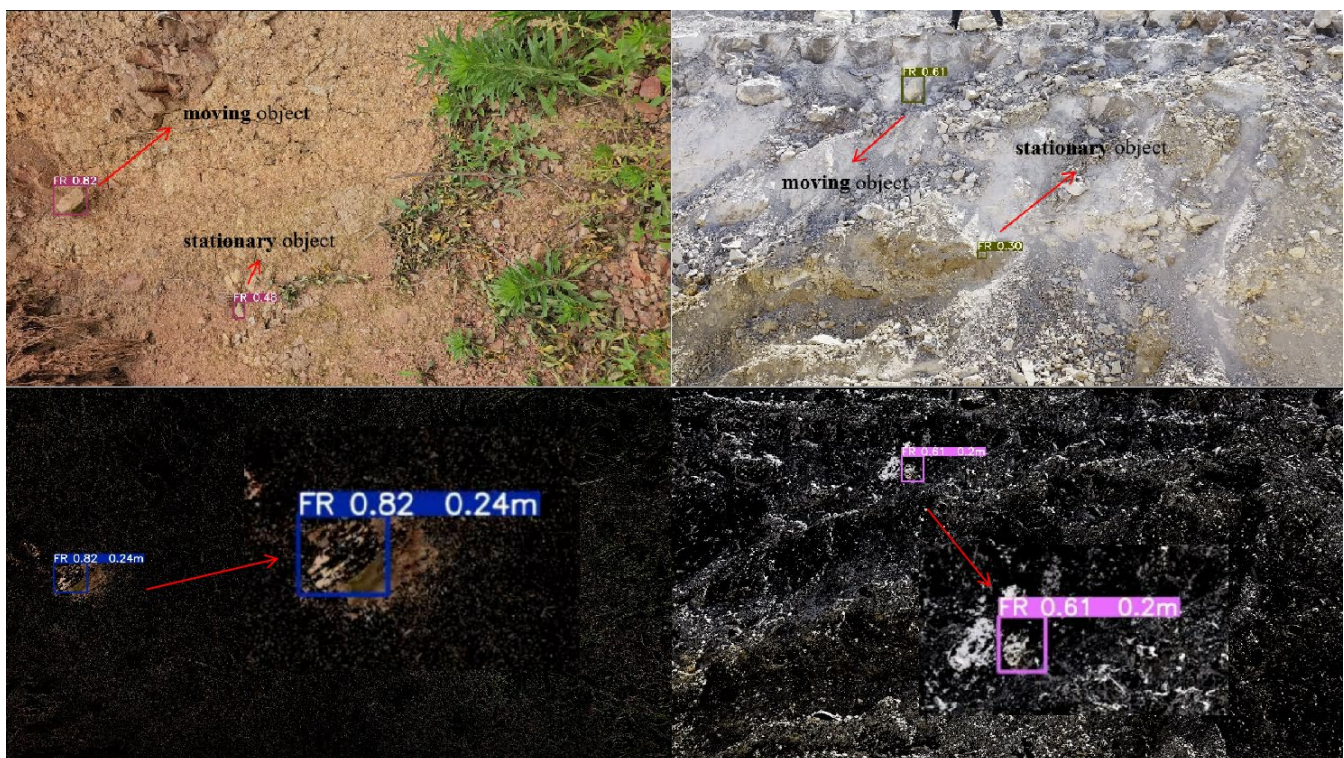


Figure 11. The final detection results of the introduction of the monocular vision algorithm: the width of the detection target is shown in the upper right corner of the inspection frame.

5. Discussion

This paper presents a dynamic rockfall detection algorithm that integrates a background subtraction algorithm with neural networks, based on existing railway slope scenarios, aimed at addressing the early warning needs for potential rockfall hazards in target scenes. Compared to traditional target detection algorithms, it offers higher precision and target matching, effectively leveraging the advantages of the background subtraction algorithm to filter out stationary, lower-risk targets. However, there are areas for improvement in the overall effectiveness:

- (1) The Gaussian mixture model method has limited interference resistance to background changes. In practical applications, it is still susceptible to weather, lighting, and terrain fluctuations, lacking stability. The background subtraction algorithm itself does not have autonomous learning capabilities; it merely delineates target areas and relies on the neural network of the YOLO model for identification. Future developments are anticipated to achieve autonomous learning capabilities.

- (2) The target information is singular. Even though rockfalls are the most common risk factor in railway slope scenarios, other potential threats to railway safety from foreign object intrusions still exist. Future work aims to aggregate other risk targets and expand the range of detectable foreign objects.
- (3) The dataset's scene diversity is limited. The dataset in this paper primarily relies on imagery from a certain railway slope in Shandong, China, which cannot broadly represent the diverse geomorphological and climatic railway operation scenarios nationwide. A dataset lacking in diversity will limit the extension to more scenarios.

6. Conclusions

The main research findings of this paper are as follows:

- (1) By comparing the detection outcomes of YOLO models using different activation functions, the optimal activation function suited to the target scenario was identified. An evaluation of the effects of weight decay, data augmentation, and the Dropout method on the YOLO model was conducted, thereby determining that the SiLU activation function exhibits the highest compatibility with this model; more robust data augmentation techniques significantly enhance the model's generalization capabilities.
- (2) Innovatively integrating a background subtraction algorithm into the improved YOLOv5 model, as described in (1), bitwise operations between the foreground mask and the original image are performed, resulting in the preservation of only the moving foreground targets. This addresses the issue of higher priority for moving targets in the target scene. A more suitable neural network dynamic detection algorithm for complex railway slope scenarios was proposed by combining the background subtraction algorithm with YOLOv5, and the dimensions of moving target information were enriched through monocular vision technology.

In summary, this paper first proposes a target detection algorithm for slope rockfalls with strong specificity through improvements to the YOLOv5 model and constructs a proprietary dataset for specific scenarios. By incorporating the background subtraction algorithm into the YOLOv5 model to address the issue of filtering moving rockfalls in complex slope scenarios, high-risk moving target detection information is obtained, and size measurement risk assessments are conducted. This forms a comprehensive neural network dynamic detection model capable of capturing dimensional information about rockfalls.

Author Contributions: Z.H. and Y.L. led the research program, and N.J. and C.L. designed the simulations. The manuscript was written by X.Z. and Q.F.; Z.H. and C.L. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the National Natural Science Foundation of China (Grant No. 52078493); the Natural Science Foundation of Hunan Province (Grant No. 2022JJ30700); the Natural Science Foundation for Excellent Young Scholars of Hunan (Grant No. 2021JJ20057); the Science and Technology Plan Project of Changsha (Grant No. kq2305006), and the Innovation Driven Program of Central South University (Grant No. 2023CXQD033). These financial supports are gratefully acknowledged.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. A Preliminary Report Confirms 17 People Missing following a Landslide on the Chengkun Railway in Ganluo County, Sichuan—Breaking News, Government of China Website. Available online: https://www.gov.cn/xinwen/2019-08/15/content_5421460.htm (accessed on 4 May 2024). (In Chinese)
2. A Train Derailment on the Beijing-Guangzhou Railway Has Resulted in 1 Death and 127 Injuries—Breaking News, Government of China Website. Available online: https://www.gov.cn/xinwen/2020-03/30/content_5497189.htm (accessed on 4 May 2024). (In Chinese)

3. Wang, K.; Gao, C.; Zhou, Y. Research on Design of Railway Bridge-shed Integrated Structure for Preventing Falling Rocks. *Railw. Stand. Des.* **2023**, *67*, 1–8. (In Chinese) [[CrossRef](#)]
4. Wang, Q.D.; Yang, Y.; Luo, Y.P.; Wei, X. Review on railway intrusion detection methods. *J. Railw. Sci. Eng.* **2019**, *16*, 3152–3159. (In Chinese) [[CrossRef](#)]
5. Liu, J. Study on Railway Intrusion Monitoring System Based on Fiber Bragg Grating Sensing. *Transp. Sci. Technol.* **2011**, 126–128. (In Chinese) [[CrossRef](#)]
6. Xu, H.; Qiao, J.; Zhang, J.; Han, H.; Li, J.; Liu, L.; Wang, B. A High-Resolution Leaky Coaxial Cable Sensor Using a Wideband Chaotic Signal. *Sensors* **2018**, *18*, 4154. [[CrossRef](#)] [[PubMed](#)]
7. Liu, Y. Research on an Improved Railway Intrusion Monitoring Scheme. *Railw. Commun. Signal Eng. Technol.* **2013**, *10*, 30–32. (In Chinese)
8. Berg, A.; Öfjäll, K.; Ahlberg, J.; Felsberg, M. Detecting Rails and Obstacles Using a Train-Mounted Thermal Camera. In *Image Analysis, Proceedings of the 19th Scandinavian Conference, SCIA 2015, Copenhagen, Denmark, 15–17 June 2015*; Paulsen, R.R., Pedersen, K.S., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 492–503. [[CrossRef](#)]
9. Garcia, J.J.; Hernandez, A.; Urena, J.; Garcia, J.C.; Mazo, M.; Lazaro, J.L.; Perez, M.C.; Alvarez, F. Low cost obstacle detection for smart railway infrastructures. In Proceedings of the IEEE Intelligent Vehicles Symposium 2004, Parma, Italy, 14–17 June 2004; pp. 670–675. Available online: <http://ieeexplore.ieee.org/document/1336464/> (accessed on 30 September 2023). [[CrossRef](#)]
10. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [[CrossRef](#)]
11. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
12. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905, pp. 21–37. [[CrossRef](#)]
13. Bojarski, M.; Del Testa, D.; Dworakowski, D.; Firner, B.; Flepp, B.; Goyal, P.; Jackel, L.D.; Monfort, M.; Muller, U.; Zhang, J.; et al. End to End Learning for Self-Driving Cars. *arXiv* **2016**, arXiv:1604.07316. [[CrossRef](#)]
14. Taigman, Y.; Yang, M.; Ranzato, M.A.; Wolf, L. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1701–1708. [[CrossRef](#)]
15. Unlu, E.; Zenou, E.; Riviere, N.; Dupouy, P.E. Deep learning-based strategies for the detection and tracking of drones using several cameras. *IPSN Trans. Comput. Vis. Appl.* **2019**, *11*, 7. [[CrossRef](#)]
16. Tong, L. A Study on Railway Obstacle Detection Using Machine Vision—CNKI. Available online: <https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD2012&filename=1012356721.nh&uniplatform=OVERSEA&v=YXyWwSZHCN3Q3TngmNiz-ps-13DvHMB4txHPxcrGKyEgW4xX6NZXIWeTPkW-fli-> (accessed on 15 February 2023). (In Chinese).
17. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [[CrossRef](#)]
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
19. Everingham, M.; Eslami, S.A.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes Challenge: A Retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [[CrossRef](#)]
20. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2017**, arXiv:1710.05941. Available online: <http://arxiv.org/abs/1710.05941> (accessed on 19 December 2023).
21. Ma, N.; Zhang, X.; Sun, J. Funnel Activation for Visual Recognition. In *Computer Vision—ECCV 2020, Proceedings of the 16th European Conference, Glasgow, UK, 23–28 August 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 351–368. [[CrossRef](#)]
22. Maas, A.L. Rectifier Nonlinearities Improve Neural Network Acoustic Models. 2013. Available online: <https://www.semanticscholar.org/paper/Rectifier-Nonlinearities-Improve-Neural-Network-Maas/367f2c63a6f6a10b3b64b8729d601e69337ee3cc> (accessed on 19 December 2023).
23. Ma, N.; Zhang, X.; Liu, M.; Sun, J. Activate or Not: Learning Customized Activation. *arXiv* **2021**, arXiv:2009.04759. Available online: <http://arxiv.org/abs/2009.04759> (accessed on 19 December 2023).
24. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
25. Stauffer, C.; Grimson, W.E.L. Adaptive background mixture models for real-time tracking. In Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, USA, 23–25 June 1999; Volume 2, pp. 246–252. [[CrossRef](#)]
26. Zivkovic, Z.; Van Der Heijden, F. Efficient adaptive density estimation per image pixel for the task of background subtraction. *Pattern Recognit. Lett.* **2006**, *27*, 773–780. [[CrossRef](#)]

27. Su, P.; Zhu, X. Research on Water Surface Target Recognition and Ranging Based on Monocular Vision. *Comput. Technol. Dev.* **2021**, *31*, 80–84. (In Chinese)
28. Feng, J.; Wang, J.; Zheng, X.; Wang, Y.; Jiang, H.; Wang, Z. Surface buoy position detection based on monocular vision. *J. Shanghai Marit. Univ.* **2023**, *44*, 17–23. (In Chinese) [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.