An intelligent railway surveillance framework based on recognition of object and railway track using deep learning

Rajiv Kapoor 1 · Rohini Goel 2 · Avinash Sharma 2

Abstract
In high speed railways, the intelligent railway safety system is necessary to avoid the accidents due to collision between trains and obstacles on the railway track. The unceasing research work is being performed to reinforce the railway safety and to diminish the accident rates. The rapid development in the field of deep learning has prompted new research opportunities in this area. In this paper, a novel and efficient approach is proposed to recognize the objects (obstacles) on the railway track ahead the train using deep classifier network. The 2-D Singular Spectrum Analysis (SSA) is utilized as decomposition tool that decomposes the image in useful components. That component is further applied to the deep classifier network. The obstacle recognition performance is enhanced by the combination of 2D-SSA and deep network. This method also presents a novel measure to identify the railway tracks. In addition, the performance of this approach is analyzed under different illumination conditions using OSU thermal pedestrian benchmark database. This system can be a tremendous support to curtail rail accidental rate and monetary loads. The results of proposed approach present good accuracy as well as can effectively recognize the objects (obstacles) on the railway track which helps to the railway safety. It also achieves a better performance with 85.2% accuracy, 84.5% precision and 88.6% recall.

Keywords Thermal imaging · Singular spectrum analysis · Singular value decomposition (SVD) · Region proposals · Faster R-CNN
1 Introduction

With the development of high speed railways, the safety necessities [37] of the railways are increasing day by day because the rapid speed of trains lead to rise in train mishaps. The obstacle on the railway tracks is one of the main factors that are responsible for the mishaps. The obstruction are pedestrians, vehicles crossing the tracks, animals meandering on the tracks [42] and some other heavy non-living objects fallen on the track from the overpass. In both developed and developing countries, a large number of railway crossing have not proper safety means like gates and lights. These issues cause challenges for the travelers as well as lead to monetary loss to railways as train cancelations and accidental compensations to aggrieved individuals. These issues can be resolved by utilizing a warning system that can significantly diminish the train mishaps by prior detection of the obstacles on the railway tracks. Various researches have been worked to detect and recognize the obstacles on the railway tracks. Zhou et al. [49] combined object detection and background learning by introducing a unified framework named DEtecting Contiguous Outliers in the Low Low-rank Representation (DECLOR) for detection of moving object.

Chen et al. [8] proposed a technique dependent on image matching and frame coupling to handle the object detection issue caused by moving camera and movement of the object. Berg et al. [5] proposed a approach for obstacle recognition on the railway track utilizing monocular thermal camera and procured a novel data collection. Sinha et al. [39] introduced an object detection method on railway tracks utilizing vibration sensors and separate out the signal from significant level acoustic noise utilizing a novel Monte Carlo based Bayesian analysis. V. Amaral et al. [2] discussed obstacle identification framework in railway level crossing utilizing utilizing 3D point clouds procured with 2D laser scanner. Wu et al. [46] introduced a coarse to fine thresholding plan on particle trajectories in video sequence of moving object captured by moving camera. Karaduman [22] proposed the obstacle identification framework utilizing laser distance meter and rail mounted camera. Manikandan et al. [29] proposed an obstacle recognition approach utilizing thermal camera and ADA boost algorithm. The rapid growth in field of machine learning and deep neural network is also contributing in the field of railway applications.

Mittal et al. [31] introduced a vision based railway track observing methodology utilizing deep learning classifier for uncontrolled real world data. Krummenacher et al. [24] introduced two machine learning strategies dependent on SVM and neural network to recognize rail wheels. Garcia et al. [13] proposed a methodology for independent train stop activity utilizing monocular vision-based methodology and Deep-Learning models. Kapoor et al. [21] demonstrated a framework using Hough transform and HSV segmentation with deep learning for recognition of railway track and obstacles on it. The scene of the railway track has something in common with the road scene; therefore the research work in the field of road information detection also helps to broaden the viewpoint. Qi Wang et al. [43] proposed a siamesed fully convolutional network (s-FCN) to segment the road regions in RGB images to improve the detection performance. Wang et al. [44] also weakly supervised adversarial domain adaption to enhance the segmentation performance from the synthetic data to real scenes using three deep neural networks. Sudha et al. [41] utilized the YOLO deep network and improved visual background extraction algorithm to detect multiple vehicles in the input videos. Zeng et al. [38] built a fusion network for segmentation of lane line and detection of road elements for autonomous driving. Nayak et al. [33] also demonstrated the use of various deep learning models for detection application particularly for COVID-19 detection in chest X-ray images.
whereas in the proposed work deep detection frameworks are used for thermal images of railway track.

Other than obstacles on the railway track, the adverse weather conditions like rain, haze, darkness, cloudy weather are also big challenge for railway safety. As these conditions have different level of illumination that directly affect the visibility of the moving train at high speed. Many times low visibility conditions can cause major loss to railways. The system should also be invariant to illumination so that it can provide proper visibility in any type of adverse weather condition. Mangale et al. [28] suggested a methodology to identify object for low illumination and different weather condition thermal images using directed acyclic graph (DAG). M. Kristo et al. [23] discussed an approach for thermal object detection in adverse weather conditions using YOLO framework.

Decomposition is a fundamental stage that is imperative for providing proper information to the classifier for image recognition. Sharma et al. [34] presented a method for the automated emotion recognition using higher order statistics (i.e. third order cumulants). The sub bands of signals are applied to particle swarm optimization to remove redundant information. Then the deep learning algorithm is used to predict the human emotions. Srirangam et al. [26] utilized the Time frequency matrix of EEG signals that is evaluated by Fourier Synchro Squeezing Transform (FSST) and Wavelet Synchro Squeezing Transform (WSST) to classify focal and non focal EEG signals using deep CNN. Chaudhary et al. [6] presented an approach for decomposition using 2D Fourier Bessel Series Expansion based empirical wavelet Transform (FBSE-EWT). The sub images from 2D-FBSE-EWT are used with machine learning methods and ResNet-50 based method to detect glaucoma. Chaudhary et al. [7] also used Fourier Bessel Series Expansion-based Decomposition (FBSED) with various pre-trained deep network classifiers to diagnose the COVID-19 using X-ray and CT images. The 2D-EMD [12] approach outperforms numerous relative methods like wavelet based image decomposition [11]. In any case, the execution of 2D-EMD is truly tedious because it is an iteration based process [20]. Whereas the 2D-SSA approach requires no iterations therefore it significantly curtails computational cost with improved efficiency [47]. The 2D-SSA approach outperforms 2D-EMD and many other number of state of art approaches in terms of accuracy and also requires very less execution time than 2D- EMD [9]. The 2-D SSA algorithm is really useful for extracting more precise discriminating patterns from original data scenes [17]. There is always a requirement of an object recognition framework that is capable of achieving high safety on the railway tracks. The achievements of the deep learning (RCNN [14], Fast/Faster RCNN [15, 36], YOLO [35] & SSD [25] etc.) in the field of object recognition motivated to analyze their performance in obstacle recognition on the railway track. The performance of the 2D-SSA [18] in the field of image enhancement and image restoration [3] further motivates to utilize its capabilities along with deep network to enhance the performance of object recognition for railway safety.

In this paper, the proposed approach automatically performs railway track identification and object (obstacle) recognition on the railway track. In this approach, a system is designed using 2D-SSA that decomposes the input frame into various discriminative components (patterns). This approach is intended to identify railway tracks by using particular component (pattern) that contains the information about tracks. Then component (pattern) having the object (obstacle) is applied to the deep networks to efficiently recognizes the obstacles on the railway track. In addition, the major contributions of the paper are:
The 2D-SSA approach is first time used to extract the railway track information to detect the railway track in railway safety applications. To the best of our knowledge, this is the earliest effort to use 2D-SSA with deep network classifier to enhance the object recognition performance. The performance of different deep networks with 2D-SSA is analyzed to choose suitable one for railway safety applications. This proposed approach also provides an efficient recognition performance under different illumination conditions when tested with OSU thermal pedestrian database from OTCBVS benchmark dataset collection [10].

The organization of paper in various subsections is as follows: Section 2 depicts the proposed strategy in detail including detection of railway tracks and recognition of obstacles on the railway track. The Section 3 describes the result and discussions. At the end, the conclusion and future scope provided in the Section 4.

2 Proposed method

The proposed technique is utilized to identify the railway track along with the obstacle on it from railway track thermal video (https://www.youtube.com/watch?v=xzGc71JFiB1). The basic block diagram of the proposed strategy is shown in Fig. 1. The frames are separated from the thermal video sequence and motion artifacts are compensated from the frame. Then the frames are decomposed into the ‘g’ distinct components having different important information of the scene using 2D- singular spectrum analysis (2D-SSA). Then the features of the railway track are extracted from ‘lth’ component to detect the railway track. Similarly, the ‘kth’ component having information about the obstacle is applied to the deep network (Faster R-CNN/ SSD/ Yolov2/Yolov5[27]) to recognize obstacle. The final predictions (railway track and obstacle) are combined to illustrate the final results at the output. The distinct steps of the proposed approach are illustrated in algorithm 1 and explained in the following subsections.

2.1 Compensation of motion artifacts in the frames

The thermal railway track video may have motion artifacts. The motion blur is the artifact of fundamental importance in the captured images when the camera is not in focus due to the train motion. In this work, the effect of motion blur in the frame is compensated by using wiener filtering [48] as it has shown good performance in removal of motion blur in the images. The motion artifacts in the form of blur are compensated by wiener filter as:

\[ I(u, v) = H_w(u, v) * F(u, v) \]

and

\[ H_w(u, v) = \frac{S(u,v)}{|S(u,v)|^2 + K} \]  

(1)

Where \( F(u,v) \) is the input frame with motion artifacts, \( H_w(u,v) \) is the wiener filter function, \( S(u,v) \) is the point spread function, \( k \) is the reciprocal of signal to noise ratio and \( I(u,v) \) is the output of wiener filter used for further processing.
Fig. 1 Basic block diagram of the proposed method used to recognize railway track and along with object on the track.
2.2 Decomposition of frames using 2D Singular Spectrum Analysis

The efficient information extracted from the thermal video frames can enhance the performance of recognition system. In this work, 2D-SSA [47] is used as an efficient decomposition tool from thermal images. When Singular Spectral Analysis (SSA) is implemented on the 2-D signal i.e. image, the steps of implementation are embedding, the Singular Value Decomposition (SVD) [40], diagonal averaging and grouping like 1-D SSA. The embedding and diagonal averaging steps have to be modified from 1-D to 2-D and rest of the stages remain identical with 1-D SSA. The steps of the 2D-SSA algorithm with mathematical explanation are discussed as below:

2.2.1 Embedding an image

In the embedding of the 2-D signal (image), the initial stage is to transform the image (2D-data) into another matrix form i.e. known as trajectory matrix [16]. The trajectory matrix not only contains all the information of an image but also retain the neighborhood information. Consider an image \( I^' \) of size \( s \times t \). To construct trajectory matrix \( T^' \), a window \( B \) of size \( m \times n \) is considered, where \( 1 \leq m \leq s \) and \( 1 \leq n \leq t \). Referring top left point of an image is reference point of the window. The reference points \( (x, y) \) ranges with \( 1 \leq x \leq -m = s - m + 1 \) and \( 1 \leq y \leq -n = t - n + 1 \). The path of movement of window \( B_{x,y} \) on image is given by reference points \( (x, y) \) as follows:

\[
(x, y) \rightarrow \begin{cases} 
(x, y + 1) & \text{if} \ y \leq t - n \\
(x + 1, 1) & \text{if} \ y = t - n + 1
\end{cases} \text{and} \ x \leq s - m
\]

(2)

Then, the elements of the each window \( B_{x,y} \) is restructured as columns by transposing rows one by one as:

\[
\overrightarrow{B_{x,y}} = (I_{x,y}, I_{x,y+1}, \ldots, I_{x,y+n-1}, \ldots, I_{x+m-1,y+n-1})
\]

(3)

where \( I_{x,y} \) is the pixel value at location \( (x,y) \) and \( \overrightarrow{B_{x,y}} \) is coloumn vector for window \( B_{x,y} \) with reference point \( (x,y) \). Finally all the column vectors \( \overrightarrow{B_{x,y}} \) are arranged into trajectory matrix \( T \) as:

\[
T = [\overrightarrow{B_{1,1}}, \overrightarrow{B_{1,2}}, \ldots, \overrightarrow{B_{1,-n}}, \overrightarrow{B_{2,1}}, \ldots, \overrightarrow{B_{2,-n}}, \overrightarrow{B_{3,1}}, \ldots, \overrightarrow{B_{m,-n}}]
\]

(4)

2.2.2 Singular Value Decomposition (SVD) of Trajectory matrix

The next step is to perform Singular Value Decomposition (SVD) of trajectory matrix \( T \). The SVD expansion of \( T \) can be obtained through the eigen decomposition of the lag covariance matrix \( C = TT^T \). The SVD of \( C \) can be represented as follows [30]:

\[
C = TT^T = \sum_{k=1}^{K} \lambda_k E_k E_k^T
\]

(5)
Where \( \lambda_k (k = 1, 2, K) \) are the eigen values of \( T T^T \) and \( E_k \) are normalized eigen vectors corresponding to \( \lambda_k \).

### 2.2.3 Process of diagonal averaging

The process of diagonal averaging is performed with the help of two step hankelization [19] process. Firstly, apply hankelization within each block and then applied between blocks. The averaging of the value belongs to the same element in the image is carried out. This enables to decompose the image into several components extracted based on SVD.

\[
I^{(k)} \leftarrow \text{DiagAveraging}(T_k) 
\]

(6)

Where, \( I^{(k)} \) is the 2D signal projected from \( T_k \) after diagonal averaging for \( (k = 1, 2, K) \) and \( T_k \) is the ‘\( k^{th} \)’ weighted orthogonal matrix which is given as:

\[
T_k \leftarrow \lambda_k E_k E_k^T 
\]

(7)

### 2.2.4 Grouping of 2D projected signals

Then the next step is grouping of the ‘\( K \)’ number of 2D projected signals \( I^{(k)} \) into ‘\( g \)’ disjoint groups using Hierarchical Cluster Analysis (HCA) [45] to obtain ‘\( g \)’ components of the input image as below:

\[
I_i \leftarrow \text{HCA}\left( \left\{ I^{(1)}, I^{(2)}, \ldots, I^{(K)} \right\}, g \right) 
\]

(8)

The ‘\( g \)’ groups of the 2D projected signals are formed as \( s^{(1)}, s^{(2)}, \ldots, s^{(g)} \) for \( 1 \leq g \leq K \). The every group \( s^{(i)} \) represents the decomposed component of the input image i.e. the sum of all the 2D projected signals of group ‘\( i \)’ which is represented as:

\[
\text{For } i \leftarrow 1 \text{ to } g \text{ do} \quad s^{(i)} \leftarrow \sum_{k=1}^{K} I^{(k)} 
\]

(9)

The final ‘\( g \)’ decomposed components of the input image are represented as ‘\( S \)’ which is given as:

\[
S = \left\{ s^{(1)}, s^{(2)}, \ldots, s^{(g)} \right\} 
\]

(10)

Once the input frame is decomposed into ‘\( g \)’ components, then the each component is analyzed to identify that which component contains the desired information i.e. information about railway track and obstacles on the tracks.
2.3 Railway track detection

2.3.1 Selection of component having railway track features

The first and most significant part in the proposed work is the recognition of railway track. The frame is decomposed into various discriminative components (patterns) using 2D Singular spectrum analysis (2D-SSA). The discriminative components contain vertical, horizontal and diagonal information of the scene. The pattern of tracks most resembles with the component having diagonal features. The \( l^{th} \) component that contains the strong pattern for the railway track is selected from the \('g'\) discriminative components of the particular frame as given below:

\[
S_{\text{Track}} = s(i)|_{i=1}
\]

Where, \( s(i)|_{i=1} \) is the \('l'\) discriminative component (pattern) having strong information about railway track.

2.3.2 Extraction of track features from selected component

Then the coordinates of the railway tracks are extracted from the \('l'\) component using thresholding segmentation. The track features are given as:

\[
S_{\text{Seg}}(i,j) = \begin{cases} 
1 & \text{if } T_{\text{low}} \leq S_{\text{Track}}(i,j) \leq T_{\text{high}} \\
0 & \text{else}
\end{cases}
\]

Where, ‘1’ and ‘0’ addresses white and dark color respectively, \( S_{\text{Track}}(i,j) \) is discriminative component (pattern) having strong information about railway track, \( S_{\text{Seg}(i,j)} \) is output image having railway track features, \( T_{\text{low}} \) and \( T_{\text{high}} \) are the lower & upper limit threshold. These coordinate locations are highlighted on the particular frame for the representation of the railway track. Hence the railway tracks are recognized in each frame of the video sequence.

2.4 Object recognition on the railway track

2.4.1 Selection of component having railway track features

Once the railway tracks are identified, the subsequent step is to recognize objects (obstacles) on the railway tracks. Initially the \('k'\) component (having horizontal information) that contains the strong pattern for the obstacles on railway track is selected from the \('g'\) discriminative components of the particular frame which is given as:

\[
S_{\text{Obstacle}} = s(i)|_{i=k}
\]

Where, \( s(i)|_{i=k} \) is the \('k'\) discriminative component (pattern) having strong information of obstacles on the railway track. After the extraction of components having obstacles information, this particular component is initialized to the deep network to recognize the obstacles.
2.4.2 Deep network for object recognition

Faster Region based Convolutional Neural Network (Faster R-CNN) [36] is utilized to perceive obstacles on the railway tracks. The Faster R-CNN comprises of two modules: initial, a deep fully convolutional network used to propose regions and second is the Fast R-CNN detector [15] that identifies the objects using region proposals. In default setup, there are anchors at position of image having 3 scales and 3 ratios. At last, the Fast R-CNN network is utilized for classification which has two fully connected layers. The one layer classifies the proposals in N+1 distinctive class (where N class and one background class for eliminating bad proposals). Another fully connected layer is utilized for better adjustment of bounding box for ‘N’ classes utilizing regression prediction. The anchors those overlap the ground truth with an Intersection over Union (IOU) greater than 0.7 are classified as foreground and the anchors don’t overlap any ground truth object (having IOU under 0.3) is classified as background. The loss function to be limited is given as:

\[
F_{Loss} (\{m_v\}, \{n_v\}) = \frac{1}{C} \sum_{v} L_c (m_v, m_v^*) + \gamma \frac{1}{R} \sum_{v} m_v^* L_r (n_v, n_v^*)
\]

(14)

Where, \(m_v\) is the predicted probability of anchor \(v\) being an object. Vector \(n_v\) indicates the parameterized coordinates of predicted bounding box. \(L_c\) and \(L_r\) are the classification and regression loss respectively.

The object identification strategy YOLOv2 [35] divides the image into \(S \times S\) grids and predicts \(B\) bounding box and \(C\) class probability for every framework cell. Each bounding box comprises of five predictions and object confidence. The object confidence addresses the dependability of existing object in the box. Nonetheless, as an overall object identification model, YOLOv2 is appropriate to cases where there are a variety of classes to be distinguished, and the distinctions among the classes are huge. At last, the non-maximum suppression (NMS) strategy is applied to dispose of redundant bounding boxes.

The most recent model of the YOLO object recognition model is now YOLOv5 [27], released by Glenn Jocher in 2020. YOLOv5 depends on the PyTorch structure. The YOLOv5s is the smallest model, and YOLOv5m, YOLOv5l and YOLOv5x are according to the increasing size. As the network size increases, its performance may likewise increases, at the expense of extra processing times. Accordingly, the bigger models may just be helpful for complex issues where enormous datasets are available.

During preparing, SSD [25] need to figure out which default boxes compare to ground truth detection and train the network accordingly. For each ground truth box, the selection is done from default boxes that vary over location, aspect ratio, and scale. The SSD training objective is driven from the MultiBox objective yet is extended to deal with various object classes. The overall objective loss function is a weighted sum of the localization loss &; the confidence loss.

\[
F_{Loss} (\{m_v\}, \{n_v\}) = \frac{1}{C} \sum_{v} L_c (m_v, m_v^*) + \gamma \frac{1}{R} \sum_{v} m_v^* L_r (n_v, n_v^*)
\]

(14)

Where, \(m_v\) is the predicted probability of anchor \(v\) being an object. Vector \(n_v\) indicates the parameterized coordinates of predicted bounding box. \(L_c\) and \(L_r\) are the classification and regression loss respectively.

The object identification strategy YOLOv2 [35] divides the image into \(S \times S\) grids and predicts \(B\) bounding box and \(C\) class probability for every framework cell. Each bounding box comprises of five predictions and object confidence. The object confidence addresses the dependability of existing object in the box. Nonetheless, as an overall object identification model, YOLOv2 is appropriate to cases where there are a variety of classes to be distinguished, and the distinctions among the classes are huge. At last, the non-maximum suppression (NMS) strategy is applied to dispose of redundant bounding boxes.

The most recent model of the YOLO object recognition model is now YOLOv5 [27], released by Glenn Jocher in 2020. YOLOv5 depends on the PyTorch structure. The YOLOv5s is the smallest model, and YOLOv5m, YOLOv5l and YOLOv5x are according to the increasing size. As the network size increases, its performance may likewise increases, at the expense of extra processing times. Accordingly, the bigger models may just be helpful for complex issues where enormous datasets are available.

During preparing, SSD [25] need to figure out which default boxes compare to ground truth detection and train the network accordingly. For each ground truth box, the selection is done from default boxes that vary over location, aspect ratio, and scale. The SSD training objective is driven from the MultiBox objective yet is extended to deal with various object classes. The overall objective loss function is a weighted sum of the localization loss &; the confidence loss.
Algorithm 1: Railway Track Detection & Object Recognition

**Input:**
Thermal Video Sequence of railway Track, \( g \): the number of component to extract \( m \) & \( n \): window size (the embedding dimension).

**Output:**
Railway tracks and obstacles.

**begin**
1. Extract frames \( I(i,j) \) from the video sequence.
2. Decompose the frame \( I(i,j) \) into elementary components using automatic SSA.
   \( T \leftarrow \text{BuildTrajectoryMatrix}(I, m, n) \)
   \( R \leftarrow \text{rank}(T) \)
   Apply Sing. SVD(\( T \))
   \( T_k \leftarrow \lambda_k E_k E_k^T \)
   \( t^{(k)} \leftarrow \text{DiagAveraging}(T_k) \)
   **End**

3. \( I_i \leftarrow \text{HCA}(\{t^{(1)}, t^{(2)}, \ldots, t^{(R)}\}, g) \)
4. \( s^{(i)} \leftarrow \sum_{k \in I_i} t^{(k)} \)
5. **Return** \( \{s^{(1)}, s^{(2)}, \ldots, s^{(g)}\} \)

**Case 1:** Railway Track Detection.
3. Select the \( i^{th} \) component having railway track features from the \( g \) output components.
4. Extract the railway track features from the \( i^{th} \) component.

**Case 2:** Object Recognition
5. Implement the deep network to identify the object from the \( k^{th} \) obstacle containing component.
6. The railway track coordinates and object information are unified to generate the final output.

**End**

### 2.4.3 Impact of braking distance of the train on the performance of object recognition

The trains have a braking distance before which they have to be alerted avoid collision with obstacles. The braking distance of a train is a distance from the point its brakes are applied to the point it comes to a stop. The braking distance ‘D’ of the train [4] is given as:
\[ D = -\frac{u^2}{2(a - g \times \tan \theta)} \]

and, \( \tan \theta = \frac{(h_2 - h_1)}{D} \) \( \tag{15} \)

Where, ‘\( u \)’ is the speed of train at the point when brakes are applied, ‘\( a \)’ is train’s deceleration rate, ‘\( g \)’ is the acceleration due to gravity and (\( h_2-h_1 \)) the gradient of the track i.e. the difference in height at which deceleration began (\( h_1 \)) and the its height at the stopping point (\( h_2 \)). Normally Indian trains decelerate at the rate of 0.5 m/s\(^2\) to 1.2 m/s\(^2\) which may increase up to 1.5 m/s\(^2\) in case of emergency brakes \[1\]. In plain regions, so the gradient of track (\( h_2-h_1 \)) is considered zero. The average speed of train is approx. between 80 and 100 Km/h \[1\]. The speed/time wise analysis along with the calculation of braking distance at speed of 100 Km/h is illustrated in Table 1.

The timely recognition of the obstacles on the track (i.e. before the braking distance) is the main aspect of the performance of the recognition system i.e. \( R_D > D \). The recognition distance \( R_D \) is the distance between obstacle and train at the time of recognition. The minimum time at which the system should give an alert, must be greater than braking time ‘\( t_B \)’ of the train. The performance of the network is improved for the timely recognition of the obstacle by optimizing the deep network parameters.

3 Result and discussion

In this section, the experimental setup for the proposed approach is discussed along with the experimental results illustrated subsequently.

3.1 Experimental setup

In this work, the frames are extracted from thermal video sequence of railway tracks. Total 749 frames of thermal video are utilized to figure out the performance of proposed work. The implementation of Faster R-CNN is performed using system with configuration Intel(R) Core(TM) i5-1035G1CPU @ 1.00 GHz, 1.19 GHz, 16 GB RAM and NVIDIA MX230 GPU.

3.2 Decomposition of for railway track and obstacle on track using 2D SSA

First of all the frames are extracted from the thermal video sequence of railway track as shown in Fig. 2(a) and motion artifacts are compensated. The results of the removal of the motion artifacts are illustrated in Fig. 2(c). Then the each frame is decomposed into several components using 2D-SSA. There are two important parameters which affect the performance of the SSA i.e. window size \( m \times n \) and the number of components ‘\( g \)’ to be decomposed from the input frame. The performance of the proposed approach is analyzed for different values of

| Table 1 | Braking time and braking distance calculation |
|---------|------------------------------------------|
| Speed of train ‘\( u \)’ (Km/h) | Deceleration rate ‘\( a \)’ (m/s\(^2\)) | Braking time (Seconds) \( t_B=(v-u)/a \) | Braking distance ‘\( D \)’ (meters) |
| 100   | 0.5       | 55.54     | 770         |
| 100   | 0.75      | 37.03     | 515         |
| 100   | 1.0       | 27.77     | 386         |
| 100   | 1.2       | 23.15     | 322         |
m × n and g but the most informative components are decomposed at window size of \( m \times n = 8 \times 8 \) and number of discriminative components (patterns) \( g = 15 \). Hence, every frame is decomposed into the 15 distinct components as shown in Fig. 2(d). The different components decomposed from input frame highlight the different information of the scene captured by the frame. In the proposed work, the desired information about railway track and obstacles are utilized for further processing.

### 3.3 Railway track detection

For the detection of the railway track, the every component decomposed from the each frame is analyzed, and then it is observed that the component \( S_{\text{Track}} = s^{(4)} \) (4th component) of every frame contains strong information about the railway tracks. Therefore, in this work, the 4th component of each frame is used for the recognition of the railway track. Then the features of the railway tracks in the 4th component are separated and utilized for the recognition of the railway track. At last, these features are superimposed on the frame to represent the tracks. The 4th component of the frame, the extracted features from the components and the final recognized tracks are shown in Fig. 3(b), (c) and (d) respectively. Similarly, the railway track is identified in all frames and hence it enables the railway track recognition.
Fig. 2 continued.
3.4 Object recognition

Once the railway track is detected and visualized, the next step is to identify object (obstacles) on the railway track. The decomposed components which have dominant information about the obstacles are utilized as input to the deep network for the recognition of the obstacles on the railway track. It is observed from the analysis of every component of each frame that the component $S_{\text{Obstacle}} = s^{(2)} \& s^{(3)}$ (2nd & 3rd component) of the frame contains the dominant information about the obstacles on the railway tracks. The 2nd component of the frame containing strong information about object (obstacle) to be recognized is shown in Fig. 4(b).

3.4.1 Training of deep network

In this work, the transfer learning technique is applied by using the pre-trained networks. Additionally, fine-tuning of the parameters and extending the training set with the sample collection are performed to improve the performance as much as possible. Through the transfer learning approach, the training is started with the implementation of the pre-trained parameters to include the useful information gathered from a previously trained network with thermal

---

**Fig. 3** (a) Input Frame (b) The 4th component of the input frame containing more dominant information of railway track (c) track features of the 4th component (d) Recognized Railway Track
railway data used for obstacle detection problem. For Faster R-CNN, 360 × 450 sized image patches were used to train the model. For the RPN stage, the bounding box scales were defined as 3 aspect ratios, which ensured that the network generated anchor boxes for each location of the feature maps. The batch size was defined as 1 to prevent memory allocation errors. For the training, the weight adjusting parameter ‘γ’ is set to 10 and Stochastic Gradient Descent with momentum value of 0.9. The process continued until 2000 iterations, which took 8 h with the learning rate at 0.001.

For the SSD network, 360 × 450 sized images were utilized for training. The sizes and perspective proportions of the default bounding boxes of each feature map layers remained as before. The RMSProp optimization technique was utilized for gradient estimations with a 0.001 learning rate and 0.9 decay factor for 2,000 iterations. The batch size was characterized as 16 & the training process took 6 h 17 min.

The Yolo-v2 model was trained with the Adam optimizer by a learning rate of 1 × 10−5 with a decay factor of 0.1, with which the validation loss didn’t diminish. 9 anchor boxes with various sizes and 3 for each phase of the network. The batch size was characterized as 16 and the entire training was done for 30 epochs. One epoch implies the feed forward and back propagation processes are finished for the entire training dataset. Training of the Yolo-v2 took around 4 h. The training procedure of Yolov5 model consisted of 15 epochs and a learning rate of 0.01. The batch size was defined as 5 which took 2 h 23 min for the thermal image dataset. Moreover, 640 of the 749 images were used for training, and 109 were used for validation.

### 3.4.2 Recognition of obstacles using deep networks

At last the different deep networks (such as Faster R-CNN, SSD, Yolov2 and Yolov5) are used to recognize obstacles on the railway tracks. When the deep networks are trained with training data, then 2nd components of the 2D SSA output of validation frames is applied to the network which gives recognized obstacles at the output. The recognized obstacle on the railway track in the 2nd component using Faster R-CNN along with railway track is shown in Fig. 4(c). Figure 4(d) shows the final recognized railway track & obstacle on the track using Faster R-CNN. The recognized obstacle on railway track in the 2nd component along with railway track using SSD is shown in Fig. 4(e).

Figure 4(f) shows the final recognized railway track and obstacle using SSD deep network on the frame. Figure 4(g) demonstrates the recognized obstacle on the railway track in the 2nd component along with railway track using YOLOv2. The final recognized railway track and obstacle using YOLOv2 deep network on the frame is shown in the Fig. 4(h). The recognized obstacle on the railway track in the 2nd component using Yolov5 is shown in Fig. 4(i) and the final recognized railway track & obstacle on the track using Yolov5 is shown in Fig. 4(j). The obstacle recognition using Yolov5 is also performed without 2D-SSA as shown in Fig. 4(l) to analyze the effect of 2D-SSA on the recognition performance of Yolov5.

### 3.5 Performance evaluation

#### 3.5.1 Evaluation metric

The performance of the proposed technique can be assessed with the assistance of various parameters. In this work, the important parameters used to analyze the performance are recognition accuracy [32], precision [32] and recall [32] which are given as:
Fig. 4 (a) Input Frame (b) The 2nd component of the input frame containing more dominant information of railway track (c) Recognized obstacle on the railway track using Faster R-CNN (d) Final Recognized Railway Track and obstacles in the frame using Faster R-CNN (e) Recognized obstacle on the railway track using SSD (f) Final Recognized Railway Track and obstacles in the frame using SSD (g) Recognized obstacle on the railway track using Yolov2 (h) Final Recognized Railway Track and obstacles in the frame using Yolov2 (i) Recognized obstacle on the railway track using Yolov5 (j) Final Recognized Railway Track and obstacles in the frame using Yolov5 (k) Input Frame (l) Recognized obstacle in the frame using Yolov5 (without 2D-SSA)
Fig. 4 continued.
\[ \text{Accuracy} = \frac{(\text{True Positive} + \text{True Negative})}{\text{Total Number of images}} \]  
(16)

\[ \text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \]  
(17)

\[ \text{Recall} = \frac{\text{True Positive}}{\text{Number of images having object class}} \]  
(18)

3.5.2 Evaluation at different IoU of faster R-CNN

On the railway track, obstacle ought to be recognized before an adequate distance (i.e. braking distance) with the goal that the speed of the train can be controlled to avoid collision. The performance of the Faster R-CNN is improved for the timely recognition of the obstacle by optimizing the Intersection over Union (IoU) parameter. All in all, the IoU value between 0.7 to 0.9 is considered as adequate range for the precise identification utilizing Faster R-CNN Network. The initial results of the proposed algorithm are acquired by identification network utilizing threshold value 0.7 as appeared in Fig. 5(a) yet the obstacle is recognize at shorter distance from the train. The performance of the system is analyzed for different values of IoU to make a tradeoff between recognition accuracy and distance from train at the time of recognition. To take care of this issue, it is observed that at the IoU threshold value ‘0.5’ the Faster R-CNN not only timely recognizes the obstacle on the track but also attains good accuracy. The final result of railway track identification and obstacle recognition at IoU =0.5 is appeared in Fig. 5(b).

The performance of the purposed system is analyzed for different IoU using above mentioned performance metrics.

Fig. 5  (a) The final results of railway track detection and obstacle recognition with Faster R-CNN at IoU=0.7 (b) at IoU=0.5

Springer
The parameters evaluated for different values of the Intersection over Union (IoU) as shown in Table 2. The accuracy, precision and recall parameters have better value at IoU ‘0.5’. Figure 6 illustrates that the graphical representation of the performance of the proposed method. It is concluded from the observations that the purposed method performs better at IoU= 0.5 with accuracy of approximately 85%.

3.5.3 Comparison with state-of-the-art methods

The performance comparison of our proposed method with the other approaches used for object recognition is done for thermal railway track data with these parameters as shown in Table 3. Initially the performance of YOLOv2, SSD, Yolov5 and Faster R-CNN framework are analyzed for recognition of obstacles, and then these approaches are used along with 2D SSA for enhancement of the recognition performance for thermal railway track data. The SSD approach performs better than the YOLOv2 framework but not than others. The Yolov5 and Faster R-CNN framework has shown better performance than both YOLOv2 and SSD. When

Table 2 Performance of the proposed method for different IoUs

| IoU  | Classification accuracy | Precision | Recall  |
|------|-------------------------|-----------|---------|
| 0.5  | 0.85                    | 0.845     | 0.88    |
| 0.7  | 0.775                   | 0.80      | 0.83    |

Fig. 6 The classification accuracy, precision and recall for different values of IoU
2D SSA is combined with these frameworks, it will boost the recognition performance. Although the performance of YOLOv2 and SSD is enhanced by utilizing 2D SSA but it unable to outperform the performance of the Faster R-CNN and Yolov5. When the Faster R-CNN and Yolov5 are used with 2D SSA, then the recognition system outperforms all the methods for thermal railway track data as discussed in comparison. The comparative plot is also shown in Fig. 7 for performance analysis of this work with other state of the art methods.

Table 4 illustrates accuracy vs. inference time analysis of the proposed work with other approaches. The performance of Faster R-CNN and Yolov5 is at par in terms of accuracy, precision and recall for thermal railway track data but.

| Method               | Accuracy | Precision | Recall  |
|----------------------|----------|-----------|---------|
| YOLOv2 [21]          | 0.725    | 0.716     | 0.743   |
| SSD [21]             | 0.738    | 0.727     | 0.751   |
| YOLOv5               | 0.832    | 0.827     | 0.869   |
| Faster R-CNN [21]    | 0.834    | 0.831     | 0.872   |
| YOLOv2 + 2D SSA      | 0.741    | 0.729     | 0.767   |
| SSD + 2D SSA         | 0.757    | 0.742     | 0.773   |
| YOLOv5 + 2D SSA      | 0.849    | 0.843     | 0.881   |
| Faster R-CNN+ 2D SSA | 0.852    | 0.845     | 0.886   |

Fig. 7 The comparative plot for performance analysis of proposed work with other state of the art approaches
Yolov5 outperforms in terms of speed. As discussed in Section 2, if the train is running at the speed of 100 Km/h, it takes 56 s. (approx.) to stop from the point when brakes are applied with deceleration of 0.5 m/s² and covers braking distance of 770 m. The proposed system detects the obstacle from the distance of 1Km i.e. the proposed system gives alert approx. 73 s. prior. As discussed in Table 4 the detection time required is 231 ms. The minimum processing time (i.e. sum of detection time and braking time) is 231ms + 56 s = 56.231 s. Hence the proposed system can comfortably avoid the possibility of collision between train and obstacle to enhance railway safety.

| Method                  | Accuracy | Inference Time |
|-------------------------|----------|----------------|
| YOLOv2 [21]             | 0.725    | 25 ms          |
| SSD [21]                | 0.738    | 63 ms          |
| YOLOv5                  | 0.832    | 9 ms           |
| Faster R-CNN [21]       | 0.834    | 209 ms         |
| YOLOv2 + 2D SSA         | 0.741    | 47 ms          |
| SSD + 2D SSA            | 0.757    | 85 ms          |
| YOLOv5 + 2D SSA         | 0.849    | 29 ms          |
| Faster R-CNN+ 2D SSA    | 0.852    | 231 ms         |

Yolov5 outperforms in terms speed. As discussed in Section 2, if the train is running at the speed of 100 Km/h, it takes 56 s. (approx.) to stop from the point when brakes are applied with deceleration of 0.5 m/s² and covers braking distance of 770 m. The proposed system detects the obstacle from the distance of 1Km i.e. the proposed system gives alert approx. 73 s. prior. As discussed in Table 4 the detection time required is 231 ms. The minimum processing time (i.e. sum of detection time and braking time) is 231ms + 56 s = 56.231 s. Hence the proposed system can comfortably avoid the possibility of collision between train and obstacle to enhance railway safety.

**Fig. 8** (a) Input thermal image under light rain condition (b) The 2nd component of the input image containing more dominant information of pedestrians (c) Recognized pedestrians in the 2nd component (d) Final Recognized pedestrians in input thermal image
3.6 Performance of proposed work under different illumination conditions

The low illumination due to adverse environmental conditions like rain, Haze, cloudy weather and darkness can cause problems in recognition of obstacles. The proposed work is also tested for different illumination conditions to analyze its illumination invariance. The OSU Thermal Pedestrian Database from OTCBVS Benchmark Dataset Collection [10] is used to test the performance of the purposed work in different illumination conditions. The dataset contains thermal images of different weather conditions like light rain, Haze, cloudy and fair sunlight. The recognition performance is analyzed with the Faster R-CNN network that was additionally trained on thermal image data for the object class. For the training and testing purposes, the Faster R-CNN detector architecture is used. Figure 8(a) shows the thermal image under light rain conditions in the afternoon with temperature: 68 °F, minimal dew point: 58 °F, humidity: 70% and visibility: 9.0 miles. The 2nd component is decomposed from the image using 2D SSA as shown in Fig. 8(b) because the 2nd component strongly contains the information about pedestrian. Then the 2nd component of the thermal image is applied to the Faster R-CNN for the recognition of pedestrian. Figure 8(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 8(d). The proposed approach performs well in the rainy conditions with thermal imaging. The thermal image captured under Haze with minimal dew point: 47 °F, humidity: 89% and visibility: 2.5 miles is shown in

Fig. 9  (a) Input thermal image under Haze condition (b) The 2nd component of the input image containing more dominant information of pedestrians (c) Recognized pedestrians in the 2nd component (d) Final Recognized pedestrians in input thermal image
Fig. 9(a). Figure 9(b) shows the 2nd component is decomposed from the image using 2D SSA because the 2nd component strongly contains the information about pedestrian. Figure 9(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 9(d). It is analyzed that the proposed system can work for recognition of objects under Haze conditions. The proposed work is also tested for thermal image having scene captured under dark cloudy conditions. Figure 10(a) shows the thermal image under dense cloudy conditions in the morning with temperature: 53 °F, minimal dew point: 46 °F, humidity: 77% and visibility: 8.0 miles. The 2nd component is decomposed from the image using 2D SSA as shown in Fig. 10(b) because the 2nd component strongly contains the information about pedestrian. Figure 10(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 10(d). This method also shows good recognition results for low illumination in dark clouds. The next illumination condition is considered when weather is clear and sunny. The thermal image captured under fair sunny conditions in the morning with temperature: 71 °F, minimal dew point: 32 °F, humidity: 45% and visibility: 10 miles is shown in Fig. 11(a). Figure 11(b) shows the 2nd component is decomposed from the image using 2D SSA because the 2nd component strongly contains the information about pedestrian. Figure 11(c) represents the output of the Faster R-CNN with recognized pedestrian. The final recognized pedestrians in the input image of light rain conditions are shown in Fig. 11(d). At last, the illumination condition in partly cloudy weather...
is considered. The thermal image captured in partly cloudy in the morning with temperature: 57 °F, minimal dew point: 37 °F, humidity: 47% and visibility: 10 miles is shown in Fig. 12(a). The 2nd component is decomposed from the image using 2D SSA because the 2nd component contains the information of pedestrian as shown in Fig. 12(b).

The output of the Faster R-CNN with recognized pedestrian is represented in Fig. 12(c). Figure 12(d) shows the final recognized pedestrians in the input image of light rain conditions. Hence the performance of the proposed work is analyzed under different illuminations in different weather conditions; it gives good illumination invariance and shows better performance in all illumination conditions.

### 4 Conclusion and future scope

This paper introduced a novel and efficient method for identification of railway track and obstacle recognition on the railway track. The proposed approach utilizes the 2D SSA to decompose the thermal image into distinct information carrying components, and then the deep learning network used the particular component to identify the obstacle on the railway tracks. Identification of the railway tracks using SSA is another significance of this work. The concatenation of Deep network with 2D SSA introduced a more efficient and robust approach.
to build a early warning framework to forestall railway accidents to enhance railway safety. As there is no significant modification is needed in the train infrastructure, so this system will be cost-effective. This will also diminish the monetary burden in terms of railway compensation. Additionally, this work is also shows good performance when tested for different illumination conditions using OSU thermal pedestrian database from OTCBVS benchmark dataset collection. This system will be extended for more efficient object recognition under distinct recognition challenges like occlusion, background clutter etc. as a future work to improve the adequacy of the framework.

**Acknowledgements** Authors are grateful to Delhi Technological University and specially Sh Anand Vardhan, CEO of MM Logic soft P Ltd for sponsoring the project.

**References**

1. Aakash DB, Ramachandran N, RastogiV (2019) Studies on the effects of braking loads on a Railway Wheel. In 2nd International Conference on Computational & Experimental Methods in Mechanical Engineering, IOP Conference Series: Materials Science and Engineering, 691, pp 3–5
2. Amaral V, Marques F, Lourenco A, Barata J, Santana P (2016) Laser-based obstacle detection at railway level crossings. J Sensors, Hindawi Publishing Corporation, pp 1–11
3. Aragon R, Zhigljavsky LA (2010) Singular spectrum analysis for image processing. Stat Interface 3(3):419–426
4. Barney D, Haley D, Nikandros G (2001) Calculating train braking distance. In Conference: Safety Critical Systems and Software 2001, Sixth Australian Workshop on Safety-Related Programmable Systems, (SCS2001), St Lucia, Queensland, Australia, 3, pp 23–29
5. Berg A, Ofjall K, Ahlberg J, Felsberg M (2015) Detecting rails and obstacles using a train-mounted thermal camera. In Image Analysis, Scandinavian Conference on Image Analysis (SCIA). Springer, Berlin, 9127, pp 492–503
6. Chaudhary PK, Pachori RB (2021) Automatic diagnosis of glaucoma using two-dimensional Fourier-Bessel series expansion based empirical wavelet transform. Bioned Signal Process Control 64:1–17
7. Chaudhary PK, Pachori RB (2021) FBSED based automatic diagnosis of COVID-19 using X-ray and CT images. Comput Biol Med 134:104454
8. Chen Y, Zhang Rh, Shang L (2014) A novel method of object detection from a moving camera based on image matching and frame coupling. PLoS ONE 9(10):e109809. https://doi.org/10.1371/journal.pone.0109809
9. Daubechies I, Lu J, Wu HT (2011) Synchronsqueezed wavelet transforms: An empirical mode decomposition-like tool. Appl Comput Harmon Anal 30(2):243–261
10. Davis JW, Keck MA (2005) A two-stage template approach to person detection in thermal imagery. Proc. 7th IEEE Workshops Appl. Comput. Vis. (WACV/MOTION) 1:364–369
11. Demir B, Erturk S (2008) Improved hyperspectral image classification with noise reduction pre-process. In 6th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland
12. Demir B, Erturk S (2010) Empirical mode decomposition of hyperspectral images for support vector machine classification. IEEE Trans Geosci Remote Sens 48(11):4071–4084
13. Garcia ME, Labayen M, Zamalloa M, Arexolaleiba NA (2020) Application of computer vision and deep learning in the railway domain for autonomous train stop operation. In IEEE/SICE International Symposium on System Integration (SII), pp 1-1
14. Girshick R, Donahue J, Darrell T, Malik J (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In IEEE Conference on Computer Vision and Pattern Recognition, pp 580–587
15. GirshickR (2015) Fast R-CNN. In Proceedings of IEEE International Conference on Computer Vision (ICCV), pp 1440-1448
16. Golyandina N (2010) On the choice of parameters in singular spectrum analysis and related subspace-based methods. Stat Interface 3:259–279
17. Golyandina NE, Usevich KD (2009) An algebraic view on finite rank in 2D-SSA. Proceedings of the 6th St. Peters burg Workshop on Simulation 1:308–313
18. Golyandina NE, Usevich KD (2010) 2D-extension of singular spectrum analysis: algorithm and elements of theory. In Matrix Methods: Theory, Algorithms, Applications World Scientific, pp 449–473
19. Golyandina N, Nekrutkin V, Zhigljavsky A (2001) Analysis of time series structure: SSA and related techniques. In: Chapman & Hall /CRC
20. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decompositionand the Hilbert spectrum for nonlinear and non-stationary time seriesanalysis. Proc R Soc Lond A Math Phys Eng Sci 454(1971):903–995
21. Kapoor R, Goel R, Sharma A (2020) Deep learning based object and railway track recognition using train mounted thermal imaging system. J Comput Theor Nanosci 17(11):5062–5071
22. Karaduman M (2017) Image processing based obstacle detection with laser measurement in railways. In 10th International Conference on Electrical and Electronics Engineering (ELECO), pp 899-903
23. Kristo M, Ivasickos M, Pobar M (2020) Thermal object detection in difficult weather conditions using YOLO. IEEE Access 8:125459–125476
24. Krummenacher G, Ong CS, Koller S, Kobayashi S, Buhmann JM (2018) Wheel defect detection with machine learning. IEEE Trans Intell Transp Syst 19(4):1176–1187
25. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In IEEE European Conference on Computer Vision (ECCV)
26. Madhavan S, Tripathy RK, Pachori RB (2020) Time-frequency domain deep convolutional neural network for the classification of local and non-local EEG signals. IEEE Sens J 20(6):3078–3086
27. Malta A, Mendes M, Farinha T (2021) Augmented reality maintenance assistant using YOLOv5. Appl Sci 11(11):4758. https://doi.org/10.3390/app11114758
28. Mangale S, Tambe R, Khambete M (2017) Object detection and tracking in thermal video using Directed Acyclic Graph (DAG). ICTACT J Image Video Process 08(01):1566–1574
29. Manikandan R, Balasubramanian M, Palanivel S (2017) Vision based obstacle detection on railway track. Int J Pure Appl Math 116(24):567–576
30. Manolakis DG, Ingle VK, Kogon SM (2005) Statistical and adaptive signal processing: spectral estimation. Signal Modeling, Adaptive Filtering and Array Processing. Artech House
31. Mittal S, Rao D (2017) Vision based railway track monitoring using deep learning. Comput Sci. ArXiv
32. Murthy CB, Hashmi MF, Bokde ND, Geem ZW (2020) Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—A. Compr Rev Appl Sci 10(9):3280
33. Nayak SR, Nayak DR, Sinha U, Arora V, Pachori RB (2021) Application of deep learning techniques for detection of COVID-19 cases using chest X-ray images: A comprehensive study. Biomed Signal Process Control 64:1–12
34. Rahul S, Ram Bilas P, Pradip S (2020) Automated emotion recognition based on higher order statistics and deep learning algorithm. Biomed Signal Process Control 58:101867
35. Redmon J, Farhadi A (2016) Yolo9000: Better, faster, stronger. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 6517-6525
36. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
37. Sabnis OV, Lokeshkumar R (2019) A novel objects detection system for improving safety at unmanned railway crossings. In Fifth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). IEEE, pp 149-152
38. Shangsheng Z, Jiangzhou Z, Xiaobo C, Yanqiang L (2021) Road information detection method based on deep learning. In 6th International Conference on Electronic Technology and Information Science (ICETIS 2021), Journal of Physics: Conference Series 1827
39. Sinha D, Feruz F (2016) Obstacle detection on railway tracks using vibration sensors and signal filtering using bayesian analysis. IEEE Sens J 16(3):642–649
40. Stewart GW (1973) Introduction to Matrix Computations. Academic, Cambridge
41. Sudha D, Priyadarshini J (2020) An intelligent multiple vehicle detection and tracking using modified vibe algorithm and deep learning algorithm. Soft Comput 24:17417–17429
42. Vazquez J, MazoM, LazaroJL, Luna CA, Urena J, GarciaJJ, Cabello J, Hierrezuelo L (2004) Detection of moving objects in railway using vision. In IEEE Intelligent Vehicles Symposium, pp 872-875
43. Wang Q, Gao J, Yuan Y (2018) Embedding structured contour and location prior in Siamesed fully convolutional networks for road detection. IEEE Trans Intell Transp Syst 19(1):230–241
44. Wang Q, Gao J, Li X (2019) Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes. IEEE Trans Image Process 28(9):4376–4386
45. Ward JH (1963) Hierarchical grouping to optimize an objective function. JASA 58:236–244
46. Wu Y, He X, Nguyen TQ (2017) Moving object detection with a freely moving camera via background motion subtraction. IEEE Trans Circuits Syst Video Technol 27(2):236–248
47. Zabalza J, Ren J, Zheng J, Han J, Zhao H, Li S, Marshal S (2015) Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging. IEEE Trans Geosci Remote Sens 53(8):4418–4433
48. Zhang M, Hu X, Xu G (2018) An adaptive restoration method for motion-blurred image based on Wiener filtering. In Fifth Conference on Frontiers in Optical Imaging Technology and Applications, 10832
49. Zhou X, Yang C, Yu W (2013) Moving object detection by detecting contiguous outliers in the low-rank representation. IEEE Trans Pattern Anal Mach Intell 35(3):597–610

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.