VADAR: A Vision-based Anomaly Detection Algorithm for Railroads

David Breuss*, Maximilian Götzinger†, Jenny Vuong‡, Clemens Reisner‡, and Axel Jantsch†
*Institute of Computer Technology TU Wien, 1040 Vienna, Austria
†Christian Doppler Laboratory for Embedded Machine Learning, Institute of Computer Technology
TU Wien, 1040 Vienna, Austria
‡Mission Embedded 1100 Vienna, Austria
david.breuss@tuwien.ac.at

***Abstract*—Detecting damages and anomalies on railroads is a tedious and expensive task. This paper proposes the Vision-based Anomaly Detection Algorithm for Railroads (VADAR), which can find rail damages and foreign objects on the trackbed in monochrome images captured by a train-mounted camera system. VADAR analyzes the input image with three Autoencoders (AEs), a segmentation network, and a one-class classifier. The detection of unknown anomalies justifies our architecture's advantage, i.e., no anomalies are necessary for training VADAR. In experiments with a dataset of over 218,000 images, VADAR achieves a detection accuracy of 95% and a recall rate of 70% for smaller and up to 100% for bigger instances of several anomaly classes. Compared with a state-of-the-art approach which is based on more expensive equipment, VADAR achieves accuracy and recall rates (for anomalies of particular interest) of about 22pps and up to 45pps higher, respectively. With a setting that achieves 83.5% accuracy, VADAR's recall rate outperforms the state-of-the-art approach for every anomaly class and object size.**

***Index Terms*—Railroad; Vision-Based; Anomaly Detection; Unknown Anomalies; Autoencoder**

I. Introduction

The railroad transport of people and goods is essential to today's urbanized society [1], but track installations wear out over time due to usage [2]. In addition, factors such as climatic influences or intentional damage also deteriorate their condition [3]. To ensure smooth rail traffic, i.e., to avoid breakdowns and – even worse – accidents, track systems must be maintained regularly. This maintenance work includes an inspection of the rails and the trackbed as well as the preparation and, if necessary, renewal of these components [4]. Experts estimate the annual maintenance costs at around 50,000 EUR/km [5], corresponding to the 15 to 25 billion EUR budget reported by the EIM-EFRTC-CER Working Group [6]. Jovanović *et al.* estimate that a 15-55% reduction is achievable through improved and more predictive maintenance [7].

Today, two methods are commonly used to inspect railroad tracks: (i) trained personnel who inspect the infrastructures only superficially but continuously, and (ii) a more thorough but infrequent inspection by slow and expensive measuring vehicles [8], [9]. This inspection alone incurs costs of about 70 million EUR per year for the 370,000 km long railroad tracks in the European Union [10]. Besides these exorbitant costs, the current practice has another severe disadvantage. Any problems, such as rail wear-outs or loose objects in the trackbed, require prompt detection and fast remedial action to avoid more severe damages, breakdowns, or accidents. However, this would require frequent inspections integrated into daily rail traffic to find anomalies in the railroad track.

Since railroad track damages and other anomalies, such as foreign objects, are usually visible [11]–[13], it constitutes a suitable application for an automated anomaly detection system based on computer vision. Wang *et al.* investigated an unsupervised method for detecting anomalies on rail tracks [14]. However, their analysis excludes rail damages, and their camera system only detects objects lying directly next to the rails. Boussik *et al.* analyzed the performance of multiple AE models on railroad obstacle detection [15]. Their analysis relies mainly on images with synthetic anomalies, except for only one single real-life scenario. Since their work focuses on railroad obstacles, they did not analyze the performance for detecting smaller objects and rail damages. Gasparini *et al.* proposed an approach to observing the railroad track based on an AE [13]. Their approach seems promising because the underlying architecture would allow the detection of unknown (i.e., unlearned) anomalies. However, their system's decision-making process entirely relies on a classification network trained with anomaly data, canceling out the advantage of its architecture. Moreover, they used a significantly limited dataset for validation and only searched for large foreign objects (like pickaxes or traffic lights), which they intentionally placed on the track. They omit rail damages which can lead to severe accidents, and smaller objects, such as dead animals, which can attract bigger animals, possibly colliding with the train. Moreover, they acquired their images with multiple RGB- and infrared cameras mounted on a drone flying close above the tracks, whereas we aim at train-mounted equipment.

Our paper's main contribution is to propose VADAR, consisting of three AEs, one segmentation network, and a one-class classifier. This architecture is advantageous compared to a state-of-the-art approach [13] as it allows the detection of anomalies that (i) are entirely unknown and (ii) include,

This work was supported in part by the Austrian Federal Ministry for Digital and Economic Affairs, in part by the National Foundation for Research, Technology and Development, and in part by the Christian Doppler Research Association.

besides foreign objects on the trackbed, rail damages.

With the help of the experiments' results, based on a dataset[1] of about 218,000 images with different-looking railroad tracks containing over 56,000 infrastructure elements (e.g., switches or sensors) and 10,000 non-intentionally placed anomalies, we prove that images captured from a cost-effective monochrome camera mounted on a regular train are sufficient for a detection accuracy of 95% while achieving a recall rate of over 70% for smaller and up to 100% for bigger objects of several anomaly classes. Compared with the state-of-the-art approach [13], equipped with a more expensive RGB camera and tested on a dataset a magnitude smaller than ours, our achieved accuracy and recall rate (for anomalies of particular interest) are about 22pps and up to 45pps higher, respectively.

The scope of this paper, including all experiments and results, refers to a camera setup analyzing images from a monochrome camera installed under the train car. Figures 4a and 4e are examples of such top-down images showing the rails, the railroad crossties, ballast, and, in the case of Figure 4a, a dead animal (an anomaly). The camera system only recorded images on rail tracks in Austria and Switzerland. The dataset consists of images under non-systematically varying lighting conditions. No images of this dataset show rainy, snowy, or foggy conditions.

II. Background and Related Work

A. Railroads Infrastructure Monitoring

Over the years of rail transport, rail facilities have evolved, and so have inspection and maintenance methods [16]–[18]. Two current state-of-the-art methods for railroad infrastructure monitoring exist: continuous albeit superficial on-site monitoring by trained personnel and a more thorough examination via measurement vehicles [8], [9]. Usually, the last-mentioned are used in most developed countries [19]. Modern vehicles use measurement techniques, such as ultrasound or methods based on light section or eddy-current [8], [17]. They are multifunctional measuring but cost-intensive and inspect the entire railroad network only in various intervals ranging from a few weeks to a year, depending on the respective rail section, i.e., more frequented routes are monitored more often than sections with less traffic. However, many problems, e.g., a broken track or an object on the trackbed, require prompt action to avoid damage and fatalities. These vehicles' long measurement intervals render early problem detection impossible [9].

Two solutions are suggested in the existing literature to tackle this issue: (i) installing sensors on the track infrastructure [20], [21], and (ii) integrating sensors on moving vehicles, such as locomotives or cars [11], [13]. The advantage of direct integration in the infrastructure itself is the possibility of continuous on-site monitoring. However, these solutions generally entail high costs for the required sensors and the communication infrastructure [22]–[25], even with the outlook of advanced wireless transmission standards, such as 5G [26].

[1]Since the dataset is not our property, we are not allowed to publish or share it. In Section IV, we describe the dataset to better discuss our results.

B. Computer Vision Based Anomaly Detection

Sometimes referred to as novelty detection, anomaly detection is the identification process of new or unknown data or signals that were not known during training [27]. Existing literature defines anomalies as observations that deviate considerably from some concept of normality [28] and divides anomaly detection algorithms into probabilistic methods and reconstruction-based models [13]. Probabilistic methods assume data follows an underlying probability density function [27]. In contrast, reconstruction-based approaches, such as AEs [29], [30], or Generative Adversarial Networks (GANs) [31], learn features from regular training data (images without anomalies) that are useful for representing regular data. In other words, such a trained model cannot sufficiently reproduce an anomaly present in an input image. Reconstruction-based approaches like AEs have shown promising results in image anomaly detection due to their ability to effectively represent high-dimensional data within a low-dimensional latent representation [32]. In railroad scenarios, some anomaly detection approaches have already been proposed that specialize in detecting certain defects or damages. Li et al. propose an electromagnetic thermography system for detecting certain rail damages [33], while [34] proposes a camera system for detecting railroad plug defects. Other works [35], [36] propose a camera and 3D camera system for inspecting railroad fasteners, respectively.

In [13], Gasparini *et al.* propose a vision-based AE approach to inspect railroad systems at night. Although their approach could allow the detection of unknown anomalies, their system uses a classifier network needing supervised training of anomalies which cancels out the advantage of the AE-based architecture. Moreover, their system only focuses on detecting and localizing foreign objects on the trackbed and omitting rail damages. Specifically, they exclusively focused their analysis of this anomaly detection approach on ten classes of construction site tools they intentionally placed on the trackbed. They analyze a quite limited dataset captured from a drone equipped with an RGB and an infrared camera in combination with an artificial light source. Our proposed system, VADAR, is also based on AEs. Section VI-A compares the performances of both systems.

C. Public Datasets

To our knowledge, three annotated railroad datasets exist:

***RailSem19*:** Zendel *et al.* propose *RailSem19* in one of their works [37]. It is a public dataset for semantic scene understanding for trains and tramways, consisting of 8,500 annotated short sequences recorded out of a train from an ego perspective. While it contains over 1,000 instances with railroad crossings and 1,200 tram scenes, it does not contain any anomalies, which is essential for testing VADAR.

***Kaggle Railway Track Fault Detection*:** The *Kaggle Railway Track Fault Detection* [38] dataset consists of 384 images showing rails and other railroad infrastructure, half of which contain anomalies. These images show entirely different perspectives (closeups from all possible sides). Therefore, they

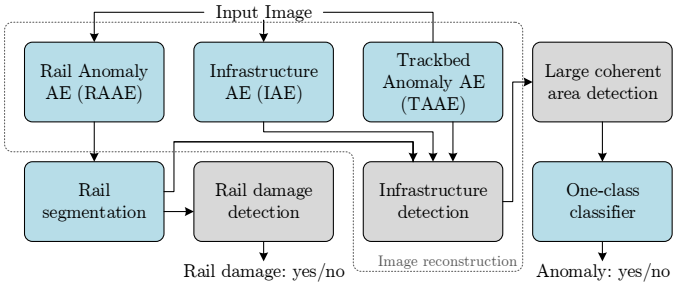


Fig. 1: The block diagram of VADAR. The blue blocks are neural network based models and the white blocks refer to other processing steps.

misfit the desired application of VADAR, detecting anomalies from a moving train.

***Vesuvio*:** Gasparini *et al.* recorded the *Vesuvio* dataset for their work [13]. They shot the sequences at night with a drone flying just above the tracks using a thermal camera, a stereo camera system, and an industrial RGB camera. To test their system, which shall detect foreign objects, they intentionally placed big objects on the track, such as a pickaxe, a traffic light, and an LPG tank. The recordings were then manually annotated with bounding boxes. It is a magnitude smaller than our recorded dataset and lacks rail damages and smaller foreign objects. Moreover, it is not publicly available.

III. Autoencoder-based Approach

VADAR (Figure 1) analyzes Reconstruction Error (RE) images based on the outputs of three different AEs. In this context, an RE image is defined as

$$I_{RE} = |I_{orig} - I_{recon}|, \tag{1}$$

where I_{orig} is an input-, and I_{recon} is a reconstructed image.

We use an additional rail-segmentation network to separate the reconstruction errors of the rails from the rest of the trackbed. This procedure enables a distinction between rail damages and foreign objects or vegetation on the trackbed. Within this section, C_{in}, C_{out}, k, s, and p refer to the number of input channels, output channels, the kernel size, stride, and padding of layers, respectively.

A. Rail Segmentation

The random nature of the ballast on the railroad tracks seems to limit the size of detectable anomalies. To overcome this limitation and also detect small rail damages, we separately analyze the rails' REs where typically no ballast is.

The accuracy of this segmentation method depends on the lighting conditions. A quality check then excludes implausible segmentation outputs by taking a convolution ρ of a predefined ground truth shape M_{truth} and the segmentation output M_{seg} and the sum σ of all pixels M_{seg} into account:

$$\rho = \max\left(M_{truth} * M_{seg}\right),\ \sigma = \sum M_{seg},\ Q = \frac{\rho}{\sigma}. \tag{2}$$

If the quality metric Q falls below a predefined threshold, the segmentation output is considered faulty, and the rail damage detection is not computed for this image.

	Encoder					Decoder				
Layer	C_{in}	C_{out}	k	s	p	C_{in}	C_{out}	k	s	p
1	1	4	7	1	3	64	64	7	1	3
2	4	8	7	2	3	64	64	7	1	3
3	8	16	7	2	3	64	64	7	2	3
4	16	32	7	2	3	64	64	7	2	3
5	32	64	7	2	3	64	64	7	2	3
6	64	64	7	2	3	64	32	7	2	3
7	64	64	7	2	3	32	16	7	2	3
8	64	64	7	2	3	16	8	7	2	3
9	64	64	7	1	3	8	4	7	2	3
10	64	64	7	1	3	4	1	7	1	3

TABLE I: The encoder and decoder of the TAAE and IAE consist of ten two-dimensional convolutional and transposed convolutional layers, respectively. The column names are defined at the beginning of Section III.

B. Image Reconstruction

Two of the three AEs are part of detecting anomalies in the trackbed: the Trackbed Anomaly Autoencoder (TAAE) and the Infrastructure Autoencoder (IAE). Both AEs are based on an encoder with ten convolutional layers and a decoder with ten transposed convolutional layers (Table I). The TAAE is exclusively trained with regular images to generate larger reconstruction errors for anomalous data. In contrast, the IAE, exclusively trained with images containing infrastructure elements, can better reconstruct such images while simultaneously achieving almost identical reconstructions of regular images and images containing anomalies. A pixel-wise comparison of the RE images of both AEs accomplishes the *Infrastructure detection* (Figure 1). Pixel values with differences greater than a predefined threshold imply that they belong to an infrastructure element and are, therefore, ignored.

To detect rail damages, the Rail Anomaly Autoencoder (RAAE) possesses a slightly different architecture (Table II) which significantly improves accuracy and false positive rate for rail damage detection (Figure 6). Since the RAAE could sufficiently reconstruct larger foreign objects on the trackbed, we only use it to detect rail damages. Figures 4b and 4f show the reconstruction images of the TAAE and RAAE, respectively.

C. Rail Damage Detection

The binary rail mask (Section III-A) enables VADAR to analyze the REs of the rail heads to detect rail damages VADAR only considers pixel values larger than a defined threshold to belong to an anomaly. A second threshold value for the minimum summed-up anomaly value allows for decreasing the false positive rate by ignoring smaller REs.

D. Trackbed Anomaly Detection

Unusually bright or dark stones of the ballast can lead to significant reconstruction errors. Thus, only large coherent areas (bigger than a threshold defined) within the RE image are considered anomalies. The output of this procedure is a binary image based on the RE image. Since factors like lighting

Layer	Encoder					Decoder				
	C_{in}	C_{out}	k	s	p	C_{in}	C_{out}	k	s	p
1	1	16	3	1	1	32	32	3	2	1
2	16	32	3	2	1	16	32	3	2	1
3	32	32	3	2	1	32	32	3	2	1
4	32	32	3	2	1	32	32	3	2	1
5	32	32	3	2	1	32	32	3	2	1
6	32	32	3	2	1	32	32	3	2	1
7	32	32	3	2	1	32	16	3	2	1
8	32	32	3	2	1	16	1	3	1	1

TABLE II: The encoder and decoder of the RAAE consist of eight two-dimensional convolutional and transposed convolutional layers, respectively. We added a batch-normalization layer before each of them. The column names are defined at the beginning of Section III.

conditions and ballast influence the total RE, the threshold value θ is defined as

$$\theta = \max\left(\text{torch.quantile}(I_{RE}, q), \theta_{min}\right), \tag{3}$$

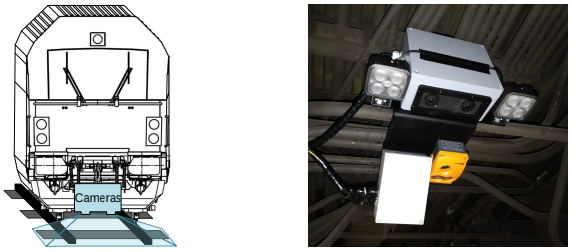
where the "torch.quantile"-function, provided by the PyTorch framework [39], returns the q^{th} quantile of I_{RE}. If the overall RE is unusually low due to bad lighting, a minimum value θ_{min} is used instead. The binary image results from setting the pixel values to 1 if the corresponding pixel values of the RE image are greater than the threshold; all other pixels are set to 0. The *Large coherent area detection* (Figure 1) utilizes the "regionprops" function of the Python package "skimage.measure" [40] to determine the largest area within the obtained binary mask. A coherent area greater than a certain threshold is considered a potential anomaly and further analyzed by a one-class classifier network. A significant contribution to the total number of false positives comes from groups of brighter or darker stones from the ballast. Therefore, an additional *one-class classifier* (Figure 1) distinguishes gravel from other objects to reduce the false positive rate. A 64-by-64 pixels big patch of the original image is the input of this one-class classifier. This image patch either includes the total detected coherent error area or parts of it. Table III shows the architecture of this classification network. If no anomalous samples are available for VADAR's training the *Large coherent area detection* output indicates anomalies and the one-class classifier is not used.

IV. Dataset used

The dataset was collected in 2019 during a project conducted in collaboration with the ÖBB (Austrian rail operator). This project aimed to define and determine a cost-efficient sensor system installable on regular trains to monitor railroad infrastructure. The sensors must fulfill the following constraints:

Standards: All parts of the system must satisfy the standards prevalent in the railway industry to be permitted for usage, i.e., homologation by official authorities [41].

Size and energy consumption: Sensors should not exceed a specific size so that regular train carriages can easily be retrofitted; they must either fit in and/or outside the train.



(a) BEV setup (b) Real BEV setup

Fig. 2: Two stereo cameras installed underneath the train, pointing at the tracks

Besides, the system's power consumption must comply with certain thresholds since trains have limited power.

Speed and external conditions: Components of the system, especially those installed on the outside of the train, must withstand the maximum operating train speed (varying from 100-230 km/h) while providing meaningful data taken under different weather and light conditions.

Cost efficiency: System costs are an essential factor since economic efficiency allows rail operators to retrofit as many trains as possible and thus increase the coverage of monitored rail tracks in the network.

Following all this guidance, two 3.2-megapixels monochromatic global shutter cameras for high-speed image acquisition were installed underneath the train carriage (Figure 2). The cameras, configured for stereo vision with a 20 cm baseline, recorded the tracks in a bird's-eye view (BEV), one stereo image pair for each meter. The data recording was conducted over three weeks and covered around 300 km of railroads in Austria and Switzerland.

A. Dataset Statistics

The annotations consist of bounding boxes created with *labelimg* [42]. Since the evaluation of VADAR should also consider the anomalies' sizes, the annotations' precision was of great importance. From the approximately 2,000,000 image dataset, we inspected and annotated about 218,000 images.

Layer	C_{in}	C_{out}	k	s	p
Convolutional 1	1	32	3	1	1
Convolutional 2	32	32	3	1	1
MaxPool	32	32	2	2	0
Convolutional 3	32	64	3	1	1
Convolutional 4	64	64	3	1	1
AdaptiveAveragePool(2,2)	64	64	-	-	-
Flatten	-	-	-	-	-
Linear 1	256	64	-	-	-
Linear 2	64	32	-	-	-
Linear 3	32	1	-	-	-

TABLE III: We added a batch-normalization layer before each convolutional and linear layer of the one-class classifier. Instead of the rectified linear unit (ReLU) activation function, the last linear layer is followed by a sigmoid function. The column names are defined at the beginning of Section III.

Class	Instances	Class	Instances
box	2,900	sensor	370
crosstie attachment	23,965	spacer	674
different rail	3,468	switch	13,313
other	3,684	switch frog	1,075
rail attachment	4,720	switch positioner	2,123

TABLE IV: The group of infrastructure elements consists of ten different classes.

Two categories of annotations exist: infrastructure elements and anomalies. In this context, infrastructure elements do not mean a rail or a crosstie but rather relatively less frequently occurring elements, such as a switch, a switch frog, or various sensors. Table IV gives detailed information about the different infrastructure annotations. The anomaly annotations are also divided into two groups: damages (e.g., rail damages) and foreign objects (e.g., a dead animal). Figure 3 gives information about the size distribution of damages, vegetation, and foreign objects. As a reference, 1,000 pixels correspond to roughly 0.13% of an image's pixels. Most vegetation that is greater than 10,000 pixels includes several smaller plants across the entire trackbed. The labeled rail damages in this dataset differ in severity and contain a variety of damages like break-outs, skid spots, and even some small-scale rail damages like indentations [43]. There are 320 additional rail anomalies larger than 10,000 pixels. These anomalies are corrugations and typically appear on multiple consecutive frames. Furthermore, each frame's type of crossties, type of ground, and ambient lighting conditions are annotated. Table V lists the number of samples for each scenario.

Railroad Crossties		Ground Types		Ambient Lighting	
Class	Instances	Class	Instances	Class	Instances
Concrete	185,128	Gravel	226,845	Daylight	227,311
Wood	42,552	Railroad Crossing	2,255	Dark	2,324
Mixed	1787	Mixed	466	Mixed	8
Metal	74	Bridge	50		
Hardrubber	26	Asphalt	9		
None	8	Unknown	18		
Unknown	68				

TABLE V: Classes defined in the BEV dataset and their corresponding numbers of images.

V. Training

All neural network models are implemented with Cuda 11.3 [44] within the PyTorch [39] machine learning framework (Version 1.11.0). We train all models on an *NVIDIA A100 40GB* [45] Graphics Processing Unit (GPU) and feed batches of 32 samples to them. We use the *Adam* optimization algorithm [46] with a learning rate of 10^{-3}, and the β-values are 0.9 and 0.999. The labeled images are divided into 80% training and 20% validation data. For all models except the one-class classifier network, only images free of anomalies are part of the training process.

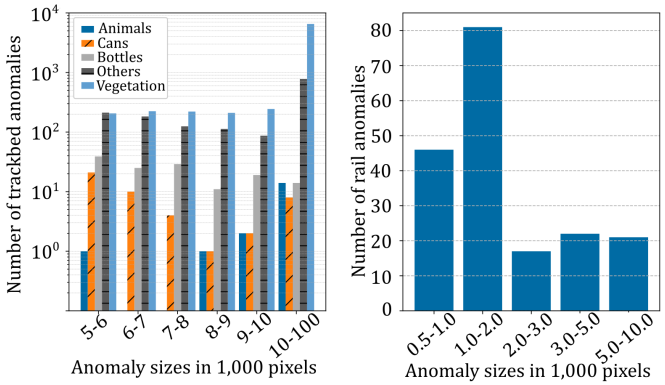


Fig. 3: Size distributions of trackbed- and rail anomalies.

A. Training the Autoencoders

For training the TAAE and RAAE, 18,191 instances of the labeled portion of the dataset are used. As proposed by Gasparini *et al.* [13], we use a loss function that includes both the mean squared error loss $L_{I,MSE}$ between the input image and its reconstruction as well as the mean squared error loss $L_{G,MSE}$ between the gradients of the input image and its reconstruction. The overall loss L is then defined as

$$L = L_{I,MSE} + L_{G,MSE}. \tag{4}$$

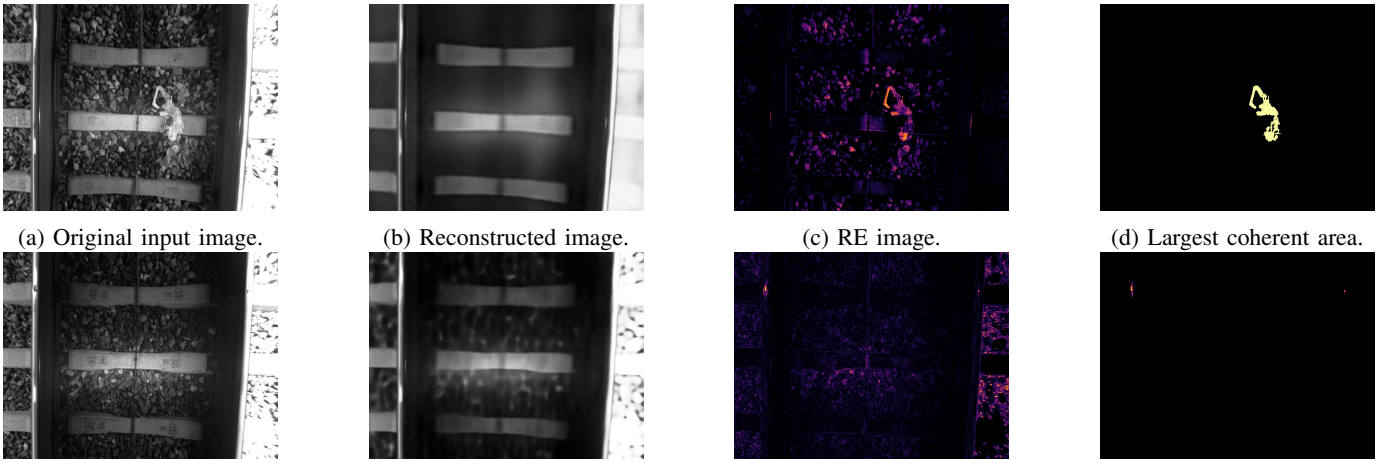
In contrast to the training of the TAAE, the training data of the IAE only includes images containing infrastructure elements. In total, 19,402 images were used for a 100-epochs-long training, applying the loss function described in Equation 4 of both AEs.

B. Training the Rail Segmentation

Since the rail segmentation network model utilizes the same network architecture as the RAAE, it was initialized with the already trained parameters of the AE model. We then used a transfer learning approach to train it for 40 epochs with 12,573 images of different parts of the dataset and corresponding manually created rail masks as the ground truth data. A manually created rail mask fits hundreds of consecutive images of a railroad track because the positions of the rails are constant. In contrast to the AE training procedure, the binary cross-entropy loss was applied as a loss function.

C. Training the One-Class Classifier

We trained the classifier with 6,000 regular images (containing only ballast) and 6,004 anomalous (containing foreign objects) patches. The regular patches were extracted from 200 labeled images of different lighting conditions without any anomalies or infrastructure elements. The anomalous patches were obtained from roughly 30% of the available anomalous images. From larger anomalous objects, multiple patches were extracted. We trained the model for 100 epochs and used the binary cross-entropy loss as the loss function.



(a) Original input image. (b) Reconstructed image. (c) RE image. (d) Largest coherent area.

(e) Original input image. (f) Reconstructed image. (g) RE image. (h) Segmented RE.

Fig. 4: These figures show a detected dead animal (first row) and a detected rail damage (second row). VADAR ignores the light reflection (in the lower left part of Figure e), which is no damage.

VI. Evaluation

We used the labeled part of the dataset to evaluate several performance metrics of this anomaly detection approach. The accuracy is defined as

$$acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{5}$$

and the false positive rate and recall rate are defined as

$$fpr = \frac{FP}{TN + FP}, \quad rec = \frac{TP}{TP + FN} \tag{6}$$

respectively. Whereby TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. Since the data obtained from a camera with a wide-angle lens leads to distorted edge areas after image transformation, we cropped 768-by-1024 pixels big center of each image. Although the camera system consists of two cameras (Section IV), this approach only considers one camera's images since both are almost congruent. Figure 4 shows two original images (one with an animal, the other with a damaged rail) and their corresponding reconstructed-, RE-, and output images.

Because rail damages and foreign objects differ in their size, the sizes of the bounding boxes are part of the evaluation. After the cropping procedure, some annotations may disappear from certain images and, thus, are ignored in the evaluation process.

A. Results

The lines between markers in all result graphs (Figures 5, 6, and 7) do not represent data points but visualize the trends. Changing the threshold value for the anomaly pixel values of the trackbed anomaly detector influences the accuracy, false positive rate, and recall rate. Figure 5 shows the results for anomalies with bounding boxes larger than 5,000 pixels, which corresponds to roughly 0.6% of the total input image pixels. Five separate experiments on the whole labeled dataset were conducted for different threshold parameters q and θ_{min}. The threshold value is defined as the maximum of the q^{th} quantile of I_{RE} and the minimum threshold value θ_{min}, as described in Equation 3. Discussions with railroad maintenance experts revealed that a low false positive rate should be targeted to increase acceptance among maintenance personnel. Figure 5 shows the improvements in accuracy and false positive rate when instead of solely using the TAAE, the IAE and TAAE are used to enable the infrastructure detection. While there are only small changes in recall rates in a few cases, the accuracy and false positive rate are improved significantly when two AEs are used. This anomaly detection approach relies on the intensity difference between the anomalous object on the trackbed and the trackbed itself. The lighting conditions affect this intensity difference and have an impact on the performance of the anomaly detector. Not every kind of vegetation is detectable by this approach. Besides the intensity difference between the plants and the elements of the trackbed, the vegetation's density is also an important factor. Because every kind of plant was labeled in this dataset, the majority of vegetation is either small or sparse. Every kind of object that did not fit into one of the other classes was labeled as "others." Most of these objects are small pieces of trash or dark pieces of wood. Because such objects and small and sparse vegetation are not of special interest for maintenance, the lower recall rates should be tolerable for most railroad maintenance applications. Through the proposed infrastructure detection approach, only roughly 6% of all images containing infrastructure elements without an anomaly lead to false positives. The classifier network further reduces the number of false positives resulting from regular images by 37%, allowing for false positive rates of under 1% while decreasing the overall recall rates of trackbed anomalies by at most 7pp.

For the rail anomaly detection evaluation, only rail head damages like break-outs, skid spots, or indentations were

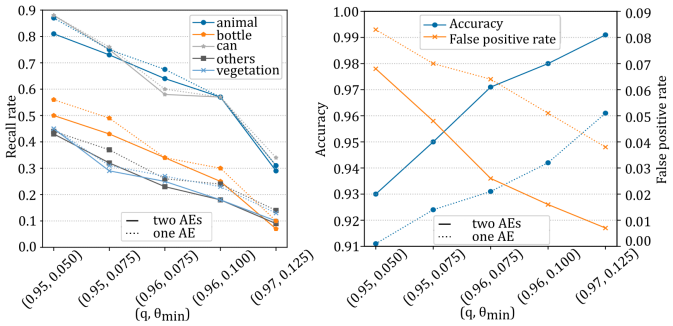


Fig. 5: Solid lines represent the results obtained using TAAE and IAE, whereas dotted lines show the usage of solely TAAE. Increasing the threshold value q and θ_{min} decreases the false positive rate and recall rates.

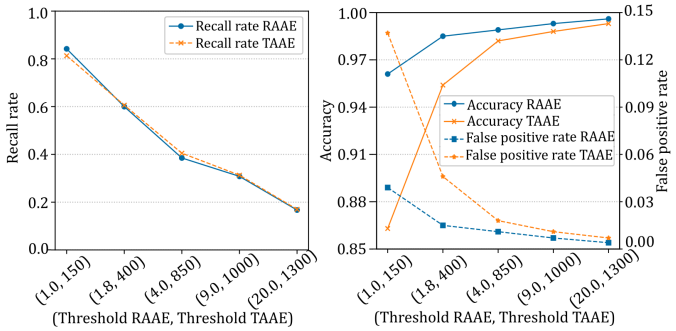


Fig. 6: While achieving the same recall rate, using the TAAE instead of the RAAE leads to a worse overall accuracy and false positive rate. Increasing the summed-up rail RE threshold decreases the false positive and recall rate.

considered. Labeled damages to the sides of the rails are ignored since the rail segmentation network only includes the rail heads in its output. Although the total sum of the thresholded reconstruction error is considered, the size of the damage is also an important factor. The bounding boxes' size was used to approximate the damages' size, and only damages larger than 600 pixels were considered, which corresponds to roughly 0.08% of the total input image pixels. Figure 6 shows five separate experiments with different threshold values and clarifies why a separate RAAE is used. The different architecture of the RAAE leads to significant improvements in overall accuracy and false positive rate while maintaining the same recall rate as the TAAE on the rail damage detection task. The annotations for rail damages do not distinguish between different types of rail damages, but an overwhelming majority of labeled rail anomalies are minor damages. The larger the threshold value, the lower the recall rate on such smaller damages. However, relatively large damages like a break-out (Figure 4e) lead to unusually high reconstruction errors, especially under advantageous lighting conditions. Because this specific break-out led to an anomaly value of 78.8, this damage would be detectable for thresholds up to this value. These more severe damages are reliably detectable while maintaining a false positive rate well below 1%. Gasparini *et al.* [13] analyzed the performance of their approach on the *Vesuvio* dataset. Since this dataset is not published, we implemented their system and tested it on the BEV dataset for comparison. Figure 7 shows the recall rates for various anomaly classes and object sizes for their approach and for two different settings of our proposed approach. The recall rates increase significantly with the object size. The setting of VADAR, with an accuracy of 83.5%, outperforms the Gasparini approach regarding accuracy and recall rate for every anomaly class and object size. For a different setting, where VADAR achieves an even higher accuracy of 95.0%, it is only outperformed by Gasparini's method regarding the recall rate for the vegetation class. Probably the reason for that is that their supervised approach benefits from the strongly overrepresented vegetation instances within the BEV dataset.

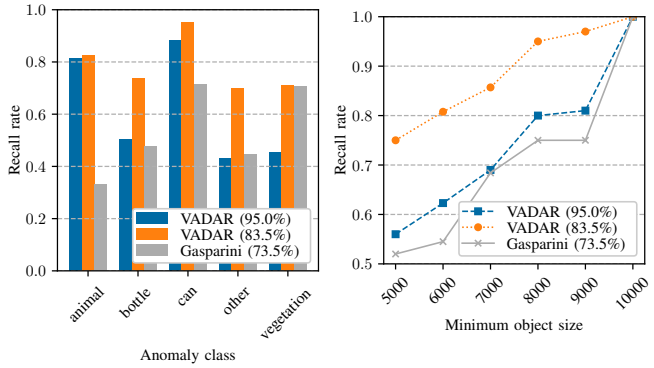


Fig. 7: The Gasparini *et al.* [13] method's accuracy reaches 72.3%, while VADAR's accuracy, in different setups, is 95.0% and 83.5% (with only higher recall rates). The recall rate for foreign objects increases with their size.

VII. Conclusions

Railroad maintenance can benefit from a cost-effective vision-based anomaly detection system integrated into daily rail traffic. The top-down perspective of the camera system and our approach, VADAR, enable the detection of foreign objects and early-stage rail damages. The false positive rate caused by infrastructure elements and the ballast is significantly reduced through VADAR's architecture, including three AEs, a rail segmentation network, and a one-class classifier. While achieving a detection accuracy of more than 95%, VADAR reaches a recall rate for rail damages of more than 80% and for objects of special interest, like animals, greater than 70%. The recall rate for animals, bottles, and cans bigger than 10,000 pixels (bounding box size), which fills roughly 1.3% of the total input image, reaches even 100%. The trade-off between the overall accuracy and recall rate could be varied and fine-tuned for a specific application. When focusing on larger objects and more severe rail damages like breakouts, a false positive rate of even 1% is achieved.

Although this approach was designed with a gray-scale dataset in mind, the same approach could also be applied

to color images. Analyzing all color channels might further improve the recall rate of anomalies with a different color than the gravel and fasteners. Besides, improving the lighting conditions with stronger artificial lighting systems or a different camera position or perspective could further improve the overall detection accuracy. Moreover, utilizing the images of both cameras of the camera system could enable another anomaly detection approach based on stereo vision.

REFERENCES

[1] European Rail Supply Industry Association. Establishing rail as the backbone of future mobility. 24(5), 2018. 1

[2] Mats Andersson. Marginal cost of railway infrastructure wear and tear for freight and passenger trains in sweden, 2010. 1

[3] Michael A Rossetti. Potential impacts of climate change on railroads. In *The Potential Impacts of Climate Change on Transportation: Workshop Summary*, 2002. 1

[4] Tomas Lidén. Railway infrastructure maintenance - a survey of planning problems and conducted research. *Transportation Research Procedia*, 10, 2015. 1

[5] Coenraad Esveld and Coenraad Esveld. *Modern railway track*, volume 385. MRT-productions Zaltbommel, 2001. 1

[6] EIM-EFRTC-CER Working Group on Market Strategies for Track Maintenance & Renewal. Report from the eim-efrtc-cer working group on market strategies for track maintenance & renewal. 2012. 1

[7] Stanislav Jovanović, Dragan Božović, and Mirjana Tomičić-Torlaković. Railway infrastructure condition-monitoring and analysis as a basis for maintenance management. 2014. 1

[8] Guido HANSPACH. Ust 02: Schienenprüfzug der neuen generation für die europäischen bahnen. *Der Eisenbahningenieur (Hamburg)*, 57(1):26–28, 2006. 1, 2

[9] Haoyu Wang, Jos Berkers, Nick van den Hurk, and Nasir Farsad Layegh. Study of loaded versus unloaded measurements in railway track inspection. *Measurement*, 169:108556, 2021. 1, 2

[10] Zdenka Popović, V. Radovic, Luka Lazarević, V. Vukadinovic, and G. Tepić. Rail inspection of rcf defects. *Metalurgija -Sisak then Zagreb-*, 52:537–540, 10 2013. 1

[11] Xavier Gibert, Vishal M. Patel, and Rama Chellappa. Deep multitask learning for railway track inspection. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):153–164, 2017. 1, 2

[12] Roger Nyberg, Narendra Gupta, Siril Yella, and Mark Dougherty. Monitoring vegetation on railway embankments : Supporting maintenance decisions. 06 2013. 1

[13] Riccardo Gasparini, Andrea D'Eusanio, Guido Borghi, Stefano Pini, Giuseppe Scaglione, Simone Calderara, Eugenio Fedeli, and Rita Cucchiara. Anomaly detection, localization and classification for railway inspection. In *ICPR*. IEEE, 2020. 1, 2, 3, 5, 7

[14] Tiange Wang, Zijun Zhang, and Kwok-Leung Tsui. A deep generative approach for rail foreign object detections via semisupervised learning. *IEEE Transactions on Industrial Informatics*, 19(1):459–468, 2023. 1

[15] Amine Boussik, Wael Ben-Messaoud, Smail Niar, and Abdelmalik Taleb-Ahmed. Railway obstacle detection using unsupervised learning: An exploratory study. In *IEEE Intelligent Vehicles Symposium*, 2021. 1

[16] Leith Al-Nazer, Thomas Raslear, Carlo Patrick, Judith Gertler, John Choros, Jeffrey Gordon, and Brian Marquis. Track inspection time study. Technical report, 2011. 2

[17] William T McCarthy. Track geometry measurement on burlington northern railroad. In *Nondestructive Evaluation of Aging Railroads*, volume 2458, pages 148–164. SPIE, 1995. 2

[18] Christian Higgins and Xiang Liu. Modeling of track geometry degradation and decisions on safety and maintenance: A literature review and possible future research directions. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 232(5):1385–1397, 2018. 2

[19] Paul Weston, Clive Roberts, Graeme Yeo, and Edward Stewart. Perspectives on railway track geometry condition monitoring from in-service railway vehicles. *Vehicle System Dynamics*, 53(7):1063–1091, 2015. 2

[20] Chayut Ngamkhanong, Sakdirat Kaewunruen, and Bruno J Afonso Costa. State-of-the-art review of railway track resilience monitoring. *Infrastructures*, 3(1):3, 2018. 2

[21] K Velmurugan and T Rajesh. Advanced railway safety monitoring system based on wireless sensor networks. *International Journal of Science, Engineering and Computer Technology*, 6(2):89, 2016. 2

[22] Abhisekh Jain, Arvind Seshadri, Ramviyas Parasuraman, et al. Onboard dynamic rail track safety monitoring system. *arXiv preprint arXiv:1212.0240*, 2012. 2

[23] Jeff Stevens. Onboard monitoring aids track maintenance. *International Railway Journal*, 53(5), 2013. 2

[24] Matthew Dick. Automating geometry measurement offers real-time benefits. *International Railway Journal*, 56(8), 2016. 2

[25] Graeme James Yeo. *Monitoring railway track condition using inertial sensors on an in-service vehicle*. PhD thesis, University of Birmingham, 2017. 2

[26] Manuel Eugenio Morocho-Cayamcela, Haeyoung Lee, and Wansu Lim. Machine learning for 5g/b5g mobile and wireless communications: Potential, limitations, and future directions. *IEEE access*, 7, 2019. 2

[27] Markos Markou and Sameer Singh. Novelty detection: a review—part 1: statistical approaches. *Signal processing*, 83(12):2481–2497, 2003. 2

[28] Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection. *CoRR*, abs/2009.11732, 2020. 2

[29] Saeed Khalilian, Yeganeh Hallaj, Arian Balouchestani, Hossein Karshenas, and Amir Mohammadi. Pcb defect detection using denoising convolutional autoencoders. In *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pages 1–5, 2020. 2

[30] Laurenz Strothmann, Uwe Rascher, and Ribana Roscher. Detection of anomalous grapevine berries using all-convolutional autoencoders. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3701–3704, 2019. 2

[31] Sertac Arisoy, Nasser M. Nasrabadi, and Koray Kayabol. Gan-based hyperspectral anomaly detection. In *2020 28th European Signal Processing Conference (EUSIPCO)*, pages 1891–1895, 2021. 2

[32] Vincent Wilmet, Sauraj Verma, Tabea Redl, Håkon Sandaker, and Zhenning Li. A comparison of supervised and unsupervised deep learning methods for anomaly detection in images. *arXiv preprint arXiv:2107.09204*, 2021. 2

[33] Hao-ran Li, Yun-han Shi, Bin Gao, Xi-yuan Zhang, Gai-ge Ru, Yongsheng Shi, Yu-hua Zhang, and Long-hui Xiong. Dynamic electromagnetic thermography system for rail inspection. In *FENDT*, 2021. 2

[34] Xinyu Du, Yu Cheng, and Zichen Gu. Change detection: The framework of visual inspection system for railway plug defects. *IEEE Access*, 8:152161–152172, 2020. 2

[35] Xavier Gibert, Vishal M. Patel, and Rama Chellappa. Deep multitask learning for railway track inspection. *IEEE Transactions on Intelligent Transportation Systems*, 18(1):153–164, 2017. 2

[36] Çağlar Aytekin, Ilkay Ulusoy, Sedat Dogru, and Yousef Rezaeitabar. Railway fastener inspection by real-time machine vision. *IEEE Transactions on Systems Man and Cybernetics - Part A Systems and Humans*, 45, 01 2015. 2

[37] Oliver Zendel, Markus Murschitz, Marcel Zeilinger, Daniel Steininger, Sara Abbasi, and Csaba Beleznai. Railsem19: A dataset for semantic rail scene understanding. In *Proceedings of the IEEE/CVF*, 2019. 2

[38] Kaggle railway track fault detection. https://www.kaggle.com/datasets/salmaneunus/railway-track-fault-detection. Accessed: 2022-10-19. 2

[39] An open source machine learning framework. https://pytorch.org/. Accessed: 2022-11-11. 4, 5

[40] scikit-image: image processing in python. https://scikit-image.org/docs/stable/api/skimage.measure.html. Accessed: 2022-11-11. 4

[41] T Hoppe, G Matschke, and R Müller. Homologation of trans-european rolling stock: An integrated approach. In *WCRR*, pages 4–8, 2006. 4

[42] D Tzutalin. Labelimg. *GitHub Repository*, 6, 2015. 4

[43] Vossloh: Rail defects. https://www.vossloh.com/en/products-and-solutions/products-at-a-glance/rail-turnouts.maintenance/schienenfehler.html. Accessed: 2022-11-11. 5

[44] Pytorch: Installing previous versions of pytorch. https://pytorch.org/get-started/previous-versions/. Accessed: 2022-11-11. 5

[45] Nvida a100 tensor-core-gpu. https://www.nvidia.com/de-de/data-center/a100/. Accessed: 2022-11-11. 5

[46] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5