IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 July 2023, accepted 20 August 2023, date of publication 29 August 2023, date of current version 20 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3309814

TOPICAL REVIEW

A Review on Deep Learning Techniques for Railway Infrastructure Monitoring

MARIA DI SUMMA[1], MARIA ELENA GRISETA[2], NICOLA MOSCA[1], COSIMO PATRUNO[1], MASSIMILIANO NITTI[1], VITO RENÒ[1], AND ETTORE STELLA[1]

[1]Institute of Intelligent Industrial Technologies and Systems for Advanced Manufacturing, National Research Council of Italy, 70126 Bari, Italy
[2]Department of Mathematics, University of Bari Aldo Moro, 70125 Bari, Italy

Corresponding author: Maria Di Summa (maria.disumma@stiima.cnr.it)

This work was supported by the Italian Ministry of Economic Development (MISE).

ABSTRACT In the last decade, thanks to a widespread diffusion of powerful computing machines, artificial intelligence has been attracting the attention of the academic and industrial worlds. This review aims to understand how the scientific community is approaching the use of deep-learning techniques in a particular industrial sector, the railway. This work is an in-depth analysis related to the last years of the way this new technology can try to provide answers even in a field where the primary requirement is to improve the already very high levels of safety. A strategic and constantly evolving field such as the railway sector could not remain extraneous to the use of this new and promising technology. Deep learning algorithms devoted to the classification, segmentation, and detection of the faults that affect the railway area and the overhead contact system are discussed. The railway sector offers many aspects that can be investigated with these techniques. This work aims to expose the possible applications of deep learning in the railway sector established on the type of recovered information and the type of algorithms to be used accordingly.

INDEX TERMS Anomaly detection, deep learning, railway, review.

I. INTRODUCTION

In recent years, we have been witnessing the explosion of a phenomenon called Industry 4.0, the new industrial revolution leading to a complete digitalization of the industrial world. We are assisting in a widespread diffusion of enabling technologies. This is linked, on the one hand, to scientific progress and, on the other, to increasingly contained costs. Scientific studies have made it possible to create increasingly complex and performing systems. Economies of scale and lower production costs have allowed the use of these technologies by a large audience of users. All this is making possible what, until a few years ago, was only a science fiction utopia.

In the railway sector, safety represents a fundamental aspect that requires investing huge resources. Business and research interests are, therefore, very high. Different companies work in this sector, often with global reaches, such as Mermec [10], Tesmec [21], and NDT Technologies [11], with stakes in developing both hardware and software solutions and ad-hoc consulting services. All these solutions need to be based on non-destructive testing and, most of the time, non-contact technologies, including electromagnetic, laser, optical, and ultrasonic technologies.

Although railway safety investments have grown exponentially in recent years, there is still room for improvement and risk mitigation [17]. The solution to this issue is to improve technological equipment for supporting human operators in the decision-making process.

Nowadays, rail transport plays a crucial role for passengers and goods carriage. The railway infrastructure includes the rail area (Figure 1) and the Overhead Contact System (OCS) (Figure 2).

The rail area includes the two rail tracks, the sleepers, which lie between the rails and keep the distance between them, the joint bars joining two rails, and the railroad fasteners, which secure the rails to the sleepers. Finally, most of the rail tracks have the ballast on which the sleepers lie. Ballast is used to support the sleepers' weight, encourage water drainage, and detain the vegetation under the rails. The

The associate editor coordinating the review of this manuscript and approving it for publication was Jesus Felez.

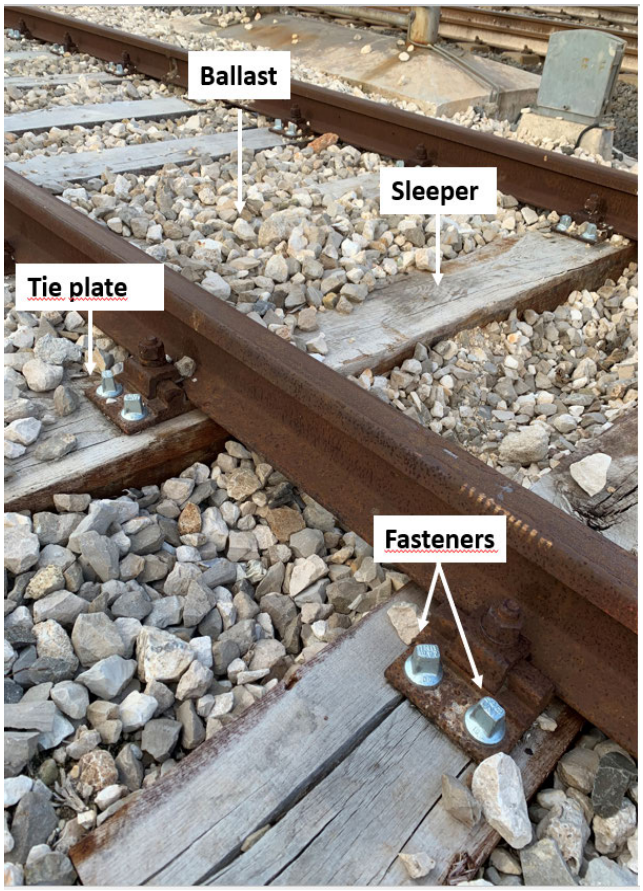


FIGURE 1. Particular of the railway infrastructure, focused on the rail area.

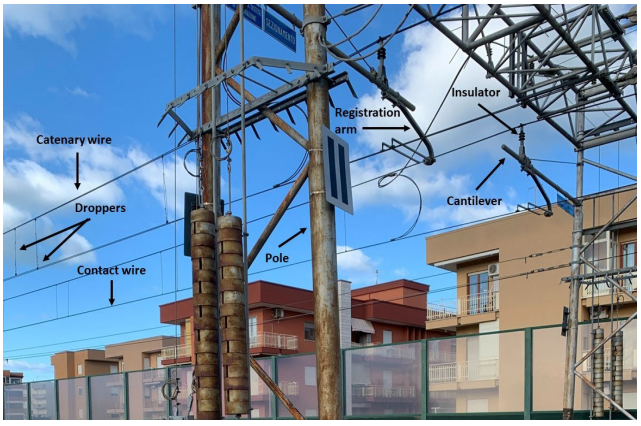


FIGURE 2. Particular of the railway infrastructure, focused on the OCS.

OCS ensures the power supply to the train: it is composed of a contact wire to which the pantograph is attached to give current to electric trains and the catenary, linked to the contact wire at symmetrical intervals by vertical cables called droppers.

The anomaly detection of railway infrastructure components is a significant challenge today since broken or missing parts on the railway track could impact the safety of the entire transport system.

Several defects can affect railway networks. Following what has been said previously, the faults can be divided mainly into two categories: defects of the rail area and deficiencies of the OCS. The first class includes the defects of the rail surface caused by the contact of the wheels on the rail surface or the temperature variations due to the heating of the track, fasteners' damage, defects of the ballast, and sleepers. The second class includes defects of the catenary, droppers, and insulators, often due to electromechanical forces or continuous exposure to the elements.

Timely detection of these faults is crucial to ensure passenger and crew safety and avoid transport delays. In the past, track inspection was performed by track patrols on foot or aboard a vehicle, but these operations were dangerous, expensive, and time-consuming. Nowadays, inspection is performed by specialized diagnostic vehicles, such as ROGER [40], comprising customized inspection, measurement, and monitoring systems manufactured to be easily personalized in every aspect.

Moreover, the data acquired during the human inspection had to be analyzed manually, and this required the use of many researchers and a very long time.

In the last decade, to optimize the time and costs for anomaly detection of railway infrastructure, researchers have begun to employ artificial intelligence (AI) techniques. First, they considered Machine Learning (ML), a field of AI that studies algorithms learning from data rather than explicit programming. These algorithms, such as Singular Value Decomposition (SVD), Principal Component Analysis (PCA), Kernel Principal Component Analysis (KPCA), or Histogram Match (HM), have the objective of instructing a machine to perform a task based on specific characteristics and examples selected by a human user; for example, in [45] the authors present an experimental study using different feature extraction techniques. In [15] the authors suggested a method designed to detect the flaws of partially worn or completely missing fasteners. The work in [31] proposes a hierarchical approach with model ensembles to improve the detection of a large-scale rail defect by observing the relationship between rail faults and track geometry anomalies. In [68], the authors identify rail surface cracks and recognize their edges by means of a bi-layer data-driven method (BDF). To find the positioning and the size of rail-head defects, in [29], ML architectures such as wavelet scattering networks and neural networks are proposed. In [56] the authors first developed a classification method to inspect rail surface faults or rail components using acceleration data acquired from an inspection vehicle. Moreover, they further propose a Convolutional Neural Network (CNN) for detecting rail joints or defects. Then a discrete wavelet function is used too. Finally, a Deep Learning (DL) approach is employed to detect rail joints or defects, by exploiting ResNet and a Fully Convolutional Network (FCN). In such a way, the performance of the classification models is

improved. The shift in used techniques highlights a general trend in the last years: researchers are showing increasing attention to using DL techniques for performing several tasks, such as classification, segmentation, or detection. DL is a field of ML that utilizes several architectures composed of artificial neural networks with many layers that can extract features gradually at different levels.

In this paper, we present a survey of various articles in which the authors use DL techniques to perform the aforementioned tasks on data representing the railway structure. We give a wide overview of the papers in which the authors employ DL techniques to classify, segment, or detect components or defects of the railway infrastructure. However, our work is not exhaustive of all the works published by researchers on this subject or of all the features of the railway infrastructure. We select papers concerning the analysis of the faults afflicting the railroad, the pantograph, or the OCS, published in the English language from 2016 to 2021. The datasets in these works are mainly composed of images, but signals or point clouds are also considered. These papers are then classified based on the component analyzed, the kind of dataset, the task accomplished by the authors, and the architecture that inspired the model. Finally, we highlight some advantageous features distinguishing a few works. In other words, we select articles in which only a limited dataset is needed to obtain good results in the experiments. Another "selling point" is whether the proposed model can be utilized for other datasets or to analyze different components. Another highlighting criterion is related to the time cost, considering where techniques to reduce the training time are exploited. Finally, methodologies able to address different defects on several objects are worth mentioning.

Although there are other review works in the literature concerning the analysis of railway infrastructure faults, this paper aims to examine articles in which the authors use DL techniques to perform anomaly detection. For example, in [9] and [53], the authors take into account papers in which a diagnosis of rail faults is obtained with ML techniques and data-driven models in general, respectively. On the contrary, our work is devoted to analyzing deep learning architecture to absolve the previously described tasks. Nonetheless, there are some existing surveys concerning DL models applied to image classification and segmentation, object and edge detection to examine surface defects in different industries, without a particular focus on the railway context [44].

The rest of this paper is organized as follows: in Section II, we describe the criteria used to classify the selected papers. Section III is devoted to the description of the content of the papers. This section is split into four subsections, based on the task absolved by the authors: classification (III-A), segmentation (III-B), detection (III-C), or other tasks not included in the previous categories (III-D). Each subsection is, in turn, split based on the railway infrastructure component taken into account in the studies. In Section IV, we mention the evaluation metrics employed by the authors. Finally, in Section V, we present our conclusions.

TABLE 1. The component of the railway infrastructure analyzed for anomaly detection.

Infrastructure area	Component	References	N. of papers
Rail Area	Rail Surface	[3], [2], [26], [14], [62], [55], [16], [61], [1], [67], [52], [42], [64], [66], [28], [63], [27], [65], [60], [47], [58], [32], [13], [50], [4], [54], [19], [20], [41].	29
	Rail Fasteners	[12], [6], [34], [43], [51], [65], [47], [35], [49], [13], [19], [20].	12
	Obstacles	[46], [24], [57].	3
	Other	[39], [41], [59].	3
No rail Area	Pantograph	[33], [5].	2
	Catenary	[7], [30], [36], [48], [38], [8], [37], [25], [23], [18].	10

II. ARTICLES' CLASSIFICATION CRITERIA

The literature review, carried out from May 2021 to November 2021, considers papers concerning the anomaly detection of the railway infrastructure using deep learning techniques published from 2016 to 2021.

The breadth of the review work suggested identifying different classification criteria, enabling to recognition of different groups of research activities targeting similar objectives. In some cases, it proved sensible to define categories and subcategories.

The first classification criterion concerns the railway infrastructure component taken into account for anomaly detection (Table 1). First, we gather the papers into two categories:

I. *Rail Area*, which includes the articles devoted to the anomaly detection of the rail components, identifying the track obstacles, and small objects on the rail area. In this case, the dataset is often obtained by a camera positioned under the train, and the images are frequently dark;

II. *No Rail Area*, which includes the papers analyzing the defects of the OCS and the pantograph's faults. If it is not acquired at night, the image that must be analyzed could be overexposed. Despite the pantograph being a train component and not an element of the railway infrastructure, in this study, we analyze the faults of its components since the right functioning of the pantograph is essential for supplying energy to the vehicle.

Successively, the papers of the first category are split into four groups based on the kind of component considered:

1) *Rail surface*, including the studies that analyze the weaknesses of the rail surface such as squats, joints, corrugations, cracks, fractures, rust, fatigue block, stripping off block, sunk-inks, and spalling. Moreover, this category also includes papers concerning faults of ballast, such as ballast degradation or loose ballast,

TABLE 2. The type of dataset used to realize the experiments.

Dataset	References	N. of papers
Images	[33], [3], [2], [26], [14], [62], [7], [46], [24], [57], [50], [55], [16], [61], [1], [67], [52], [42], [64], [66], [28], [60], [27], [6], [34], [43], [51], [65], [47], [19], [20], [41], [30], [36], [48], [38], [5], [8], [37], [25], [23], [18], [58], [35], [49], [32].	46
Signals	[13], [4], [63], [54], [39].	5
Other	[12], [59].	2

TABLE 3. The task performed in the study.

Task	References	N. of papers
Classification	[33], [3], [2], [26], [13], [12], [4], [63], [27], [34], [43], [51], [65], [47], [19], [20], [48], [37], [25], [35], [14], [49], [32].	23
Segmentation	[62], [12], [7], [50], [52], [64], [28], [27], [51], [65], [47], [19], [20], [48], [5], [18], [32].	17
Detection	[26], [46], [24], [57], [55], [16], [61], [1], [67], [52], [42], [66], [28], [60], [6], [51], [34], [65], [47], [54], [19], [20], [41], [30], [36], [48], [38], [5], [8], [37], [25], [23], [18], [32].	34
Other	[13], [7], [39], [59], [58], [35].	6

sleepers and railway subgrade defects, the segmentation of rail area, the classification of railway track materials, conductive objects across the rail joints, mechanical defects, and electrical noise;

2) *Rail fasteners*, including the papers considering the faults of rail fasteners, both for ballast or ballastless tracks. Examples included skewed or partially fractured spring bar fasteners and missing element fasteners;
3) *Obstacles*, containing papers detecting obstacles placed on the track, such as people, trees, trains, bags, boxes, helmets, cardboards, or rail-side plastic bags, which could cause smoke or fire;
4) *Others*, containing papers in which the authors analyze several faults not included in the previous categories. More in detail, here we consider papers in which the authors analyze vehicle-body vibrations [39], the railway assets, including switches and signals [41], or vehicle onboard equipment defects [59].

Papers of the ''No rail area'' category are then split into two sub-categories:

1) *Pantograph*, which includes the recognition of the pantograph slide plate faults (mild, excessive, groove, and slide eccentric wear) and the pantograph offset;
2) *Catenary*, including papers analyzing the faults of the catenary system, an essential element in the OCS, and its components.

The second classification criterion is based on the type of dataset employed in the study (Table 2). We recognize three categories:

1) *Images* captured by cameras mounted on an inspection vehicle, UAV vehicle, or frames extracted by videos;
2) *Signals*, such as acoustic emission signals, acceleration signals, current signals, radar signals, or track inspection data [39];
3) *Others*, including 3D point clouds, 3D rails profiles obtained by preprocessing the three-dimensional data acquired from a laser camera, or vectors of binary variables.

The third classification concerns the task performed by the authors (Table 3). We split papers into four categories:

1) *Classification*, includes documents in which the proposed architecture aims to distinguish between different defects or objects. In this category, we have papers that realize classification directly on the available dataset and reports that classify defects or things after the use of detection or segmentation identifies them;
2) *Segmentation*, which contains articles performing segmentation of the rail area, track components, obstacles, or defects;
3) *Detection*, comprising papers where the model identifies a defect or obstacle, providing a bounding box around them, sometimes including a label with the type of defect or object;
4) *Others*, containing papers not included in the categories mentioned above. For example, studies in which the authors expand the dataset using generative models to increase the training set size, papers related to the enhancement of the images of a rail infrastructure component, vehicle-body vibration prediction, or the detection of vehicle onboard equipment defects.

We observe that the majority of the papers performed multiple tasks. This is, in our opinion, likely due to the fact that the works attempt to solve real problems in the field instead of constructing toy ones for ''pure'' research purposes. For example, there are several articles in which the authors first realize segmentation to extrapolate the rail or the component to be analyzed and then realize detection or classification on the images previously segmented. In this way, they reduced noise due to other background components. Therefore, the same references can be found repeatedly in the table depending on the papers' multiple tasks. Category ''Segmentation'', in addition to containing the works in which the authors segmented the defects or objects present in the railway network, also includes papers in which, before carrying out other tasks, the authors proposed innovative algorithms to position or crop the rails in the images.

Since this work is DL-oriented, there was an interest in discovering what were the most common DL architectures or frameworks employed. The bar chart in Figure 3 contains a classification of the papers based on the type of neural network on which the proposed model is based. The architectures described in the articles are based on these nets, which are then improved, optimized, and adapted to the case study. Moreover, we also indicate neural networks used as the backbone of the proposed model.

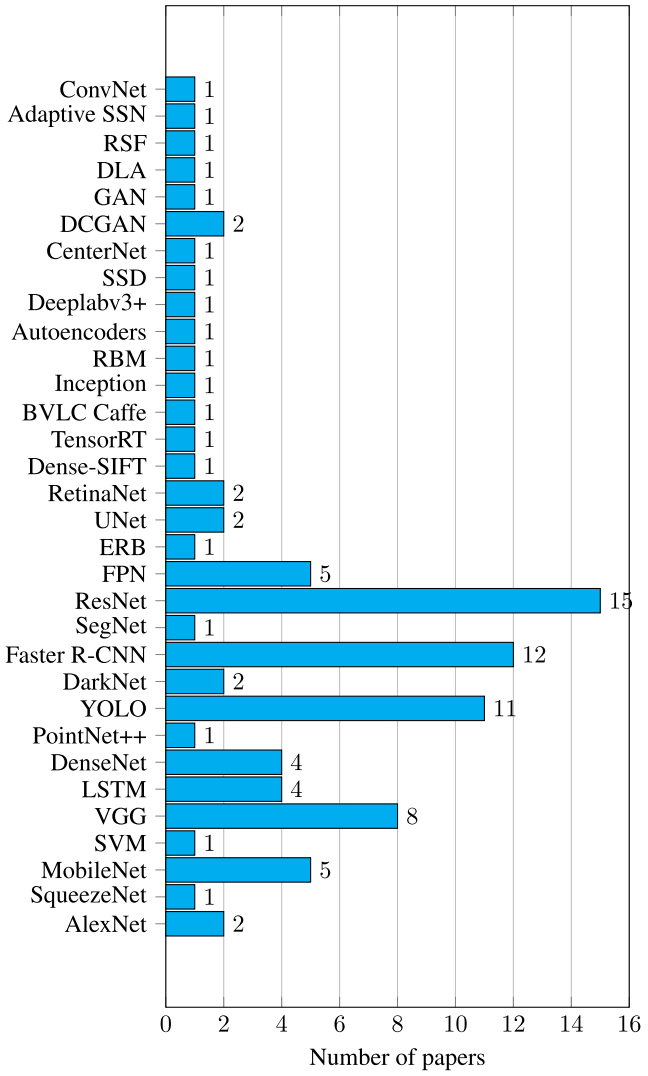


FIGURE 3. The reference architecture and the backbones inspired the model chosen by the authors to accomplish the task.

TABLE 4. Results observed versus the component analyzed for the anomaly detection.

Component	Results Observed: Limited dataset	Adaptability	Computational speed	Reduced training time	Various defects/ object
Rail Surface	[26], [55], [61], [42], [64], [66], [27], [58].	[67], [42], [66].	[55], [16], [52], [42], [66], [28], [60], [65], [47].	[3], [2], [62], [55], [67], [66], [27], [47].	[3], [2], [26], [14], [62], [16], [67], [42], [4], [28], [60], [32].
Rail fasteners	[6], [51].	[49].	[6], [51], [65], [47].	[51], [47], [49].	[12], [6], [51] [65], [47], [20], [35].
Obstacles			[46], [24], [57].	[46], [24], [57].	[46], [24].
Pantograph			[48]		[33].
Catenary			[48], [38], [37], [25].	[30], [23], [18].	[36], [48], [37] [25], [18].
Other	[13], [41].		[50], [54], [41], [59].	[54], [20], [39], [41].	[13], [54], [19] [20], [39], [41], [59].

The distribution of the defective components across the results reached is listed in Table 4. More in detail, we also identify works achieving particular objectives, such as:

1) Trainable with *Limited dataset*, containing references in which all or part of the experiments carried out require a number of samples less than a predefined amount, with the limit empirically set to 550 in this study;
2) *Adaptability*, meaning that the proposed architectures are trained and tested on different datasets to accomplish the same task ([42], [66], [67]). This category also contains papers whose model is extended to a different flaw identification ([67]); on the contrary, papers in which this adaptability is only mentioned and not validated by experimental results are excluded;
3) *Computational speed*, containing papers in which the researchers highlight the computational speed of the proposed model in addition to the accuracy achieved, utilizing suitable evaluation metrics, such as FLOPS, inference time, inspection time, frame per seconds, average time for image recognition;
4) *Reduced Training Time*, containing papers in which the authors used transfer learning, multitask learning, or other techniques to reduce the training cost;
5) *Various defects/objects*, including papers in which defects or detected objects are classified into two or more categories. As a rule of thumb, documents in which the authors use the obvious defect classification as whether normal or faulty are excluded.

III. REVIEW RESULTS

This section describes the articles cited in the previous areas, dividing the analysis according to the task performed by the authors.

A. CLASSIFICATION

The purpose of this section is a survey of papers in which the authors realize the classification of the railway infrastructure components or other classes of objects in the railway area.

In particular, we split this paragraph into four subsections based on the railroad element analyzed in the papers. Starting from the rail area, in Section III-A1, we present papers in which rail surface, ballasts, and sleepers defects are classified, and in Section III-A2, we review articles in which fasteners faults are studied. Concerning the OCS, in Section III-A3 and Section III-A4, we summarize papers in which the pantograph and catenary's weaknesses are classified, respectively.

1) CLASSIFICATION OF RAIL SURFACE'S DEFECTS

Here we present a survey of papers concerning the classification of rail surface, ballast, and sleeper defects.

The wheels' rubbing and impact over time, contact forces, and material aging cause several rail surface defects, such as corrugations, cracks, and squats. These damages could affect rail safety, and faults such as sunkinks could cause train derailment. Therefore, automated and fast detection of rail surface imperfections is fundamental to guarantee safety and allow timely intervention by the rail patrol. The images acquired by the cameras arranged on the inspection train are often affected by uneven illumination and changes in reflection; moreover, each image could contain different defects randomly distributed. Detection and classification of other blemishes are challenging for researchers who have begun to use deep learning techniques to address this topic in the last decades. Reference [3] aims to realize defect detection with high accuracy, combining the features extracted from two architectures: SqueezeNet and MobileNetV2. The authors chose these two neural networks since they are smaller and faster than other networks. The presented model consists of three steps: first, images captured by cameras mounted below a locomotive are preprocessed to reduce noise. In particular, filtering and increasing contrast are applied to the images; in the second step, the rail is found and cropped by the rail position algorithm. Finally, the classification of rail surface defects is made by combining the two nets and providing the training set to each one of the deep neural networks, in such a way as to extract the features by means of 16 residual blocks, a convolution block, a ReLU, and a global average pooling block. Then, the relief algorithm is involved to get the weights of these features. Successively, several features with high weight are selected, concatenated, and supplied to the support vector machine (SVM) to classify the objects into four classes: healthy, light squat, joint, and severe squat. Despite this model being less accurate than other methods, it does not require a lot of parameters to set. It allows the classification of multiple rail surface defects for images affected by irregular illumination and differences in reflection. In [2], the authors propose a two-step process to classify railroad imperfections, caused by abrasion due to the rubbing of the train wheels in the time, employing two different datasets. In the first step, train tracks are cropped from images captured in a large area, using Otsu's method. Moreover, data augmentation is applied to the NEU dataset and pre-processing to the author's dataset to clear and resize the images. In the second step, the classification task is realized by means of a hybrid system assembled with VGG-16 and transfer learning. In the proposed fine-tuning method, every one of the layers is not frozen and the size of the entrance image is rearranged before the layers of an input image are put into frozen transfer learning training, and the system is entered into the intermediate feed layer. This model allows us to save time since classification is not instantly weighted as in transfer learning, without reducing the success rate. Results are compared between CNN, Transfer learning VGG-16, and the proposed model success rate.

To detect fractures, squats, rust, and corrugations, in [26], the authors propose a multi-robot system. A Raspberry Pi camera supports the ultrasonic sensors to capture images, and a CNN model is analogized to ML algorithms. Moreover, this model is expanded to a multi-robot environment using the Low Energy Adaptive Clustering Hierarchy (LEACH) protocol, which helps to minimize energy consumption. These multiple frameworks use an IoT-based cloud server to communicate and coordinate the robots. Concerning fracture and crack detection, first, images are preprocessed to remove noise pixels. Successively, the Speeded-up robust features model is employed to extract features, catching feature points of specific objects in an image that may differ for rotation, scale, or pixel intensities. Ultimately, the number of key points is computed and a multi-directional fracture is then detected. Finally, a CNN realizes the classification. The same procedure is applied for the squats. Concerning rust detection, the acquired images are converted into the HSV color space. Successively, images are classified as rust if white pixels are greater than a threshold value. Finally, the processed images are updated on the IoT server. The proposed method outperforms Artificial Neural Network (ANN), CNN, random forests, and SVM, trained and validated on the same dataset, realizing a real-time search of defect detection of the railroad.

The model proposed in [27], named TrackNet, focuses on one class of track fault, i.e. rail discontinuity. Here, images are captured by a visual-based track inspection system (VTIS) for rail surface, composed of a high-speed camera mounted under an inspection train. First, a U-Net is applied to crop rail tracks from the images and locate the Region of Interest (ROI) to realize the semantic segmentation. After that, several picture processing tools allow cropping of the part of the probable faulty region in the images. The images obtained in such a way are saved and then fed to a neural network that categorizes the images into either True or False alarms. Concerning the classification step, the weights of the DenseNet and ResNet are initialized from an architecture trained on the ImageNet dataset, and the network model is trained using stochastic gradient descent. The last fully connected layer gives the two classes corresponding to the true or false alarm. Data augmentation and K-fold cross-validation are also used. The ROI segmentation of the first step allows the network to focus on a small portion of the

track, reaching better performance. Nevertheless, this work analyses only one class of defects of rails.

In [47], the authors present a real-time fault detection method for the track elements. The suggested model consists of two stages. First, an improved lightweight instance segmentation network is offered for the segmentation and location of fasteners and rails. Second, an approach based on geometric features is proposed to detect fastener defects. This method alleviates the scarcity of defective fastener images and overcomes the problem of the subtle differences between classes. The architecture consists of four modules: instance segmentation of track components, fastener defects detection, rail defects detection, and TensorRT acceleration. A modified YOLOACT network is exploited to extract the track elements. Concerning the weaknesses of fasteners, pixel lengths are evaluated to establish if the fastener is faulty or not. For the rail defects, an instance segmentation network is applied. MobileNetV2 is used to classify the defect type of rail surface if there is a defect. Finally, TensorRT is utilized to accelerate the instance segmentation. The proposed method ensures high detection accuracy and fast inference speed. However, its robustness could be improved by analyzing rare defect classes.

In [14], a CNN is presented to classify the images of the rail surface captured by a camera mounted on a measurement vehicle into six classes: regular, weld, light squat, moderate squat, severe squat, and joint. Here, three deep convolutional neural network (DCNN) performances are compared. These DCNNs, named small, medium, and large, differ for various combinations of parameters, such as the dimensions of filters, the number of feature maps, the number of layers, and the number of nodes of fully connected layers. To solve the problem of class unbalancing, the authors apply undersampling to the standard type. The proposed architecture guarantees feature extraction effortlessly, which is essential due to the huge quantity of images. These data are labeled manually by the authors: to improve the automatic detection of defects, the employing of autoencoders could be considered in future works. In [32], the authors utilize a two-stage deep learning method. The first step consists of rail detection and the second step includes the detection and localization of five rail surface faults: abrasions, scars, cracks, corrugations, and normal. Concerning the first step, the detector uses an anchor-free module to find rails in the raw images acquired by a line-scan camera at the bottom of the rail inspection vehicle. It comprises a backbone network, ResNet, and two subnetworks: the first one is a small Fully Connected Neural Network (FCN) that realizes a binary classification into rail and background part on the backbone's outcome; the second one indicates the width offset of the rail bounding box. In the second step, the authors apply ConvNet with a sliding window characterized by a specified size and step in the image containing only a rail. The authors chose this architecture since the proportions of rails are substantial, and the defects on the surface have various shapes: abrasions are not vast but very long; on the contrary, scars occupy a few tens of pixels. ResNet18 is used to classify five types of defects. Finally, to localize the faults, the authors exploit the class expected in each window and the location of each window. In particular, they merge the neighboring imperfections of the same class in the longitudinal orientation and attach the cracks in the horizontal orientation. Results obtained using other backbone networks are compared. Despite the different shapes of the defects analyzed, the proposed approach guarantees a robust performance. However, the model could fail to detect minor faults located at the boundary of the sliding window.

In the papers mentioned above, the datasets analyzed are composed of images. In other studies, samples consisting of signals different from images (electric signals, acoustic signals) are employed to classify rail surface defects. In particular, in [63], the authors propose a method founded on CNN and probability to detect rail defects, analyzing multiple acoustic emissions (AE) obtained by a test system composed of a test sample, a tensile testing instrument, and an AE data acquisition device. AE techniques are more precise and accurate, guaranteeing dynamic detection in real time. A CNN based on a single event is not able to consider the relationships among numerous AE occurrences. On the contrary, several successive dangerous AE signals can be obtained when a fault occurs in the rail. By exploiting the probability of various events, the authors improve the classification accuracy and annihilate the error rate of a single event. The first stage of the architecture proposed in this paper consists of a classification task realized by a CNN: first, the AE signals are preprocessed by the Fast Fourier Transform, and the resulting data is the input of a CNN. In the second stage, the probability of each sample belonging to a category is obtained. In this work, the strong point is represented by the integration between a CNN and an AE system, the limit, as in many cases, is the specific application contest, ie that of monitoring the state of the rails. Reference [4] deals with recognizing rail imperfections such as cracks, joints, cracked sleepers, defective fasteners, hanging sleepers, and local irregularities (mud spot) of the track, exploiting the wavelength variations that occur in the presence of short-wave defects. Also in this article, as in the previous one, the objective is to monitor the state of the rails, the limitation is represented by the fact that the tests have always been conducted in the same configuration, meaning speed, direction of the vehicle, starting point for measurement and so on. An oval track-vehicle scale model, including cracks, rail joints, mud spots, and the entrance/exit to a bridge, is used to simulate a railway. When the train passes a fault, the sensor perceives a growth of vertical acceleration. The produced signal is analyzed to identify and classify eight different classes of defects: braking of the vehicle, acceleration of the vehicle, no-fault (train stopping), no-fault (train moving), rail crack, rail joint, entrance/exit to the bridge, and mud spots. A recurrent Neural Network (RNN) is involved to increase the cardinality of the dataset. The proposed architecture contains two LSTM layers, a dropout, and a final classification layer.

This model is finally compared with the Bagged Decision Tree.

Often, deep learning algorithms show higher accuracy than traditional machine vision algorithms, but they cannot realize real-time detection and defect localization. To solve these problems, in [16], the authors propose a novel object detection algorithm that uses as backbone network MobileNet (MobileNetV2 and MobileNetV3) and additional detection layers with multi-scale feature maps inspired by You Only Look Once (YOLO) and Feature Pyramid Network (FPN). The images are acquired by the line scan cameras placed on the inspection train. This model can detect and classify three kinds of defects, fatigue block, corrugation, and stripping off the block, which are respectively identified by a green, red, and blue bounding box. This work aims to monitor the state of the rails for increased safety issues, but the task it faces is to return the information retrieved in real-time. This aspect is obvious for vision systems but not for those of DL. Obviously, as can be imagined, the limit of this system is that the data must necessarily be collected accurately and labeled.

We end this subsection by describing two papers in which the authors classify ballast and sleeper defects, such as crumbling and chipping on a tie. In [19], the authors propose an FCN composed of four convolutional layers to classify ten classes of materials, including ballast, rough, wood, concrete, smooth concrete, medium concrete, chipped concrete, crumbling concrete, rail, lubricator, and fastener. The ground truth data is annotated using a custom annotation tool that allows assigning a material category to each tie and its bounding box. The tool also allows for defining polygons enclosing crumbling, chips, or ballast regions. To indicate whether an image includes a damaged tie, the scores for each class, chip, or crumbling, and then for the whole picture are computed. The detectors report an alarm if the score calculated exceeds the detection threshold. Finally, by considering the label corresponding to the highest score, the segmentation map is obtained. The accuracy of the proposed model is computed by dividing each tie into four pictures and evaluating the score on each image independently. The use of DCNN ensures a better performance than shallow machine learning since these architectures can acquire more elaborate patterns, reducing overfitting despite a limited training dataset. In [20], the authors, in addition to detecting defects of sleepers using the same architecture in [19], classify rail fasteners (Section III-A2).

2) CLASSIFICATION OF RAIL FASTENERS' DEFECTS

In this subsection, we proceed with a survey of the papers in which deep learning architectures are employed to classify fasteners' defects (Figure 4).

In [19], the authors use the result of the fastener detection model to classify fastener defects (Section III-A1). Reference [20] deals with the detection of defects on sleepers and rail fasteners. The focus of this work is the inspection of the fasteners of the sleepers to the rails. By using a network that simultaneously performs multiple tasks, it has been possible to benefit from the scalability of convolutional neural networks. In fact, despite having a limited dataset, a common problem when using these algorithms, we have taken advantage of the use of intermediate characteristics. In particular, they split the problem into two parts: fault detection (good, broken, and missing fastener) and semantic segmentation (crumbling and chipping concrete ties and different material classes). The dataset images are collected by a moving vehicle, and the proposed architecture is an FCN based on [19]. There are four convolutional layers for the material classification task into ten classes. In addition, the fastener detection task includes five convolutional layers which accomplish two tasks: coarse level and fine-grained classification. The fasteners images in the training set are divided into five rough classes: the first one contains missing and broken fasteners, the following three categories include PR-clips, e-clips, and fast clips, and the final class includes everything else. The classifier divides fastener vs background at the coarsest level and includes missing fasteners. Once the fasteners' location is detected, the fasteners are classified into good and broken and then organized into the fastener type: PR clip, e-clip, fast clip, c-clip, and j-clip. The second task is based on the SVM maximum margin principle. Combining the outcome of the binary classifiers, a single score is generated. This value indicates how correctly the system is able to determine if an image contains a fastener without defects.

FIGURE 4. Particular of the fastening system, focused on the tie plate.

In [34], rail track images, including supporting frames, video recorder, and lightings, are collected, and the YOLOv3 model has been used to classify eleven types of fasteners. After describing the existing literature concerning rail inspection, the authors propose using Artificial Intelligence (AI) in conjunction with the back-end server to reduce the burden of rail workers. The pictures captured by a video camera, that allows night vision, are transmitted to the back-end DL server, which accepts the front-end image data

and identifies and archives them. Moreover, this device is also able to record the speed, time, and GPS information. In order to reduce the positioning error provided by GPS, the authors suggest using a hundred-meter virtual detection circle (VDC) and they choose YOLOv3 since it is lightweight and efficient. The images used are captured by railway workers and labeled as normal/defective under the supervision of these experts. In this article a DL-based automatic system for the inspection of rail fastening systems is proposed, the limitation in this case is represented by the need to manually label the data.

The information system used to monitor fastener defects in [43], is mounted on the detector vehicle and supplies fastener inspection in real-time. Four video cameras, which continually take photos of the tracks, are placed on the base of the car. Successively, a workstation processes the video stream. In the obtained pictures, parts with fasteners are recognized, formed into data packets, and fed to a variant of the VGG network, composed of six convolutional layers and five max-pooling layers. The results for the classification are sent to the operator's decision support system (DSS), and the information about the track condition is sent to the Railway Diagnostic Center. Six categories of rail fastener defects are classified, and data augmentation is utilized to expand the size of the training set, in such a way as to alleviate the scarcity of defective data.

In [51], two essential steps are necessary for the detection of fastener faults. The first one consists of precisely setting the fastener's location in the pictures captured by a camera. The second step consists of classifying complete, broken, and missing fasteners. Concerning the first one, the location of the fastener is achieved by searching the track's edges and the concrete rail bed. The method used allows segmenting the fastener from the other components by exploiting the spatial relation of the fastener, the track, and the backing plate. Concerning the second step, the authors present an algorithm that guarantees the extraction of local features, the Dense Scale Invariant Feature Transform (Dense-SIFT). Then, Dense-SIFT features extracted from each picture are mapped to the Bag-of-Visual-Word (BOVW) and used to classify fasteners, by constructing their bag of visual words. Moreover, spatial pyramid decomposition is introduced to overcome the image's lack of spatial location information. In the second part of the paper, to facilitate the fastener classification method, a VGG16 network is used. Finally, faster R-CNN is employed to increase the detection rate and efficiency for fastener defect detection. This work gives an exhaustive comparison of several methods to classify fasteners' weaknesses, reaching a high classification accuracy in a reduced time. Reference [49] aims to classify railway fasteners to reduce the high false-positive rate due to ballasts covering the fasteners or non-uniform illumination of the images. The pictures captured by a camera mounted under the train are first resized and normalized, and then fasteners are extracted and annotated. The dataset has been acquired from two passengers and a freight railway line. The authors use two deep learning architectures, pre-trained on ImageNet, AlexNet, and ResNet, to classify fasteners as normal, normal with ballast occlusion, normal with non-uniform illumination, and abnormal. More in detail, the authors create four models, two for each architecture: in the first, they freeze the convolutional layer and realize the finetuning of the last two fully connected layers and the classification layer; in the second, they unfreeze the convolution layer and tune all the parameters. These models are trained, evaluated, and tested on the images captured on the first railway line; successively, the trained model is applied to the pictures of the second line to test the generalization ability. Undersampling of the standard fasteners is used to balance the two classes, normal and abnormal. Finally, the model is compared with the other four architectures. ResNet with unfreezing layers reaches the best classification on each category and a good generalization result on the new test set. Since the proposed model is tested on the second line, this work guarantees applicability to other datasets.

Reference [65] aims to detect rail and fastener defects simultaneously. Firstly, the authors utilize a modified YOLOv5 architecture to localize fasteners and rail in the images collected from a camera mounted on a particular inspection train. The method proposed here uses different networks simultaneously to perform different tasks. This is a fast and non-destructive method but its accuracy needs to be improved.

This model involves a Ghost bottleneck [22] that optimizes the original cross-stage partial (CSP) backbone network of the YOLOv5, reduces the computational cost, and increases the inference speed of the network. The Mask R-CNN is a network that realizes object detection in two-stage, localizing and segmenting the area of the defects. Finally, an architecture based on ResNet is used to classify the fasteners into regular, loosening, and broken.

Reference [12] develops a real-time inspection system to classify three kinds of fastener defects, skewed spring bar, missing components, and partial fracture spring bar. A structured-light-based fastener examination procedure, called Intelligent Rail Checker (IRC), is employed to obtain fasteners' dense and precise point clouds. Data are sent to the deep learning module, where the point cloud of fasteners is segmented into different parts and transmitted to the data storage module. More in detail, a point cloud extraction algorithm, which consists of a sliding-window-based method, is suggested to extract the point cloud of the fasteners from the initial point cloud. Then, an automatic annotation technique based on region growth is suggested to construct a dataset composed of a ballast rail fastener point cloud. The authors expand the dataset by data augmentation. The performance of three deep learning point cloud segmentation networks, POINTNET++, POINTSIFT, and POINTCNN are compared on the semantic segmentation dataset. POINTNET++, which achieves the best performance, has been embedded in the Deep learning module of IRC to find defective fasteners. This point cloud-based inspection model looks very promising because it can be completely automatic but

unfortunately, we do not yet have well-balanced datasets with the right presence of broken and intact components.

The class of defective and defect-free fasteners is often imbalanced, and there are infrequent real defective fasteners on the railroad. This dataset feature can worsen the performance of the architectures employed to classify faulty fasteners. To unravel this problem, in [35], the researchers apply a deep learning approach named four discriminators cycle-consistent adversarial network (FD-Cycle-GAN) to augment the number of faulty fasteners. The proposed approach that improves the existing GAN is composed of two generators and four discriminators. The two generators produce fake defect fastener pictures and fake defect-free images. Concerning the four discriminators, DX1 and DX2 differentiate authentic defect-free fastener images from the generated defect-free pictures, while DY1 and DY2 aspire to distinguish real defect fastener pictures from the generated image. After generating defects, the fasteners are classified into defect-free, missing, and damaged, training the VGG-16 network with a dataset composed of generated defect fasteners and authentic images. This article addresses a widespread problem, the lack of balanced datasets by generating images of defective objects in a synthetic way, but the problem that persists is that the system is trained with synthetic data and not with real data.

To ensure safety during the inspection of railway tracks, it is critical to establish if trains occupy sections of a railway track. Therefore, a detection system is designed to report the railway section as occupied in case of fault. Reference [13] proposes an artificial recurrent neural network (RNN) called the long-short-term memory (LSTM) for imperfection diagnosis in railroad track circuits. A track circuit utilizes the rails in a track area as conductors that link a transmitter at one end of the part to a receiver at the other end. When a train enters the area, the short circuit caused by the wheels generates the current flow via the receiver. Where the relay is not energized, the current decreases to a level and the section is indicated as occupied. The faults evaluated in this paper are conductive objects, insulated joint faults, mechanical rail defects, electrical nuisance, and ballast degradation. Since there is data available to train the network, a generative model is proposed in the paper. Faults are diagnosed from temporal and spatial reliances by considering the signals coming from multiple track circuits in a geographic region. To evaluate the spatial reliances, the input is given by the electrical current signals coming from five different track circuits. An LSTM is used to inspect temporal dependencies since it can discover long-term time reliances by presenting several memory cells into the model.

3) CLASSIFICATION OF PANTOGRAPH'S DEFECTS

In the previous subsections, we analyzed defects of the rail area. On the other hand, this subsection provides an exposition of the paper concerning the classification of the flaws of the pantograph, a critical component of the railway infrastructure supplying power to the locomotive.

In [33], the authors analyze the damages that afflict the pantograph sliding plate due to strong electromechanical effects. In the paper, the images of the sliding plate captured by several cameras mounted in the tunnel are first normalized in size and color and then preprocessed using the deep learning framework CAFFE. Data are split into five classes, including regular, mild, excessive, groove, and slide eccentric wear. To classify the sliding plate defects, the authors first exploit the deep learning architecture AlexNet, since it ensures an increased accuracy rate in the extraction of the features. Then, to improve the accuracy obtained, the authors optimize AlexNet, setting the structure parameters and hyper-parameters. Finally, a CNN architecture, PanNet, is proposed based on AlexNet and deeper in the hierarchy to reach the best accuracy. The proposed architectures can detect different classes of defects with good performances. However, parameters are set based on many experiments; therefore, this model could be improved by finding a theoretical way to individuate parameters with limited training data.

4) CLASSIFICATION OF OCS DEFECTS

Here we review papers concerning the classification of OCS defects. Firstly, we consider the catenary a critical component of the OCS. Catenary support instruments are important to ensure energy for high-speed trains, and their features must be worked. In particular, the split pin is a protective element of the catenary and dramatically influences the functioning of the catenary during the train ride. In [48], the authors classify the faults of this component, analyzing catenary pictures acquired by HD cameras on a detection vehicle photographed at night. Since split pins are small components compared with the whole picture, a two-stage localization approach is utilized: thimble-up (T_UP), mast bracket (MB), thimble-down (T_DOWN), and clevis are first located, and then, split pins are identified in the pictures. The improved YOLOV3 algorithm localizes the joint components in the first localization, such as MB, T_UP, T_DOWN, and clevis. Since the split pin is relatively tiny in the image and the picture is unclear after localization, a deblurring module is integrated into the model to increase the accuracy of following semantic segmentation and classification. Successively, split pins are extracted carefully. The deeplabv3+ algorithm is used in this localization phase to realize semantic segmentation. Concerning the classification stage, the split pins are separated into three based on their different shapes and positions in the catenary. Each split pin type is categorized into three conditions: loosening, missing, and normal. The authors carry out several experiments in order to verify the model's adaptability under other environmental conditions and the proposed method reaches an elevated accuracy, although split pins are very small components of the OCS.

Typically, the regular functioning of the OCS relies on the right sliding connection between the pantograph and the catenary, and the dropper ensures adequate sliding contact. In [37] the authors propose a dropper defect detection method established on depthwise divisible convolution, analyzing

the images of the OCS captured in different and complex backgrounds by several cameras mounted on the top of the train. The technique consists of two steps: a dropper progressive location network (DPLN) and a dropper fault recognition network (DFRN). DPLN used to enhance the pantograph consists of a pantograph location network (PLN) and a dropper location network (DLN). It utilizes MobileNet to create an OCS picture feature extraction block. This architecture can identify numerous droppers in the OCS images. The dropper closer to the photo camera is chosen as input for the procedure based on depthwise separable convolution. DFRN can identify the dropper as normal, broken, slack, and missing. The problem is that the limited amount of data available does not allow for superior performance compared to other methods [25] deals with the detection of fault droppers and current-carrying rings. The data set comprises images extracted from videos obtained from the inspection vehicle. The deep learning method proposed consists of three steps: first the detection of current-carrying rings, second the match of rings, and third the classification of droppers. First, a modified CenterNet identifies the defective and the standard current-carrying rings. Here, the feature pyramid network (FPN) block is implanted in the DLA network to adapt to the miscellaneous scales of rings. In the second step, the regular current-carrying rings are chosen and divided into top and bottom rings. Therefore, the bottom and the top rings are checked via the relative position, and the dropper region judges the region between the matched rings. The third step consists in classifying with ResNet-34, ResNet-50, ResNet-101, and a DenseNet-169, the dropper as normal or faulty (absent, loose, or broken). The most promising results in classification accuracy have been achieved by ResNet34 and ResNet-50.

B. SEGMENTATION

Here we describe papers whose aim is to segment defects of the railway infrastructure components or the components themselves. We also include in this subsection papers in which the authors propose an innovative rail positioning algorithm before carrying out other tasks. According to Section III-A, we split this section into four subsections based on the element of the railroad segmented in the research. In particular, Section III-B1 and Section III-B2 are devoted to the segmentation of rail surface imperfections and fasteners, respectively. Concerning the ''no rail area'', Section III-B3 and Section III-B4 deal with the segmentation of pantograph and catenary components, respectively.

1) SEGMENTATION OF RAIL SURFACE'S DEFECTS

To realize an efficient rail inspection is fundamental to detect rail surface defects accurately. These faults take up a small section of the images captured by the inspection vehicle, and therefore their segmentation is necessary to accomplish other tasks. Reference [62] aims to identify the no-service rail with three categories of rail surface defects: spot-shaped, strip-shaped, and block-shaped. To this end, the authors propose a method with image-level labels that highlight defect characteristics such as area and shape. The segmentation method consists of two steps: First, a classification network is equipped with a pooling convolution module, which exploits various pooling functions based on different defect categories and generates the sub-category maps. Then, the pseudo-pixel-level tags are acquired by the sub-category activation maps and the preliminary size information (area and shape). Finally, in the second step, a fully supervised segmentation network, MCnet, is trained using the pseudo-pixels-level labels generated in the first step. Most of the pictures in the data set are artificial, and few are natural images. In this work, a pixel segmentation system for the defects of the railway surface without service is the segmentation system. The limitations of the application are represented by the use of many synthetic images.

In [52], the authors propose a deep learning architecture to segment and identify rail surface defects (RSDs) on the images caught by a camera installed on an unmanned aerial vehicle (UAV). The proposed method addresses the following issues: position, size, and angle of the same rail would not be identical for all the pictures due to the UAV being affected by airflow; the rail surface takes up a large area, and the edge of the rail surface is analogous to the web; finally, the reflection of rail surface fluctuate a lot. The authors propose a hybrid system that integrates FCN and image processing (IP) algorithms to realize rail surface segmentation and RSD inspection. Firstly, the RBGNet exploits the relationship between rail edge and rail object information for rail surface segmentation. Successively, defects are detected by an IP-based architecture combining local Weber-like contrast (LWLC) and the Maximum Entropy algorithm. The proposed architecture reaches a high detection rate, and it guarantees adaptability to different environments. Moreover, the accuracy has been improved, integrating information acquired by the rail track object and rail edge.

Reference [64] provides a deep extractor (DE) that integrates FCNs and conditional random field (CRF). More in detail, first the DE employs CNNs to acquire a high-level feature map and successively utilizes conditional random fields (CRFs) inference to refine coarse and weak pixel-level label predictions and realize fine-grained segmentation. A bilateral FCN (biFCN) with two constituents is suggested in the first stage. The two components are an inverse FCN (iFCN) and a regular FCN (rFCN). rFCN employs a standard encoder-decoder approach to extract high-level features from the whole picture. In contrast, iFCN utilizes a decoder-encoder process to increase missed defect information. The output of two feature maps is joined to obtain a high-level feature map. The architecture is tested on publicly available datasets: the Rail Surface Defect Dataset (RSDD) consists of two data sets captured on express railways and common/heavy haul rails. This approach bypasses offline post-processing and realizes end-to-end deep learning. This guarantees better performance than models exploiting low-level features.

Reference [28] proposes a rail inspection system (RIS) based on a deep multimodel architecture that combines a segmentation stage and a Faster Recurrent Convolutional Neural Network (Faster RCNN) for objective location. This deep multimodel RIS (DM-RIS) realizes an end-to-end parallel model with high detection speed, robustness, and precise segmentation. Finally, the investigation is executed on extensive and various rail samples, divided into six categories, depending on the physical structure of the rail. The proposed parallel architecture DM-RIS consists of an upward path, FRGMM, and a downward path, a Faster RCNN. The first path segment defects using the Gaussian Mixture Model, a method founded on pixel sample statistics, and a Mixture Model Based on MRF, which incorporates local information. On the other hand, Faster RCNN indicates the position of the fault in a bounding box. Integrating spatial information, the model gives the defects edge with high accuracy. In addition, it is robust under different environmental conditions when the images are affected by an imbalance reflection, inadequate illumination, exterior noise, or rust.

An efficient railroad faults detection method is presented in [50]. In this work, a rail area detection system is proposed. The model consists of two phases. The first step comprises the extraction of the rail area using a convolutional neural network (CNN) that can realize a pixel-level classification of the rail picture. The dataset consists of multichannel pictures obtained by the onboard camera at different times. The network's output is a binary image containing the rail and non-rail areas. The suggested architecture consists of an encoder-decoder model inspired by SegNet. The encoding part includes the convolutional layers utilized to obtain the rail area features in the feature map. The cascade sampling layers are employed to down-sample the picture to lower the resolution of the feature map and extract more additional global features. The authors propose a dilated cascade connection to solve the problem of the loss of information during the upsampling process, achieving a multi-scale feature extraction and reasonable accuracy. Therefore, decoding includes deconvolution layers to upsample the encoder's feature maps to reach the input resolution. As the edge of the rail area is not precise, the second step is devoted to optimizing the boundary of the railroad by the improved polygon fitting method. The limitations of the system are the speed of the algorithm and the not-so-good performance in far areas.

Concerning the above-mentioned papers devoted to the classification of rail surface defects, we remember that in [19] (Section III-A1) and [20] (Section III-A2), the authors realize segmentation of sleepers defects in addition to accomplish the classification task. Moreover, [65] (Section III-A2) and [47] (Section III-A1) perform the segmentation of the area of rail defects; [27] (Section III-A1) carry out the segmentation of the rail track. Finally, in [32] (Section III-A1), an anchor-free detector is proposed to crop the tracks in the original images, and then rail surface defects are located and classified.

2) SEGMENTATION OF RAIL FASTENERS' DEFECTS

In [12] (Section III-A2), the fasteners point cloud is segmented before the classification and in [51] (Section III-A2) and [47] (Section III-A1) fasteners are also segmented from the other parts of the images.

3) SEGMENTATION OF PANTOGRAPH'S DEFECTS

The optimal contact between the pantograph and the power grid is essential for the proper operation of rail transport. Among all the imperfections, the pantograph offset strongly reflects the state. Reference [5] aims to reconstruct the shape of the pantograph, replaced with a line between the left and right horns, in a 3D world coordinate system. The first step is locating the pantograph horn region: a DL method named Single Shot MultiBox Detector (SSD) is proposed. The SSD model is based on a forward-propagation CNN, producing several predefined rectangular boxes, possibly including object examples in each rectangular box. The last predicted value will be acquired by non-maximum suppression. The second step consists of corner-point extraction: first, the detected region is half-cropped, and image processing is applied. The images are then divided into two classes based on their background pixel value, and the edge of the horn is extracted. Then, Harris corner detection will locate possible corner points in the edge map. Among them, the correct point is that with the largest ordinate. After acquiring the coordinates of the 2D corner points in the pictures taken by the left and the right cameras, the binocular stereo vision method is evaluated to reconstruct 3D coordinates. Therefore, the authors used a 3D coordinate line of the left and right horns to illustrate the pantograph and the angle between two spatial lines to calculate the pantograph offset: if the pantograph offset is beyond the threshold set before, the program will report the anomaly.

4) SEGMENTATION OF OCS DEFECTS

Concerning the inspection of the OCS, video and images captured from the inspection vehicle are often affected by low visibility due to fog, haze, and other natural factors. Reference [7] proposes an enhancement technique for the catenary images with scarce visibility. First, the authors use a network to estimate the haze transmission map in the atmospheric scattering model. The architecture comprises two branches: multi-scale feature extraction and multi-feature fusion. The first one is the full convolution operation and the dilated convolution, which can apprehend multi-scale context via various dilation rates. The branch of multi-feature fusion combines multi-scale features to evaluate the transmission map. Finally, three concatenated layers connect the initial haze image with additional feature maps and compensate for the loss of information. The second step consists of obtaining the restored image. To this end, the improvement of the transmission map is achieved, and the atmospheric light value is estimated after multiple segmentation of the sky and

non-sky regions. Therefore, the restored catenary image has an improvement in clarity.

In the OCS, the cantilever is employed to preserve the contact wire height and stagger. The cantilever comprises the catenary suspension system and the contact wire support, connected through the swivel clevis (SC). Reference [18] offers an imperfections detection process that integrates an adaptive SC segmentation network (Adaptive SSN) and local operators, intending to inspect two kinds of flaws, including the split pin looseness and the SC crack. This approach alleviates the problem of scarcity of defective samples and shifts in data distribution. This strategy includes three main phases: critical components localization, adaptive SC components segmentation, and fault detection. Concerning the first step, the Faster R-CNN is employed to localize the OCS parts, including the SC. The second step involves a two-stage segmentation pipeline (adaptive SSN). Firstly, a fully convolutional regional proposal network (RPN) accepts the picture as input and generates region proposals that may include the target objects. Successively, the lightweight heads are employed to indicate the bounding box coordinates, classification scores, and segmentation masks for the region proposals. The network heads and the RPN have ResNet50 as the same backbone network. The authors suppose that the segmentation outcome is unreliable if the geometric features obtained by the segmentation masks are inconsistent with the previous geometric features. Successively, the human annotator is invited to annotate the unreliable SCs, add them to the training set, and finally, the architecture is retrained employing the new training set to adjust to the data distribution. Once obtained, the segmentation masks and geometric features can be exploited to detect SC defects. This model could be applied in the future to other components of the OCS.

Finally, in [48] (Section III-A4), the semantic segmentation of split pins is realized before classifying missing, loosening, and standard split pins.

C. DETECTION

This section provides an exposition of papers in which the proposed models aim to realize the detection of defects or objects on the railroad. More in detail, here we list the architectures that identify the defects or the objects and enclose them into a bounding box. Often, this task is followed by a classification step, in which defects or objects are ranked into several categories. Also, this part of our paper is divided into subsections according to the component of the rail area (Section III-C1 and III-C2) or the OCS (Section III-C4 and III-C5) detected in the reviewed article.

1) DETECTION OF RAIL SURFACE'S DEFECTS

Automated rail surface defect detection ensures passenger safety and low maintenance operations costs. The model proposed in [55] is based on YOLOv3 architecture. More in detail, it consists of two steps. The first one concerns image acquisition, which happens through a camera equipped with a device to improve the light source and annihilate the interference of other light origins. The second step is image processing: once the images have been acquired, they are sent to a server in which the YOLOv3 algorithm realizes the defect detection. This architecture comprises a feature extraction layer (Darknet-53) and a processing output layer, similar to an FPN network. In this algorithm, the images are divided into multiple cells, and the bounding box is obtained starting from the coordinates of the center point using a clustering method. Finally, a binary cross-entropy loss is utilized to indicate the category of the object contained in the bounding box. This architecture can detect minor rail surface weaknesses with high accuracy in a reduced time. Nevertheless, some issues linked to the learning rate could be solved in future works.

The best mean average precision has been reached using MobileNetV3 as the backbone.

Reference [61] is devoted to a novel rail surface defect (RDS) for restricted samples with a line-level tag. More in detail, since defective image contains thousands of pixels, they can be regarded as sequences of pixels, and then it is possible to classify pixel lines. In such a way, few images are sufficient to collect and label lines of pixels and simplify the labeling work. Therefore, binary classification of pixel lines is applied to two sub-datasets: one captured from express rails and the second from common/heavy rails. In particular, the authors utilize the OC-IAN method to realize the RDS on the express rail. It consists of five modules: a one-dimensional CNN as a feature extractor, a long and short-term memory (LSTM) network as a context information extractor, an attention mechanism applied to process information, linear layers used to classify, and a filter module that purifies noise. On the other hand, OC-TD is used to detect heavy or common rail imperfections, and it consists of four modules, the same as OC-IAN, without the attention module. In such a way, the architecture can classify defect-line and defect-free lines. The authors also compare the proposed method with other models, showing that their architecture gives better outcomes. However, this approach presents some deficiencies, such as the false recognition of defects, missing detection of minor faults, and not recognizing some portions of the flocculent defects.

In [1], the authors propose an automatic method of identifying defects on the tracks to complement the manual methods to increase safety. The proposed model consists of two steps: first, the images are processed, resized, and denoised. Successively, segmentation is applied; it is followed by morphology, which collects image structures. Then the images will be filtered with the necessary features, and cracks will be segmented. The photos are classified into defective and non-defective. The limitation of this system is represented by the small dataset used for training.

Reference [67] identifies five classes of rail surface defects, such as shelling, corrugations, squats, faults, and grinding marks. The limitation of this method is linked to the dataset. In fact, every image shows only a specific defect.

This task is realized by analyzing railway images captured by a camera installed on the test train. Since each image contains only one flaw, the identification of the rail surface multi-flaw consists of a multi-class classification scheme. The suggested framework comprises a rail extractor and a cascading rail surface defect identifier. The rail extractor aims to extract the rails from the image, reducing the background noise. The extraction step is obtained by two methods, a generic global threshold and Otsu's method. The cascading rail surface defect identifier first uses the unhealthy rail detector (URD), which filters the standard rails in the set of extracted rails, exploiting as a backbone DenseNet-169. Successively, rails recognized as unhealthy are given to the rail flaw classifier (RFC), which identifies the defect class, merging a feature reduction module and a feature joint learning module.

Two types of rail surface imperfections, cracks, and spalling, are detected in [42], analyzing images captured by the monorail image acquisition system (MIAS). The proposed framework, named CCEANN, realizes offline multi-scale defect detection. Firstly, a histogram-based railroad extractor cut the rail ROI to screen background elements. The defect detection on the cropped rail images employs two steps: CSFA-Hourglass and CASIoU-CHEM. The first is used to extract multi-scale features of faults and then combines these features as the input of the second part. In the second stage, imperfections are located by adaptable bounding boxes based on center-point estimation merged with CASIoU-guided fine coordinate compensation, in which the gradient propagation tool and CASIoU loss will be highlighted. The limits of this system are the slowness and the computational complexity.

To detect the crack of rail surface, in [66], the authors propose a deep transfer learning framework (DTL) that combines two pre-trained networks. First, they define the crack-size threshold to differentiate the ground-truth cracks from the noises in railway pictures. The authors underline the impact of this value on detection performance. Successively, the images are preprocessed to enhance contrast, manually tagged with bounding boxes, and the bounding boxes are filtered based on the crack thresholding. YOLOv3 and RetinaNet are trained on the training and validation set filtered at the previous step. Finally, the filtered test data set is given to the pre-trained network, and the detection ensemble is applied. This ensemble scheme selectively integrates detected boxes of YOLOv3 and RetinaNet models. Exploiting transfer learning, this strategy requires a restricted number of training images. In addition, recall and average precision obtained are satisfying.

To identify three types of rail surface defects, including corrugation, tripping off block, and fatigue block, in [60], the authors propose a deep convolutional network named MOLO, whose backbone exploits MobileNetV2 and combines YOLOv3 multi-scale architecture with the notion of regression. MOLO comprises a conv2d layer, a bottleneck module, and an upsample layer, and it includes a single-scale input/multi-scale output structure. MOLO refers to YOLO, turning target detection topics into target regression issues. The prediction information contains the target object type, confidence, and the bounding box's width-height corresponding to the target object and the center coordinates. Data augmentation is applied to improve the training. This model achieves better robustness and accuracy compared to YOLOv3.

Concerning the detection of railway subgrade imperfections, [54] deals with a modified Faster R-CNN to recognize subgrade settlement, mud pumping, ballast fouling faults, and water abnormality. A vehicle-mounted ground penetration radar obtains the dataset (GPR). At first, the original radar data are processed to reduce the noise signal due to the interference generated by the catenary and other signals. The authors improve the classic Faster R-CNN by employing four techniques to realize automatic defect detection. The feature cascade merges the outcome features of swallow convolutional layers with those of deep convolutional layers to create a new multisized feature; the Adversial Spatial Dropout network generates complex positive instances which are problematic to classify since the sample's dimensionality is restricted, and the shape and proportions of faults are diverse; the Soft Non-Maximum Suppression (Soft-NMS) suppresses redundant detection boxes, reducing the cardinality of false-positive outcomes and improving the performance; finally, data augmentation is applied to unravel the problem of the limited instances in the dataset. The authors chose VGG16 as the primary network. The proposed model achieves better performance than the SVM+HOG architecture and the Faster R-CNN. Moreover, the robustness is guaranteed by the applicability to the detection of different fault classes.

In [41], the authors use transfer learning to detect railroad imperfections (loose ballast, sunkinks) and railway assets (signals and switches) (Figure 5). They proposed a fully automatic defect detection system trained with specially recorded data. Unfortunately, the limitation of this method is due to the lack of sufficient data to train the system on a large scale. The dataset comprises track videos recorded from a camera installed on the roof of a few locomotives. The article is divided into two parts: the first deals with track imperfections detection, and the second is railway assets mapping. To detect sunkinks and loose ballast, ResNet-50, and Inception v3, pretrained on ImageNet data, are compared after replacing some layers in the original models. The models are chosen in such a way as to avoid false negatives for sunk-inks detection and false positives for loose ballast. Concerning the second task, the switch place allows the operator to pay attention, and signal color detection onboard can be used in the operator assistance system. The same procedure as the sunkink architecture has been tested in the switches dataset. In this case, Inception v3 is preferable. The Faster R-CNN model and Single Shot Detector (SSD), trained on the COCO dataset, are used for signal detection. Finally, a track health index has been evaluated to observe

FIGURE 5. An example of railway switch.

the warnings for various railroad faults for a whole track in an area.

Finally, we recall that in [16], [19], [20], [26], [28], [32], and [52] summarized in the previous sections, detection is performed in addition to other tasks.

2) DETECTION OF RAIL FASTENERS' DEFECTS

Reference [6] compares several machine learning and deep learning methods utilized to inspect rail fasteners' defects. HD Cameras mounted on an unmanned aerial vehicle (UAV) capture the images, and three categories of objects are used to train the models: standard fasteners, fractured fasteners, and missing fasteners. The authors first use traditional visual methods, such as the HOG feature with SVM classifiers, but they obtain unfulfilling results due to heterogeneous background, external light influence, and minor components. The authors compare different deep learning methods to improve the results: YOLOv3, an improved YOLOv3 obtained by deleting the feature maps suitable for large target detection, the Faster R-CNN, and the FPN. Faster R_CNN archives good performance.

Also in [19], [20], [34], [47], [51], and [65], detection of rail fasteners is obtained.

3) RAIL TRACK OBSTACLES DETECTION

In this subsection, we focus our attention on deep learning models that realize the detection of obstacles or small objects on the railway track: automatic real-time detection of these objects can support manual driving and allow the timely intervention of the emergency system.

Several methods for obstacle detection founded on computer vision and DL are described in the existing literature, but they require high memory for onboard devices. In [46], the authors observe that to overcome this problem is appropriate to use an edge-computing-based method that improves the cloud-based solutions previously used. This system includes edge nodes responsible for model training, cloud nodes accountable for storing and labeling data sets, and edge devices in the train, which capture the original images, send the photos, and acquire the results from the remote serves. A YOLOv3 architecture is adapted, concatenating a feature fusion module with a detection module to detect obstacles. The authors chose this network to ensure real-time and accurate performance.

Identifying obstacles in the railway area is crucial to ensure safety, particularly for the train driverless. Accordingly, a high-speed detection approach is essential to realize obstacle detection since many picture details need to be analyzed during driverless train runs. On the other hand, high accuracy is necessary since false detection will induce loss of life. In [24], the authors adopt a modified Faster R-CNN named Mask R_CNN. First, ResNet101 is employed as the backbone feature extractor, and, successively, Mask R_CNN utilizes a two-stage approach. In the first step, the region proposal network proposes promising objects bounding boxes. Mask R_CNN produces a binary mask for each ROI while indicating the category in the second stage. Finally, data augmentation and transfer learning are applied. The authors chose this model for its increased accuracy as a two-stage approach and its ability to color the detected target. The data are obtained by selecting the frame from a video acquired in the subway. The obstructions are split into two types: movable obstructions (people) and immovable obstructions, such as helmets, bags, cardboard, and boxes. This approach presents a higher precision rate and a more rapid inspection speed with respect to one-stage methods. However, the inspection speed of this architecture could be additionally improved.

During the quick run of high-speed trains, an exterior substance such as plastic bags on the rail side can enter the bottom bogies and cause smoke, short-circuits, or fires. Since the manual examination is time-consuming, expensive, and not very precise, in [57], the authors employ an automated inspection strategy formed of a high-speed linear camera, high-speed array camera, and a processing unit. This automated system recognizes the abnormal parts of the images, but there is a high error rate. To overcome this issue, the researchers suggest an improved inspection approach based on YOLOv3. Firstly, DenseNet is utilized as the backbone network instead of the classic Darknet-53 to strengthen feature propagation. In addition, three spatial pyramid pooling (SPP) architectures are added to the feature pyramid network. Moreover, a normalization phase after each

SPP layer is added to improve the convergence and stop overfitting. The three prediction layers are added at the end of dense blocks. Finally, transfer learning that exploits the ImageNet data set and data augmentation technologies are exploited to reduce training time and extend the data set.

4) PANTOGRAPH FAULTS DETECTION

In [5], the detection of the corner points is applied in addition to the segmentation task.

5) OCS FAULTS DETECTION

In this subsection, we review several papers concerning the detection of the OCS's components faults, such as the insulator, the bird damage prevention, and fasteners of the current-carrying ring, the catenary, and the catenary clevis.

The insulator is a critical element of the catenary which guarantees the insulation between the catenary and the earth. Before inspecting the insulator surface, it is fundamental to localize the insulator from the complex background of the images. Therefore, the method proposed in [30] consists of two phases: object detection and defect detection. The first task is obtained using a Faster R-CNN, which allows the extraction of six crucial parts, including insulators of the catenary images acquired by the inspection vehicle. Since the grey-scale levels of the insulators are comparable to other catenary components and the defective samples are fewer than the standard images, the authors suggest a deep multitask neural network (DMNN) founded on autoencoders. The DMNN includes a deep material classifier (DMC) and a deep denoising autoencoder (DDAE). The DMC allows for segmenting the insulators; on the other hand, the purpose of the DDAE is to detect imperfections in the insulator picture, reconstructing overlapping patches of the insulator images. Once the DMNN slides via the insulator photo, the DMC produces a classification score map and the DDAE a reconstruction error map. When the anomaly score computed in a particular patch exceeds a predefined threshold, the patch is classified as abnormal. By comparing the proposed method with other DL detection models, the authors show that the Faster R-CNN VGG16 achieves higher average precision than that achieved by other methods, except for stable arm support.

In [36], the authors present an improved Faster R-CNN network to detect the defects of bird-preventing and fasteners, which guarantees the cantilever's connection to the catenary support. Currently to monitor the state of the catenary countless images are viewed by human operators who may be subject to eye strain. The pictures of the catenary support device are acquired by the inspection train at night; hence, they need to be brightened or darkened, depending on the circumstances. Moreover, the dimensions of bird-preventing and fasteners vary significantly, and fasteners are usually tiny. To solve these problems, the authors propose to add a top-down-top feature pyramid fusion system to a Faster R-CNN. The bottom-up pathway is a feed-forward computation of the backbone ConvNet; the top-down architecture is used to build the semantic features by upsampling spatially coarser feature maps from the higher pyramid levels. On the other hand, the down-top pathway propagates the low-layer information. The defects classified by the proposed structure are divided into eight categories. Obviously the system needs to be tested and evaluated on datasets of general use to express all its potential.

Brace sleeve screws are indispensable components in the catenary supporting apparatus of high-speed railways: the fall of the brace sleeve screw could cause several safety problems. In [38], a modified Faster RCNN is presented to realize a high-precision anomaly detection for this critical component, whose dimensions are too small. The dataset is composed of pictures of the support of the catenary acquired by the inspection vehicle. The improved Faster CNN proposed by the authors is executed with VGG-16 and ResNet-101 as a primary network model. This improved network includes two concepts: the proposal map, utilized in the method of region proposal map (RPN), which generates object bounding boxes, and the discriminant map utilized to define the kinds of objects in the proposal region. This proposed method is very important for many little components but has been tested in different conditions.

Due to the different aspects, the current-carrying ring could be broken, and the defective ring would impact the energy supply of the catenary. Reference [8] proposes a defect diagnosis approach founded on improved RetinaNet to detect the ring. The improved model architecture has two embedded blocks: the spatial attention map block (SAMB) and the channel weight map block (CWMB). The modules are embedded in the feature network and feature pyramid network (FPN) stages. The multi-scale feature maps produced by FPN are arranged into the CWMB to improve the essential features in channels. Since the defective current-carry ring is small and the background and obstacles are diverse and abundant in the picture, the SAMB is employed to enhance the local features of each feature map in the spatial attention. In the FPN, the channels of feature maps from each layer are condensed with the depth of the network, which has different information in each channel. The CWMB weighs the variable significance of channels and exploits helpful details among the global feature maps. Pictures are chosen randomly from the videos captured by the inspection train. The backbone networks are VGG19, ResNet50, and Res2Net50.

In [23], an automated visual inspection approach is suggested to find the rupture of the cross-link clevises of the high-speed railway catenary. The images of catenaries are acquired by CCD cameras installed on the top of an inspection vehicle. The methodology of clevis fracture inspection can be split into two phases, clevis extraction and fracture detection. First, the clevises are captured from the catenary photo employing a method founded on the Faster R-CNN network. Here, ResNet101 pre-trained on the ImageNet dataset is assumed as the backbone of the Faster R-CNN. The model of the original faster R-CNN is changed by extending the ROIs by the regional proposal

network to a more extensive scale. Successively, the detection of fractures is founded on the detection of cracks. It is performed by acquiring the edge information of the clevis sub-image using the region-scalable fitting (RSF) structure, which can segment pictures with intensity non-homogeneous. Cracks are detected by computing the wavelet entropy inside the crucial zones and applying morphological filtering. To evaluate the performance of the proposed model in clevis extraction, the authors compare the improved Faster R-CNN with other deep learning networks, showing that it achieves a higher average precision. However, there are some situations in which the model fails to extract clevis. This occurs when other fittings occlude some clevises, or when the component is underexposed, due to defective activation of cameras.

We end this subsection by recalling that [18], [25], [37], and [48] realize the detection of catenary elements faults as described before.

D. OTHER

This section deals with a review of papers in which the authors apply deep learning techniques to analyze the railway infrastructure elements, realizing tasks not included in the previous categories. In particular, here, we consider papers in which the authors use generative models to increase the datasets, such as discovering the enhancement of the catenary images [7], predicting the vehicle-body vibrations, or inspecting the vehicle onboard equipment defects.

Researchers must unravel the issue of the shortage of rail surface faults in track inspection images for a suitable application of deep learning techniques. To this end, in [58], the authors propose an approach founded on a Deep Convolution Generating Adversarial Network (DCGAN) to extend and enhance the defects sample, consisting of images. The DCGAN is given by a CNN and a Generative Adversarial Network (GAN), which improves the GAN. In this paper, first, the authors prove that DCGAN does not unravel the issue of the vanishing gradient completely; successively, after applying a noised-guided optimization, they show that the real sample distribution and the developed ones have a non-negligible overlapping and the Jensen-Shannon divergence makes the two distributions closer until they are totally coincident. In [13], a generative model is proposed to increase the sample size, and in [35], the authors create new samples to alleviate the problem of classes unbalancing.

Predicting vehicle-body vibrations is advantageous for finding railroad imperfections, and the expected accelerations can be utilized as an additional index for evaluating track quality. Reference [39] provides an approach to indicate vehicle-body vibration established on DL: by integrating CNN and long short-term memory (LSTM), a CNN-LSTM architecture is suggested. While CNN-LSTM creates a point-wise prediction, it requires as input a part of track geometry. Here, this information allows CNN to acquire shape features. The three input channels are given by average alignment, average longitudinal level, and cross-level, and they are validated to reach the best performance among all possible mixtures. The lateral and the vertical vehicle-body acceleration (VVBA and LVBA) include the two-channel outcome. CNN-LSTM assumes the format of alternating convolutional and pooling layers. Each LSTM cell conforms to an individual pair of VBAs. After LSTM layers, two fully connected layers are assumed as a classifier, which finally creates the predicted VBA. In addition, the authors show that the CNN-LSTM model achieves higher accuracy than the fully-connected neural network assumed in the performance-based track geometry (PBTG) technology and the basic LSTM.

Reference [59] deals with a deep learning architecture to detect onboard vehicle equipment (VOBEs) defects for high-speed trains. VOBEs are fundamental components for high-speed railways (HSRs) since they transmit to the trains important information such as the length of the track circuits, speed limits, and running speeds. The fault diagnosis of VOBE consists of finding out the reasons, called fault reasons, that force the train to break down. The fault reasons are based on fault evidence detected by onboard drivers. The authors define two vectors: the fault evidence vector and the fault reason vector. The first one is given as input to the suggested architecture, and the outcome is the detect fault reason vector. The model consists of a Deep Belief Network (DBN) composed of many stacked restricted types of Boltzmann machines (RBM). The performance reached by the proposed architecture exceeds the k-nearest neighbors algorithm and the artificial neural network with back propagations, improving the accuracy of defects diagnosis for VOBE.

IV. EVALUATION METRICS

Comparing the performance of papers devoted to solving different problems using different data is challenging and, perhaps, even trivial in a few cases. Nonetheless, it might help in pointing out the kind of results that are achievable with these modern neural network architectures.

To evaluate the performance of the proposed methods, the authors exploit different metrics based on the task accomplished. More in detail, the most used metric for classification is accuracy (A), given by the ratio between the number of samples correctly classified and the cardinality of the dataset. In particular, if data are classified into two classes, the following formula gives the binary accuracy:

$$A = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP and TN denote the true positive and true negative, i.e., the number of positive and negative samples correctly classified. On the other hand, FP and FN indicate the number of positive and negative samples misclassified as belonging to the other class.

Moreover, pixel accuracy (PA) and mean intersection over union ($mIoU$) are often used to evaluate segmentation

TABLE 5. The metrics most employed to evaluate the models' performance.

Task	Dataset	Ref.	Metrics
Classification	Authors' dataset	[3]	$A = 97.1\%$
	NEU[1]	[2]	$A = 99.27\%$[2]
	Authors' dataset: FUB		$A = 98.33\%$
	Authors' dataset: Crack	[26]	$A = 98.57\%$ (training)
			$A = 95.04\%$ (validation)
	Authors' dataset: Rust		$A = 90.32\%$ (training
			$A = 82.76\%$ (validation)
	Authors' dataset: pan5-sG256	[33]	$A = 90.625\%$
	Authors dataset	[14]	$A = 92.93\%$ (small DCNN)
			$A = 92.77\%$ (medium DCNN)
			$A = 93.04\%$[3] (large DCNN)
	Authors dataset	[13]	$A = 99.7\%$
	Authors dataset	[12]	$A = 99.74\%$
	Authors dataset	[4]	Accuracy N.A.
	Authors dataset	[63]	Accuracy N.A.
	Authors dataset	[27]	$A = 90.34\%$
	Authors dataset: Line-1	[49]	$A = 97.6\%$[4]
	Authors dataset: Line-2		$A = 94.23\%$
	Authors dataset	[32]	$A = 96.55\%$
	Authors dataset	[34]	Accuracy N.A.
	Authors dataset	[43]	$A = 97.2\%$
	Authors dataset	[51]	$A = 99.26\%$
	Authors dataset	[65]	$A = 98.65\%$
	Authors dataset	[47]	$A = 93.5\%$
	Authors dataset	[19]	$A = 93.35\%$
	Authors dataset	[20]	$A = 95.02\%$
	Authors dataset	[48]	$A = 98.72\%$
	Authors dataset	[35]	$A = 93.25\%$
	Authors dataset	[37]	$A = 94.92\%$[5]
	Authors dataset	[25]	$A = 99\%$
Segmentation	Authors dataset: NRSD-MN Man-made	[62]	$PA = 71\%$
			$mIoU = 37\%$
	Authors dataset: NRSD-MN Natural		$PA = 70\%$
			$mIoU = 42\%$
	Authors dataset	[12]	PA N.A.
			$mIoU$ N.A.
	Authors dataset	[7]	PA N.A.
			$mIoU$ N.A.
	Authors dataset: BH-rail-dataset	[50]	$MPA = 99.15\%$
			$mIoU = 98.46\%$
	Authors dataset	[52]	PA N.A.
			$mIoU$ N.A.
	RSDD Type I[6]	[64]	PA N.A.
			$mIoU$ N.A.
	RSDD Type II [7]		PA N.A.
			$mIoU$ N.A.
	Authors dataset	[28]	PA N.A.
			$mIoU$ N.A.
	Authors dataset	[32]	PA N.A.
			$mIoU$ N.A.
	Authors dataset RSDD	[65]	$PA = 99.72\%$
			$mPA = 94.37\%$

TABLE 5. *(Continued.)* **The metrics most employed to evaluate the models' performance.**

			$mIoU = 87.52\%$
	Authors dataset	[27]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[51]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[47]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[19]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[20]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[48]	$PA = 91.26\%$ $mPA = 81.54\%$ $mIoU = 77.92\%$
	Authors dataset	[5]	*PA* N.A. *mIoU* N.A.
	Authors dataset	[18]	*PA* N.A. $mIoU = 89\%$[8]
Detection	Authors dataset: Crack	[26]	*mAP* N.A.
	Authors dataset: Rust		*mAP* N.A.
	PASCAL VOC 2012 COCO 2017	[46]	*mAP* N.A.
	Authors dataset	[24]	$mAP = 95.7\%$
	Authors dataset	[57]	$mAP = 94.5\%$
	PASCAL VOC	[55]	*mAP* N.A.
	Authors dataset	[16]	$mAP = 87.40\%$ (M2-Y2) $mAP = 82.91\%$ (M3-Y3)
	RSDD - Type I RSDD - Type II	[61]	*mAP* N.A. *mAP* N.A.
	Authors dataset	[1]	*mAP* N.A.
	D_I industrial partner D_{II} open-source	[67]	*mAP* N.A.
	Authors dataset	[52]	*mAP* N.A.
	Authors dataset RSDD NEU RSDDS-113	[42]	$AP = 92.45\%$ $AP_L = 76.74\%$[9]
	RSDD Type I	[64]	*mAP* N.A.
	RSDD Type II		*mAP* N.A.
	Authors dataset[10] RSDD Type I	[66]	$AP = 78\%$[11]
	Authors dataset	[28]	*mAP* N.A.
	Authors dataset	[60]	*mAP*=87.4%
	Authors dataset	[32]	*AP*= 99.99%[12]
	Authors dataset	[6]	$mAP = 95.78\%$[13]
	Authors dataset	[34]	$mAP = 84.08\%$
	Authors dataset	[51]	$mAP = 97.9\%$
	Authors dataset	[65]	$mAP = 99.68\%$
	Authors dataset	[47]	$mAP = 96.13\%$
	Authors dataset	[54]	*mAP* N.A.
	Authors dataset	[19]	*mAP* N.A.
	Authors dataset	[20]	*mAP* N.A.
	Authors dataset	[41]	*mAP* N.A.
	Authors dataset	[30]	*mAP* N.A.
	Authors dataset	[36]	$mAP = 81.2\%$
	VOC2007		$mAP = 80.1\%$

TABLE 5. *(Continued.)* **The metrics most employed to evaluate the models' performance.**

	Authors dataset	[48]	$mAP = 95.26\%$[14] $mAP = 99.06\%$[15]
	Catenary-5000	[38]	$mAP = 85.34\%$
	Authors dataset	[5]	mAP N.A.
	Authors dataset	[8]	$mAP = 70.4\%$
	Authors dataset	[37]	$mAP = 89.10\%$[16]
	Authors dataset	[25]	$AP = 92.7\%$
	Authors dataset	[23]	$AP = 94.42\%$ (left)[17] $AP = 93.72\%$ (right)
	Authors dataset	[18]	mAP N.A.
Other	Authors dataset	[13]	$t - SNE$
	Authors dataset	[7]	$E = 6.9602$[18] $AG = 4.9448$[19] $e = 5.8148$[20] $r = 3.7983$[21]
	Authors dataset	[39]	$MAE/g = 0.0062$[22] $RMSE/g = 0.008$[23] $PCC = 0.898$[24]
	Authors dataset	[59]	$A \approx 90\%$
	Authors dataset	[58]	$IS = 1.96$[25]
	Authors dataset	[35]	$FID = 77.539$[26] (damaged fastener) $FID = 134.886$ (missing fastener)

1 Northeastern University Dataset.
2 Best accuracy value obtained with respect to batch size and every epoch training time.
3 Best accuracy value obtained with respect to the activation function.
4 Best mean accuracy value reached by the "ResNet_all" model and obtained by the average of four class accuracy.
5 Obtained by DFRN.
6 Express rails.
7 Common/heavy rails.
8 Achieved by "Adaptive SSN" at Stage 3.
9 Average Precision on large defects.
10 From the China Railway Corporation.
11 Best result obtained by DLT method with respect to different confidence and crack size threshold.
12 Best precision at intersection over union (IOU) reached by "ResNet50-FPN" with threshold 0.8.
13 Best accuracy reached by FPN with respect to other models. This is a comparative study.
14 First localization.
15 Second localization.
16 Obtained by DPNL.
17 Average Precision for clevis extraction.
18 Information Entropy
19 Average Gradient
20 Rate of New Visibility
21 Mean of Visual Edge Gradient
22 Mean Absolute Error
23 Root Mean Square Error
24 Pearson Correlation Coefficient
25 Inception Score
26 Fréchet Inception Distance

algorithms. The following formulas give them:

$$PA = \frac{TP}{G},$$

and

$$mIoU = \frac{TP}{G + T - TP},$$

where TP denotes the number of true positives (correct prediction pixels for foreground). G and T indicate the number of pixels in the foreground of the ground-truth and prediction map, respectively [62]. Moreover, mean pixel accuracy (MPA) is also used in [50].

Finally, mean average precision (mAP) is employed to evaluate the performance of the detection algorithm. More in detail, the average precision (AP) over 11 spaced recall levels $[0, 0.1, \ldots, 1]$ is given by:

$$AP = \frac{1}{11} \sum_{r \in \{0, 0.1, \ldots, 1\}} P_{interp}$$

and

$$mAP = \frac{sum(AP)}{N}.$$

is the average AP in the classes [24].

Many articles also contain evaluation metrics for each class of defects analyzed or different values for each experiment realized. For example, the researchers provide additional values for each batch size or activation function choice.

Finally, we observe that papers on the same task analyze datasets differently. Therefore, the performance of the proposed architecture is influenced by the characteristics of the data, such as cardinality, the unbalancing between classes, the image quality, and the shape or size of the defect or object analyzed.

In Table 5, we present some relevant metrics described by the authors, highlighting the task accomplished (first column) and the kind of dataset (second column). Concerning the values assumed by the metric, we chose the best performance reached from the model. Concerning the papers in which the proposed method is used to accomplish different tasks, we split the metrics based on the study.

For layout needs, in Table 5 we shorten ''not available'' with ''N.A.'' and the ''Classification'', ''Segmentation'' and ''Detection'' tasks with ''Class.'', ''Seg.'', and ''Det.'', respectively.

V. CONCLUSION AND FUTURE WORK

In this work, an in-depth analysis of artificial intelligence techniques was carried out, focusing on identifying anomalies in the railway infrastructure. The study showed that the scientific community has increasingly relied on deep learning techniques compared to machine learning ones. This growing interest in deep learning techniques has been caused by the fact that deep learning is able to operate on raw and, sometimes, unstructured data. In fact, the peculiar characteristic of an approach established on DL is to identify by itself the distinctive features suitable for performing a specific task without the need for training by a human operator. On the other hand, an aspect that should be highlighted is that the deep-learning ability to extract features without guidance is a double-edged sword: they are indeed like black boxes within which it is not possible to trace the selection path that the learning process selected. Therefore, in strategic and highly complex sectors such as railway safety, the presence of a human user is still central and must not be completely replaced by a machine. In fact, the experience acquired by the human being provides an ability to perceive events, that are difficult to emulate by the machine. For this reason, it is essential to exploit both the advantages of using AI-based techniques that can assist experienced operators in this context, with the aim of developing semi-supervised solutions. In fact, in this context, deep learning is proving to be the most reliable solution currently available.

The review is focused on the monitoring of the railway infrastructure for preventive maintenance purposes.

Comparing the papers shows that most analyzed components relied on the rail area (Table 1), particularly on the site where the train is moving. This is probably due to the fact that anomalies concerning the rail surface or the fastening system could mainly affect transport safety, causing severe accidents and even derailments in particular. Another possible explanation concerns the fact that the rail track is more prone to friction and rubbing, and this causes broken or lost components. Table 2 shows that most of the datasets analyzed are images rather than video or signals, and Table 3 shows that the attention of researchers is devoted to the detection task. In addition, Table 4 suggests that researchers do not pay much attention to the parameter ''adaptability'': only a few works test the proposed model on a different dataset.

Nonetheless, 46 of the 53 papers analyzed achieved remarkable objectives, as listed in Table 4. Finally, from Table 5, two fundamental aspects are deduced. First, it is impossible to draw a unique comparison between all the papers taken into account since no particular dataset was analyzed. Still, the authors own most of the dataset. Second, not all documents exploit the same metrics, although they report high values regarding performance evaluation. Therefore, possible future work in this area would be achieved by creating a unique reference dataset that the scientific community could use.

The conducted systematic analysis allowed us to identify multiple common functional clusters, thanks to the context awareness achieved by deep learning models, useful for different research objectives. The obtained result is an in-depth overview of the solutions to real-world problems approached with deep learning in the railway sector. All this represents the basis for subsequent studies in which the best-performing approaches will be selected based on the context specifications, and these same algorithms will be tested and compared to understand the advantages and disadvantages of the most promising techniques. Once the most suitable algorithms have been identified to solve specific tasks in the railway sector, the goal is to integrate them into more complex systems, for example, virtual and augmented reality systems, with the aim to increase and improve support for a human operator.

ACKNOWLEDGMENT

The authors would like to thank Michele Attolico, Giuseppe Bono, and Paola Romano involved in the VRail project ''Virtual and Augmented Reality for Railways'' Prog n. F/190009/02/X44–CUP: B81B19001410008–COR:1666959.

REFERENCES

[1] C. Akhila, C. A. Diamond, and A. M. Posonia, ''Convolutional neural network based online rail surface crack detection,'' in *Proc. 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2021, pp. 1602–1606.

[2] A. Aydin, M. U. Salur, and I. Aydin, ''Fine-tuning convolutional neural network based railway damage detection,'' in *Proc. 19th Int. Conf. Smart Technol. (EUROCON)*, Jul. 2021, pp. 216–221.

[3] I. Aydin, E. Akin, and M. Karakose, "Defect classification based on deep features for railway tracks in sustainable transportation," *Appl. Soft Comput.*, vol. 111, Nov. 2021, Art. no. 107706.

[4] S. Bahamon-Blanco, S. Rapp, Y. Zhang, J. Liu, and U. Martin, "Recognition of track defects through measured acceleration using a recurrent neural network," *Int. J. Comput. Methods Exp. Meas.*, vol. 8, no. 3, pp. 270–280, 2020.

[5] L. Chang, Z. Liu, and Y. Shen, "On-line detection of pantograph offset based on deep learning," in *Proc. IEEE 3rd Optoelectron. Global Conf. (OGC)*, Sep. 2018, pp. 159–164.

[6] P. Chen, Y. Wu, Y. Qin, H. Yang, and Y. Huang, "Rail fastener defect inspection based on UAV images: A comparative study," in *Proc. 4th Int. Conf. Elect. Inf. Technol. Rail Transp. (EITRT)*, B. Liu, L. Jia, Y. Qin, Z. Liu, L. Diao, and M. An, Eds., Singapore: Springer, 2020, pp. 685–694.

[7] Y. Chen, B. Song, X. Du, and N. Guizani, "The enhancement of catenary image with low visibility based on multi-feature fusion network in railway industry," *Comput. Commun.*, vol. 152, pp. 200–205, Feb. 2020.

[8] Y. Chen, B. Song, Y. Zeng, X. Du, and M. Guizani, "Fault diagnosis based on deep learning for current-carrying ring of catenary system in sustainable railway transportation," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106907.

[9] M. C. Nakhaee, D. Hiemstra, M. Stoelinga, and M. van Noort, "The recent applications of machine learning in rail track maintenance: A survey," in *Reliability, Safety, and Security of Railway Systems. Modelling, Analysis, Verification, and Certification*, S. Collart-Dutilleul, T. Lecomte, and A. Romanovsky, Eds., Cham, Switzerland: Springer, 2019, pp. 91–105.

[10] A Company. *MERMEC Company Profile*. [Online]. Available: https://www.mermecgroup.com

[11] NDT Company. *NDT Company Profile*. [Online]. Available: https://www.ndttech.com/

[12] H. Cui, J. Li, Q. Hu, and Q. Mao, "Real-time inspection system for ballast railway fasteners based on point cloud deep learning," *IEEE Access*, vol. 8, pp. 61604–61614, 2020.

[13] T. de Bruin, K. Verbert, and R. Babuška, "Railway track circuit fault diagnosis using recurrent neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 523–533, Mar. 2017.

[14] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska, and B. De Schutter, "Deep convolutional neural networks for detection of rail surface defects," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2584–2589.

[15] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, and L. Chen, "Automatic fastener classification and defect detection in vision-based railway inspection systems," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 877–888, Apr. 2014.

[16] J. H. Feng, H. Yuan, Y. Q. Hu, J. Lin, S. W. Liu, and X. Luo, "Research on deep learning method for rail surface defect detection," *IET Electr. Syst. Transp.*, vol. 10, no. 4, pp. 436–442, Dec. 2020.

[17] "Report on railway safety and interoperability in the EU—2020," Eur. Union Agency Railways, Tech. Rep., 2020.

[18] S. Gao, G. Kang, L. Yu, D. Zhang, X. Wei, and D. Zhan, "Adaptive deep learning for high-speed railway catenary swivel clevis defects detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 1299–1310, Feb. 2022.

[19] X. Giben, V. M. Patel, and R. Chellappa, "Material classification and semantic segmentation of railway track images with deep convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 621–625.

[20] X. Gibert, V. M. Patel, and R. Chellappa, "Deep multitask learning for railway track inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 153–164, Jan. 2017.

[21] TESMEC Group. *TESMEC Company Profile*. [Online]. Available: https://www.tesmec.com/it

[22] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1577–1586.

[23] Y. Han, Z. Liu, Y. Lyu, K. Liu, C. Li, and W. Zhang, "Deep learning-based visual ensemble method for high-speed railway catenary clevis fracture detection," *Neurocomputing*, vol. 396, pp. 556–568, Jul. 2020.

[24] D. He, K. Li, Y. Chen, J. Miao, X. Li, S. Shan, and R. Ren, "Obstacle detection in dangerous railway track areas by a convolutional neural network," *Meas. Sci. Technol.*, vol. 32, no. 10, Oct. 2021, Art. no. 105401.

[25] C. Huang and Y. Zeng, "The fault diagnosis of catenary system based on the deep learning method in the railway industry," in *Proc. 5th Int. Conf. Multimedia Image Process. (ICMIP)*. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 135–140.

[26] S. Iyer, T. Velmurugan, A. H. Gandomi, V. N. Mohammed, K. Saravanan, and S. Nandakumar, "Structural health monitoring of railway tracks using IoT-based multi-robot system," *Neural Comput. Appl.*, vol. 33, no. 11, pp. 5897–5915, Jun. 2021.

[27] A. James, W. Jie, Y. Xulei, Y. Chenghao, N. B. Ngan, L. Yuxin, S. Yi, V. Chandrasekhar, and Z. Zeng, "TrackNet—A deep learning based fault detection for railway track inspection," in *Proc. Int. Conf. Intell. Rail Transp. (ICIRT)*, Dec. 2018, pp. 1–5.

[28] X. Jin, Y. Wang, H. Zhang, H. Zhong, L. Liu, Q. M. J. Wu, and Y. Yang, "DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1051–1065, Apr. 2020.

[29] Y. Jin, "Wavelet scattering and neural networks for railhead defect identification," *Materials*, vol. 14, no. 8, p. 1957, Apr. 2021.

[30] G. Kang, S. Gao, L. Yu, and D. Zhang, "Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 8, pp. 2679–2690, Aug. 2019.

[31] A. Lasisi and N. Attoh-Okine, "Machine learning ensembles and rail defects prediction: Multilayer stacking methodology," *ASCE-ASME J. Risk Uncertainty Eng. Syst., A, Civil Eng.*, vol. 5, no. 4, Dec. 2019.

[32] X. Li, Y. Zhou, and H. Chen, "Rail surface defect detection based on deep learning," in *Proc. 11th Int. Conf. Graph. Image Process. (ICGIP)*, Z. Pan and X. Wang, Eds., vol. 11373, 2020, Art. no. 113730K.

[33] Y. Li and X. Wei, "Pantograph slide plate abrasion detection based on deep learning network," in *Proc. 3rd Int. Conf. Electr. Inf. Technol. Rail Transp. (EITRT)*, L. Jia, Y. Qin, J. Suo, J. Feng, L. Diao, M. An, Eds. Singapore: Springer, 2018, pp. 215–224.

[34] Y.-W. Lin, C.-C. Hsieh, W.-H. Huang, S.-L. Hsieh, and W.-H. Hung, "Railway track fasteners fault detection using deep learning," in *Proc. IEEE Eurasia Conf. IoT, Commun. Eng. (ECICE)*, Oct. 2019, pp. 187–190.

[35] J. Liu, Z. Ma, Y. Qiu, X. Ni, B. Shi, and H. Liu, "Four discriminator cycle-consistent adversarial network for improving railway defective fastener inspection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 10636–10645, Aug. 2022.

[36] J. Liu, Y. Wu, Y. Qin, H. Xu, and Z. Zhao, "Defect detection for bird-preventing and fasteners on the catenary support device using improved faster R-CNN," in *Proc. 4th Int. Conf. Electr. Inf. Technol. Rail Transp. (EITRT)*, B. Liu, L. Jia, Y. Qin, Z. Liu, L. Diao, and M. An, Eds. Singapore: Springer, 2020, pp. 695–704.

[37] S. Liu, L. Yu, and D. Zhang, "An efficient method for high-speed railway dropper fault detection based on depthwise separable convolution," *IEEE Access*, vol. 7, pp. 135678–135688, 2019.

[38] Z. Liu, Y. Lyu, L. Wang, and Z. Han, "Detection approach based on an improved faster RCNN for brace sleeve screws in high-speed railways," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 7, pp. 4395–4403, Jul. 2020.

[39] S. Ma, L. Gao, X. Liu, and J. Lin, "Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction," *IEEE Access*, vol. 7, pp. 185099–185107, 2019.

[40] MERMEC. *Roger Inspection Vehicle Web Page*. [Online]. Available: https://www.mermecgroup.com/measuring-trains-br-and-systems/recording-cars/105/roger-400.php

[41] S. Mittal and D. Rao, "Vision based railway track monitoring using deep learning," 2017, *arXiv:1711.06423*.

[42] X. Ni, Z. Ma, J. Liu, B. Shi, and H. Liu, "Attention network for rail surface defect detection via consistency of intersection-over-union (IoU)-guided center-point estimation," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1694–1705, Mar. 2022.

[43] S. P. Orlov, R. V. Girin, and A. V. Piletskaya, "Intelligent information processing system for monitoring rail tracks," in *Proc. 3rd Int. Conf. Control Tech. Syst. (CTS)*, Oct. 2019, pp. 233–236.

[44] S. Qi, J. Yang, and Z. Zhong, "A review on industrial surface defect detection based on deep learning technology," in *Proc. 3rd Int. Conf. Mach. Learn. Mach. Intell. (MLMI)*. New York, NY, USA: Association for Computing Machinery, Sep. 2020, pp. 24–30.

[45] Y. Santur, M. Karaköse, and E. Akin, "Random forest based diagnosis approach for rail fault inspection in railways," in *Proc. Nat. Conf. Electr., Electron. Biomed. Eng. (ELECO)*, Dec. 2016, pp. 745–750.

[46] S. Li, H. Zhao, and J. Ma, "An edge computing-enabled train obstacle detection method based on YOLOv3," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–9, Oct. 2021.

[47] Z. Tu, S. Wu, G. Kang, and J. Lin, "Real-time defect detection of track components: Considering class imbalance and subtle difference between classes," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[48] J. Wang, L. Luo, W. Ye, and S. Zhu, "A defect-detection method of split pins in the catenary fastening devices of high-speed railway based on deep learning," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9517–9525, Dec. 2020.

[49] P. Dai, X. Du, S. Wang, Z. Gu, and Y. Ma, "Rail fastener automatic recognition method in complex background," in *Proc. 10th Int. Conf. Digit. Image Process. (ICDIP)*, Aug. 2018, Art. no. 1080625.

[50] Z. Wang, X. Wu, G. Yu, and M. Li, "Efficient rail area detection using convolutional neural network," *IEEE Access*, vol. 6, pp. 77656–77664, 2018.

[51] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, and Y. Li, "Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study," *Eng. Appl. Artif. Intell.*, vol. 80, pp. 66–81, Apr. 2019.

[52] Y. Wu, Y. Qin, Y. Qian, F. Guo, Z. Wang, and L. Jia, "Hybrid deep learning architecture for rail surface segmentation and surface defect detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 37, no. 2, pp. 227–244, Feb. 2022.

[53] J. Xie, J. Huang, C. Zeng, S.-H. Jiang, and N. Podlich, "Systematic literature review on data-driven models for predictive maintenance of railway track: Implications in geotechnical engineering," *Geosciences*, vol. 10, no. 11, p. 425, Oct. 2020.

[54] X. Xu, Y. Lei, and F. Yang, "Railway subgrade defect automatic recognition method based on improved faster R-CNN," *Sci. Program.*, vol. 2018, pp. 1–12, Jun. 2018.

[55] S. Yanan, Z. Hui, L. Li, and Z. Hang, "Rail surface defect detection method based on YOLOv3 deep learning networks," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2018, pp. 1563–1568.

[56] C. Yang, Y. Sun, C. Ladubec, and Y. Liu, "Developing machine learning-based models for railway inspection," *Appl. Sci.*, vol. 11, no. 1, p. 13, Dec. 2020.

[57] Z. Yao, D. He, Y. Chen, B. Liu, J. Miao, J. Deng, and S. Shan, "Inspection of exterior substance on high-speed train bottom based on improved deep learning method," *Measurement*, vol. 163, Oct. 2020, Art. no. 108013.

[58] L. Yifan, M. Yongzhi, and L. Banghuan, "Research on enhancement method of track defect sample based on deep convolution generative adversarial network," in *Proc. IEEE 2nd Int. Conf. Civil Aviation Saf. Inf. Technol. (ICCASIT*, Oct. 2020, pp. 331–335.

[59] J. Yin and W. Zhao, "Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach," *Eng. Appl. Artif. Intell.*, vol. 56, pp. 250–259, Nov. 2016.

[60] H. Yuan, H. Chen, S. Liu, J. Lin, and X. Luo, "A deep convolutional neural network for detection of rail surface defect," in *Proc. IEEE Vehicle Power Propuls. Conf. (VPPC)*, Oct. 2019, pp. 1–4.

[61] D. Zhang, K. Song, Q. Wang, Y. He, X. Wen, and Y. Yan, "Two deep learning networks for rail surface defect inspection of limited samples with line-level label," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6731–6741, Oct. 2021.

[62] D. Zhang, K. Song, J. Xu, H. Dong, and Y. Yan, "An image-level weakly supervised segmentation method for no-service rail surface defect with size prior," *Mech. Syst. Signal Process.*, vol. 165, Feb. 2022, Art. no. 108334.

[63] X. Zhang, K. Wang, Y. Wang, Y. Shen, and H. Hu, "An improved method of rail health monitoring based on CNN and multiple acoustic emission events," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2017, pp. 1–6.

[64] Z. Zhang, M. Liang, and Z. Wang, "A deep extractor for visual rail surface inspection," *IEEE Access*, vol. 9, pp. 21798–21809, 2021.

[65] D. Zheng, L. Li, S. Zheng, X. Chai, S. Zhao, Q. Tong, J. Wang, and L. Guo, "A defect detection method for rail surface and fasteners based on deep convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–15, Jul. 2021.

[66] Z. Zheng, H. Qi, L. Zhuang, and Z. Zhang, "Automated rail surface crack analytics using deep data-driven models and transfer learning," *Sustain. Cities Soc.*, vol. 70, Jul. 2021, Art. no. 102898.

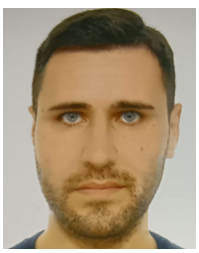
[67] L. Zhuang, H. Qi, and Z. Zhang, "The automatic rail surface multi-flaw identification based on a deep learning powered framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 12133–12143, Aug. 2022.

[68] L. Zhuang, L. Wang, Z. Zhang, and K. L. Tsui, "Automated vision inspection of rail surface cracks: A double-layer data-driven framework," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 258–277, Jul. 2018.

MARIA DI SUMMA received the bachelor's and Ph.D. degrees from Politecnico di Bari. She has been a Researcher with STIIMA-CNR, since 2010. Her research interests include virtual, augmented, and mixed reality, people and posture tracking, intelligent systems, AI paradigms for behavior learning and understanding, feature extraction for pattern recognition, and anomaly detection. She served as the Industrial Session Chair for EuroVR 2016. She has collaborated, covering different roles, actively in several regional, national, and international projects, focuses on advanced manufacturing and ambient assisted living.

MARIA ELENA GRISETA was born in Mola di Bari, in 1988. She received the Ph.D. degree in computer science and mathematics from the University of Bari, in 2019. From 2020 to 2021, she taught probability and statistics with the Computer Science Institute, University of Bari. From 2021 to 2023, she was a Research Fellow with the National Research Council, Bari, Italy. Currently, she is a Researcher in probability and statistics with the Mathematical Institute, University of Bari. Her research interests include artificial intelligence techniques applied to anomaly detection in different fields, such as medicine and railway transportation.

NICOLA MOSCA received the degree (summa cum laude) in computer science from the University of Bari, Bari, Italy, in 2004, and the master's (by research) degree in transport systems engineering from the University of South Australia, in 2012. He was a Contract Researcher with the Institute of Intelligent Systems for Automation, Bari, from 2004 to 2009. He applied for a contest organized by the Italian Apulia Region, won a scholarship for attending the master's by research, in 2009. He has been a Researcher with CNR, since 2014. His research interests include human–computer interaction, 3-D visualization of sport events, video analysis, computer vision for quality control, video surveillance, and intelligent transportation systems.



COSIMO PATRUNO received the B.S. and M.S. degrees in automation engineering from the Polytechnic University of Bari, Italy, in 2010 and 2013, respectively. From May 2013 to January 2022, he was cooperated with the National Research Council of Italy (CNR), as a Research Fellow. Since February 2022, he has been a Permanent Researcher with STIIMA-CNR. His research interests include 3-D data analysis, signal and image processing, computer vision, machine learning, and perception systems for robotics.

MASSIMILIANO NITTI has been a Researcher with CNR, since 2000. He has 20 years of experience in software design and development on the Intel platform and Windows/Linux operating systems, interfacing of both frame and linear cameras, adaptation of image processing algorithms to realistic working conditions, and development of algorithms on FPGA and GPU. He is the coauthor of numerous articles in international scientific journals and conferences and the author of book chapters and international patents. His research interests are related to artificial vision and robotics, with a particular focus on high-performance computing architectures, the development of image processing algorithms, and the development of prototypes and demonstrators for real-time validation of research results.

VITO RENÒ received the master's degree (Hons.) in computer engineering and the Ph.D. degree in electrical and information engineering from Politecnico di Bari, in 2011 and 2017, respectively. He was defending a thesis about "Computer Vision and Robust Background Modeling" and "3-D Modeling, Reconstruction and Analysis of Environments Assisted by Multi-Sensorial Data Processing" during the master's and Ph.D. studies, respectively. He is currently a Researcher with STIIMA-CNR, and has involved in research activities in the fields of computer vision and pattern recognition. He is the coauthor of more than 40 scientific papers and one international patent. He is deeply curious and enthusiast about artificial intelligence, with a pinch of multi-disciplinary and synergic applications.

ETTORE STELLA is the Research Director with ISSIA-CNR, working on computer vision and robotics. Since 2004, he has been an Associate Professor of computer science with the TLC Engineering Faculty, University of Basilicata, Matera. He is a scientific chief of several research projects. He is the coauthor of more than 100 papers on international journals, proceeding of conferences, and international patents. His activities include industrial automation, robotics, computer vision, high-performance computing, and soft-ware design and development.

Open Access funding provided by 'Consiglio Nazionale delle Ricerche-CARI-CARE-ITALY' within the CRUI CARE Agreement