Improving Rail Tracks Defect Classification based on a Cascade Swin-transformer Model

FIELD OF THE INVENTION

[0001] The present invention relates to a computer-implementable system that classifies defects in railway tracks, more specifically, the computer-implementable system employs a two-stage Transformer model to detect defects in railway tracks based on false-alarm-labelled images.

BACKGROUND OF THE INVENTION

[0002] Rail track defect detection often relies on specialized inspection systems, and one such system is the On-board Rail Inspection System (ORIS). ORIS is engineered to identify defects in rail tracks while trains are operational, making it as a valuable tool for ensuring rail network safety and maintenance. However, a notable hurdle encountered with this system is its propensity to generate a considerable number of false alarms. This issue becomes particularly evident when attempting to distinguish between images of normal, well-maintained rail tracks and those exhibiting defects. The high false alarm rate poses challenges in terms of system efficiency and accuracy.

[0003] One of the essential tasks to safeguard the railway is through railway inspection. The ability to detect defects at the early stage enables timely rail maintenance or replacement, and it will also prevent catastrophic failures, such as rail derailments and rail breaks. A desirable defect detection system should be efficient, non-destructive, robust, and reliable. Axle box acceleration (ABA) measurement is a common method to identify railway anomalies.

[0004] There are two ways to detect the defect on the rail track: Non-destructive and destructive methods. Non-destructive methods involve inspecting, testing or evaluating materials, components, or assemblies for discontinuities or differences in characteristics without destroying the serviceability of the part or system. On the other hand, destructive methods are not typically employed for regular track maintenance or inspection because they result in damage. However, destructive methods might be used in controlled environments for research, material testing, or analysing failed components post-incident. Furthermore,

destructive and non-visual detection techniques, including acoustic, ultrasonic, eddy-current, and magnetic flux leakage tests, are employed in identifying internal anomalies like voids and inclusions.

[0005] Previously, simulated ABA signals and the model of wheel-rail interaction were analysed. The signals are processed by the short-time Fourier transform, continuous wavelet transform, empirical mode decomposition, and power spectral density to estimate the likelihood and severity of different rail surface defects. By utilizing algorithms that incorporate both the windowed Fourier transformer and wavelet transformer, different rail defects can be detected from data collected by an instrumented freight car. Another common measurement is the air-coupled ultrasonic rail inspection system. However, the mismatch of high acoustic impedance between air and steel severely undermines the detection performance. There is a need in the industry to build an efficient and non-destructive system and method with pre-trained model for robust and reliable detection performance.

[0006] On the other hand, a 2D model is created to predict the performance of the system, which can work with different defect sizes in various positions. The 2D laser displacement sensor is used for monitoring rail corrugation. However, the pitching vibration under the measurement environment may lead to measurement error in the laser sensor. The model may be able to take the deviation angle and the offset into account and calibrate the dataset under pitching vibration using Laser ultrasonic technology (LUT). LUT can be used by matching pursuit (MP) to develop a non-contact inspection system for rails. It can sense Rayleigh waves generated by laser excitation to detect defects under various locations such as subsurface horizontal, subsurface vertical, surface horizontal, and surface edge. Additionally, electromagnetic tomography can facilitate rail crack detection. The back-projection algorithm can be applied in electromagnetic tomography to classify rail cracks. However, the above technologies require expensive equipment which only experts would be able to handle.

[0007] An efficient way for classifying or detecting rail surface defects uses computer vision. Computer vision approaches, such as those using Convolutional Neural Networks (CNN) or traditional image processing, are adept at spotting surface irregularities like cracks, squats, and corrugation due to their distinct visual features. For example, image data are collected from a real-time inspection system, such as a line-scan camera on the rail surface. It is less time-intensive and more effective. Traditional machine learning algorithms such as SVM (Support

Vector Machine) and Random Forest have been applied in identifying railway anomalies with computer vision. Previous studies have successfully demonstrated using a pre-train SVM to identify the difference between a sample tie plate and a referenced tie plate in the sliding window from machine vision technology. Also, the images can be collected from unmanned aerial vehicles (UAV) and random forest classification can be conducted to detect rail crack. In this work, the images are noisy and require denoising.

[0008] Whilst denoising method using wavelet transform and median filtering (WTCMF) can be used in image pre-processing for rail surface defects detection, a solution is to enhance the edge detection ability and to alleviate the influence of noise. There are two other major obstacles in the visual inspection system for rail surface detection. The images are usually taken under ambient lighting and the defects appear in different shapes and sizes.

[0009] The coarse-to-fine model (CTFM) takes the inputs in three different scales from sub-image, region to pixel level. It filters the defect-free range in the region level, locates the defect region using phase-only Fourier transforms in the sub-image level, and extracts the shape of the defect at the pixel level. The use of a line-scan camera can be combined with Spectral Image Differencing Procedure (SIDP) to identify rail cracks and defects on irregular textures.

[0010] Traditional computer vision methods can detect defects and classify the defect classes of rail track images. For example, morphological operation method involves a method using morphological feature extraction and image processing, combined with the Hough transform, to detect rail defects has been disclosed. Specific defects like head check breakages, apletilik, and undulations can be identified. Another disclosed technology is a real-time rail track system to use the H-value of color images for swift pre-processing and employs morphological processing to eliminate redundancy. It uses the contour of the direction chain code to identify defect shapes quickly in the system which could be achieved in real-time detection. However, the traditional image processing method produces a high false alarm rate and demands extensive domain knowledge.

[0011] In the Convolutional neural network approach (CNN-based method), YOLOX involves using results in various defects that affect train operation and safety. Current defect detection models struggle with poor illumination and insufficient data. A prior literature proposes an enhanced YOLOX model combined with image processing techniques to highlight and detect

rail defects efficiently. RetinaNet presents a deep learning (DTL) framework for detecting rail surface cracks using a limited number of training images. Utilizing pre-trained YOLOv3 and RetinaNet models from the COCO dataset, the study applied these transferred models to rail track images, then employed an ensemble scheme to enhance detection. When benchmarked against several models and traditional detection methods used a dataset of 102 rail images, the DTL model showcased superior recall and precision. YOLOv3 showed in detecting smaller cracks, while RetinaNet was suitable to detect large cracks.

[0012] One application of neural networks for defects detection is depicted in PRC Invention Patent Application No. 112649513A wherein it provides an image recognition-based artificial intelligence damage judging method for a railway. US Patent Application No. 20230290135A1 may have disclosed a system and technique to generate a robust representation of an image wherein the input tokens of an input image are received, and an inference about the input image is generated based on a vision transformer (ViT) system comprising at least one self-attention module to perform token mixing and a channel self-attention module to perform channel processing. However, neither the image recognition-based artificial intelligence damage judging method nor the self-attention module provided any practical solution to the long existing problem of high false alarm and high computational demand.

[0013] PRC Invention Patent Applications Nos. 114973199A and 115661726A may have disclosed a rail transit image recognition in detecting obstacles based on a convolutional neural network and image and an autonomous video acquisition and analysis method for rail train workpiece assembly respectively. However, both patents lack of effective classification module to accurately classify defects and thus reducing false alarm rate.

[0014] False alarms, or erroneous defect identifications, can lead to various problems. They may trigger unnecessary maintenance activities, causing disruptions to the railway network and incurring additional costs. Moreover, excessive false alarms can strain the human operators' tasks with monitoring the system, leading to reduced trust in the system's alerts and potentially causing legitimate defects to be overlooked. Consequently, there is a pressing need for innovative systems and methods that can effectively address the issue of false alarms in rail track defect detection.

SUMMARY OF THE INVENTION

[0015] It is an objective of the present invention to provide a system and method to reduce the false alarm rate and enhance the defect detection rate in the existing railway track inspection technologies.

[0016] It is also an objective of the present invention provides a system and method that allow cross-domain knowledge transfer for Transformer models with more extensive datasets.

[0017] Accordingly, these objectives can be achieved by following the teachings of the present invention, which relate to a computer-implementable system and method for detecting and classifying defects in railway tracks images using Artificial Intelligence (AI). The system comprises a receiving module configured to receive image inputs, a pre-processing module configured to pre-process the image inputs, a two-stage Transformer model configured to detect and classify defects and false alarms based on the image inputs, and an output module to display results containing images with defects, false alarms or unknown.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The features of the invention will be more readily understood and appreciated from the following detailed description when read in conjunction with the accompanying drawings of the preferred embodiment of the present invention, in which:

[0019] **Fig. 1** illustrates an overview of the flow of the present invention;

[0020] **Fig. 2** illustrates four examples of defect classification in the present invention; and,

[0021] **Fig. 3** illustrates a flowchart of the two-stage Transformer model in the present invention with confidence threshold selection.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

[0022] For the purposes of promoting and understanding the principles of the invention, reference will now be made to the embodiments illustrated in the drawings and described in the following written specification. It is understood that the present invention includes any alterations and modifications to the illustrated embodiments and includes further applications of the principles of the invention as would normally occur to one skilled in the art to which the invention pertains.

[0023] Traditional image processing techniques often suffer from high false alarm rates due to their sensitivity to variabilities and their lack of adaptability. First, they are sensitive to various factors like lighting conditions, shadows, noise, and environmental changes. These variabilities can significantly impact the accuracy and reliability of image processing results, making it challenging to maintain a low false alarm rate. Second, traditional methods lack adaptability. Unlike deep learning methods, which can be trained on extensive datasets to learn and adapt to specific scenarios, traditional techniques remain static. They cannot evolve or improve their performance when faced with new data or novel defect types without manual intervention, which limits their practicality in dynamic real-world applications.

[0024] To address these shortcomings, modern computer vision approaches, particularly deep learning-based methods, have gained popularity. These methods can automatically learn and adapt from new data, making them more robust in handling variabilities and reducing false alarms. By using neural networks and large-scale datasets, they can extract relevant features and patterns from images, allowing them to be more versatile and adaptive to changing conditions. Additionally, deep learning methods can continually improve their performance over time, making them a valuable tool in various image processing applications, including object recognition, defect detection, and image analysis.

[0025] More particularly, the present invention teaches a computer-implementable system for detecting defects in railway tracks images using Artificial Intelligence (AI), the system comprising: a receiving module configured to receive image inputs; a pre-processing module configured to pre-process the image inputs; a two-stage Transformer model configured to detect defects or false alarms based on the image inputs; and an output module to display results containing images with defects, false alarms or unknown.

[0026] Due to the recent advancements in computer vision, it has been suggested that transformer models are often utilised due to its self-attention mechanism. It offers superior feature extraction capabilities compared to traditional convolutional neural networks, especially in defect detection tasks. Among the variations of transformer models, the Swin Transformer stands out. A Swin Transformer diverges from standard transformer structures by incorporating a self-attention window. This addition not only reduces computational demands but also boosts model performance through its hierarchical architecture and shift window attention mechanism. Leveraging its capabilities to effectively extract features and accurately detect defect rail track images, Swin Transformers have been employed in the present invention due to its flexible backbone and architecture.

[0027] In an embodiment of the present invention, the two-stage Transformer model is employed, wherein the two-stage Transformer model is at least two Swin Transformer models, wherein the models further comprising a first Swin Transformer model and a second Swin Transformer model arranged in a sequence. These cascaded Swin Transformers have optimal confidence selection, wherein the cascaded Swin Transformer design is improved to integrate two sequential Swin Transformers.

[0028] In another embodiment of the present invention, the two-stage Transformer model has a pre-determined threshold of confidence with a value as 0.7.

[0029] Another embodiment of the present invention is that the first Swin Transformer model uses a pre-trained model on ImageNet.

[0030] Another embodiment of the present invention, wherein the second Swin Transformer model has a classification head, modified from the first Swin Transformer model to output at least one class of defect. The classification head is modified to include 12 classes of defects of which are further disclosed below.

[0031] Another embodiment of the present invention, wherein the at least one class of defect including but not limited to error images, grinding marks, joints, mark, squat, weld, word, head check error, missing fastener, rail head anomaly, corrugation or others.

[0032] Further to the above, the first Swin Transformer model is a binary classifier and the

second Swin Transformer model is a multiclass classifier. For simplicity, the Swin Transformer model's classifier head is modified to incorporate the 12 classes of defects. Generally, the binary classifier is used to predict the input image. If the prediction confidence is larger than 0.7, the multiple class classifier is used to predict the specific class because these images are not hard sample for model. If the prediction confidence is not larger than 0.7, the binary classifier is further used to predict the binary class label for these hard sample. The confidence in 0.5 to 0.9 is opted, and amongst the selected confidence, 0.7 is the best threshold based on false alarm recall and precision. The specific classes of images from the second Swin Transformer model / the multiclass classifier to determine defect or normal class are shown in **Fig. 2**.

[0033] Another embodiment of the present invention, wherein the system is implementable into existing inspection systems including but not limited to On-board Rail Inspection System (ORIS).

[0034] Another embodiment of the present invention, wherein the image inputs are obtained from image files, a database having a plurality of false alarm images or images from a sensor camera.

[0035] More specifically, the two-stage Swin Transformers in the present invention are tasked with different tasks. For example, in the first stage, the first Swin Transformer model is tasked with predicting both false alarms and defect rail track images. To refine these predictions, the optimal thresholding selection is applied to filter out low confidence of false alarms and defect rail track images. The specific defect classes are detected in the second Swin Transformer model. The input rail track images at the second Swin Transformer model input are the high confidence score of false alarm images selected by the first Swin Transformer model output. This method achieves a high defect detection rate and reduce the false alarm rate from the output of the inspection system.

[0036] The present invention teaches a method **100** for detecting and classifying defects in railway tracks images using Artificial Intelligence (AI), the method comprising: inputting one or more images with defects in railway tracks **102**; pre-processing the images **104**; analysing the images for defects detection or false alarms using a two-stage Transformer model **106**; and outputting results **108**. An illustration of the method is further depicted in **Fig. 1**.

[0037] An embodiment of the present invention, wherein the step of inputting one or more images with defects in railway tracks **102** further comprising selecting images with defects from a database, image files, or images from a sensor camera.

[0038] Another embodiment of the present invention is the step of pre-processing the images **104** further comprising denoising the images with defects in railway tracks.

[0039] Another embodiment of the present invention, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model **106** further comprising: extracting feature maps from the inputted images using a first Transformer model; predicting a defect or a false alarm in each feature map and each image to a classifying head of the first Transformer model; scoring each predicted image; and filtering high-scored images with confidence scores exceeding a pre-determined threshold from low-scored images with confidence scores below the pre-determined threshold. This is further illustrated in **Fig. 3** along with confidence threshold selection.

[0040] Another embodiment of the present invention, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model **106** further comprising: passing the high-scored images from the first Transformer model to a second Transformer model; extracting feature maps from the high-scored images; applying the mathematical function to a classifying head of the second Transformer model to classify a specific defect class; and labelling the image with the specific defect class.

[0041] Another embodiment of the present invention, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model **106** further comprising: passing the low-scored images from the first Transformer model to the second Transformer model; extracting feature maps from the low-scored images; applying the mathematical function to the classifying head of the second Transformer model to classify the images; and labelling the image as "unknown" or "false alarm" based on the confidence score.

[0042] Another embodiment of the present invention, wherein the step of passing the low-scored images from the first Transformer model to the second Transformer model further comprising generating an improved false alarm recall rate.

[0043] Another embodiment of the present invention, wherein the step of labelling the image with the specific defect class further comprising selecting the specific defect class by the classifying head from a group of pre-determined defect classes.

[0044] Another embodiment of the present invention, wherein the step of filtering high-scored images with confidence scores exceeding the pre-determined threshold from low-scored images with confidence scores below the pre-determined threshold further comprising: selecting an optimum confidence score as the pre-determined threshold; initiating an optimal thresholding selection with the pre-determined threshold selected; and segregating the images with confidence scores exceeding the pre-determined threshold, wherein, images with confidence scores that exceed the pre-determined threshold are considered high-scored images and vice versa.

[0045] Another embodiment of the present invention, wherein the step of outputting the results **108** further comprising: determining whether the images consisting of defects, false alarms or unknown; and displaying the results based on the analysis by the two-stage Transformer model.

[0046] Another embodiment of the present invention, wherein the step of determining whether the images consisting of defects, false alarms or unknown is determined by the classification head of the second Transformer model.

[0047] Another embodiment of the present invention, wherein the step of selecting the optimum confidence score as the pre-determined threshold is selected from a range of 0.5 to 0.9 with a step of 0.05. It is a thresholding selection technique that is applied after training the first Swin Transformer model. The high confidence of predicted results and the low confidence of predicted results are split by the optimal confidence threshold.

[0048] Another embodiment of the present invention, wherein the step of predicting the defect or the false alarm in each feature map and each image by applying the mathematical function to the classifying head of the first Transformer model further comprising training the first Transformer model using training images containing defects and false alarm.

[0049] The present invention provides various advantages and more specifically, the present

invention can reduce the false alarm rate and enhance the defect detection rate in the existing railway track inspection system. The Swin Transformers employed in the present invention has a self-attention mechanism and a hierarchical architecture that enable more accurate detection of defect railway track images, thereby reducing the false alarm rate. Furthermore, the cascaded structure of the first Swin Transformer with optimal thresholding further refines predictions whilst the second Swin Transformer targets specific defect classes as defect classification.

[0050] Another advantage of the present invention is that the present invention employs non-destructive technology to detect defects on the railway tracks. Such technology provides advantages over invasive technology as it is generally less expensive compared to destructive methods, as it does not require extensive excavation or replacement of the rail tracks. The present invention also quicker to implement, as it does not require the removal or replacement of the rail track. This allows for faster detection of defects and reduces downtime for repairs.

[0051] The present invention also causes minimal disruption to train services, as it does not require the closure or diversion of rail tracks. This ensures smoother operations and prevents inconvenience to commuters. It is also relatively safer for workers, as it eliminates the need for physical contact with the rail tracks. This reduces the risk of accidents or injuries during the defect detection process.

[0052] Additionally, the present invention allows continuous monitoring of the rail track, enabling the detection of defects in real-time or at regular intervals. This proactive approach helps prevent major failures or accidents and ensures timely maintenance and repair. It also preserves the integrity of the rail track, as it does not require removal or replacement of any components. This helps extend the lifespan of the tracks and minimizes the need for large-scale repairs or replacements.

[0053] Moreover, the present invention is also more environmentally friendly, as it does not generate excess waste or require the disposal of materials from the rail track. This reduces the carbon footprint associated with defect detection and maintenance activities.

[0054] Table 1 shows the first Swin-Transformer model results. It presents results on false alarm recall and precision rates at various confidence levels. The column labelled "Detected num" signifies the number of predictions with a probability exceeding the confidence level.

The "True Negative" column corresponds to the false alarm category, while the "True Positive" column pertains to the defect category.

Table 1

Confidence	Detected num	True Negative	True Positive	False Negative	False Positive	False alarm recall	Precision
0.5	940	784	98	23	35	0.950	0.971
0.6	871	764	76	13	18	0.926	0.983
0.7	784	730	47	3	4	0.884	0.995
0.8	693	676	16	1	0	0.819	0.998
0.9	256	255	1	0	0	0.309	1.0

[0055] **Table 2** illustrates the model comparison with other popular models, such as vision Transformer, ResNet152 and Efficientnet. Comparison metric is top1 accuracy.

Table 2

	Accuracy	Parameters' number	Flops
EfficientNet-b8_8xb32	0.862	87M	7G
VIT-BASE-p32_64xb64	0.863	89M	4.3G
ResNet152	0.886	60M	11.5G
The present invention	0.933	87M	15G

[0056] **Table 3** illustrates the predicted results selected from the low-confidence threshold.

Table 3

Confidence	Detected num	True Negative	True Positive	False Negative	False Positive	Correct
0.5	117	53	32	15	17	81 (0.692)
0.6	82	41	18	12	11	57 (0.695)
0.7	47	30	7	7	3	37 (0.787)
0.8	15	9	3	3	0	12 (0.8)
0.9	2	2	0	0	0	2 (1.0)

[0057] **Table 4** illustrates the predicted results selected from the high-confidence threshold. It is the result of multiple class classifier for high confidence prediction sample from binary classifier. Accuracy means the ratio of model predicting the correct multiple class label.

Table 4

Confidence	Detected num	True Negative	True Positive	False Negative	False Positive	Correct
0.5	767	729	35	2	1	761 (0.992)
0.6	757	725	29	2	1	751 (0.992)
0.7	744	716	26	1	1	740 (0.994)
0.8	667	649	16	1	1	665 (0.997)
0.9	170	166	4	0	0	170 (1.0)

[0058] **Table 5** illustrates cascaded model results with different confidence levels. It is the result of combining these two Transformers. The results show that the cascaded architecture has improved the false alarm recall by 3.7% with 0.7% precision loss.

Table 5

Confidence	Detected num	True Negative	True Positive	False Negative	False Positive	False alarm recall	Precision
0.5	901	783	79	18	21	0.949	0.977
0.6	866	771	65	15	15	0.934	0.980
0.7	831	760	54	10	7	0.921	0.987
0.8	799	739	50	6	4	0.895	0.991
0.9	786	732	47	3	4	0.887	0.995

[0059] Based on the models presented in **Table 2**, columns for Architecture, Adaptability, Hierarchical Representation, Computational Efficiency, and Data Augmentation Dependency are further tabulated in **Table 6**. These dimensions have been chosen for comparative analysis because they offer crucial insights into the models' design, functionality, and performance.

Table 6

Model	Architecture	Hierarchical Representation	Computational Efficiency	Data Augmentation Dependency	Adaptability
Swin Transformer	Transformer-based	Yes	High	Low	High
EfficientNet	CNN-based	Yes	Optimized	Moderate to High	Moderate
VIT	Transformer-based	No	Moderate	High	Moderate
ResNet-152	CNN-based	Yes	Moderate	High	Moderate

[0060] The present invention explained above is not limited to the aforementioned embodiment and drawings, and it will be obvious to those having an ordinary skill in the art of

the prevent invention that various replacements, deformations, and changes may be made without departing from the scope of the invention.

CLAIMS

WHAT IS CLAIMED:

1. A computer-implementable system for detecting defects in railway tracks images using Artificial Intelligence (AI), the system comprising:

   a receiving module configured to receive image inputs;

   a pre-processing module configured to pre-process the image inputs;

   a two-stage Transformer model configured to detect defects and false alarms based on the image inputs; and

   an output module to display results containing images with defects, false alarms or unknown.

2. The system according to claim 1, wherein the two-stage Transformer model is at least two Swin Transformer models, wherein the models further comprising a first Swin Transformer model and a second Swin Transformer model arranged in a sequence.

3. The system according to claim 1, wherein the two-stage Transformer model has a pre-determined threshold of confidence with a value as 0.7.

4. The system according to claim 2, wherein the first Swin Transformer model uses a pre-trained model on ImageNet.

5. The system according to claim 2, wherein the second Swin Transformer model has a classification head, modified from the first Swin Transformer model to output at least one class of defect.

6. The system according to claim 5, wherein the at least one class of defect including but not limited to error images, grinding marks, joints, mark, squat, weld, word, head check error, missing fastener, rail head anomaly, corrugation or others.

7. The system according to claim 1, wherein the system is implementable into existing inspection systems including but not limited to On-board Rail Inspection System (ORIS).

8. The system according to claim 1, wherein the image inputs are obtained from image files, a database having a plurality of false alarm images or images from a sensor camera.

9. A method (100) for detecting and classifying defects in railway tracks images using Artificial Intelligence (AI), the method comprising:

inputting one or more images with defects in railway tracks (102);

pre-processing the images (104);

analysing the images for defects detection or false alarms using a two-stage Transformer model (106); and

outputting results (108).

10. The method according to claim 9, wherein the step of inputting one or more images with defects in railway tracks (102) further comprising selecting images with defects from a database, image files, or images from a sensor camera.

11. The method according to claim 9, wherein the step of pre-processing the images (104) further comprising denoising the images with defects in railway tracks.

12. The method according to claim 9, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model (106) further comprising:

extracting feature maps from the inputted images using a first Transformer model;

predicting a defect or a false alarm in each feature map and each image by applying a mathematical function to a classifying head of the first Transformer model;

scoring each predicted image; and

filtering high-scored images with confidence scores exceeding a pre-determined threshold from low-scored images with confidence scores below the pre-determined threshold.

13. The method according to claim 12, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model (106) further comprising:

passing the high-scored images from the first Transformer model to a second Transformer model;

extracting feature maps from the high-scored images;

applying the mathematical function to a classifying head of the second Transformer model to classify a specific defect class; and

labelling the image with the specific defect class.

14. The method according to claim 12, wherein the step of analysing the images for defects detection or false alarms using the two-stage Transformer model (106) further comprising:

passing the low-scored images from the first Transformer model to the second Transformer model;

extracting feature maps from the low-scored images;

applying the mathematical function to the classifying head of the second Transformer model to classify the images; and

labelling the image as "unknown" or "false alarm" based on the confidence score.

15. The method according to claim 14, wherein the step of passing the low-scored images from the first Transformer model to the second Transformer model further comprising generating an improved false alarm recall rate.

16. The method according to claim 13, wherein the step of labelling the image with the specific defect class further comprising selecting the specific defect class by the classifying head from a group of pre-determined defect classes.

17. The method according to claim 12, wherein the step of filtering high-scored images with confidence scores exceeding the pre-determined threshold from low-scored images with confidence scores below the pre-determined threshold further comprising:

selecting an optimum confidence score as the pre-determined threshold;

initiating an optimal thresholding selection with the pre-determined threshold selected; and

segregating the images with confidence scores exceeding the pre-determined threshold, wherein, images with confidence scores that exceed the pre-determined threshold are considered high-scored images and vice versa.

18. The method according to claim 9, wherein the step of outputting the results (108) further comprising:

determining whether the images consisting of defects, false alarms or unknown; and

displaying the results based on the analysis by the two-stage Transformer model.

19. The method according to claim 18, wherein the step of determining whether the images consisting of defects, false alarms or unknown is determined by the classification head of the second Transformer model.

20. The method according to claim 17, wherein the step of selecting the optimum confidence score as the pre-determined threshold is selected from a range of 0.5 to 0.9 with a step of 0.05.

21. The method according to claim 12, wherein the step of predicting the defect or the false alarm in each feature map and each image by applying the mathematical function to the classifying head of the first Transformer model further comprising training the first Transformer model using training images containing defects and false alarm.

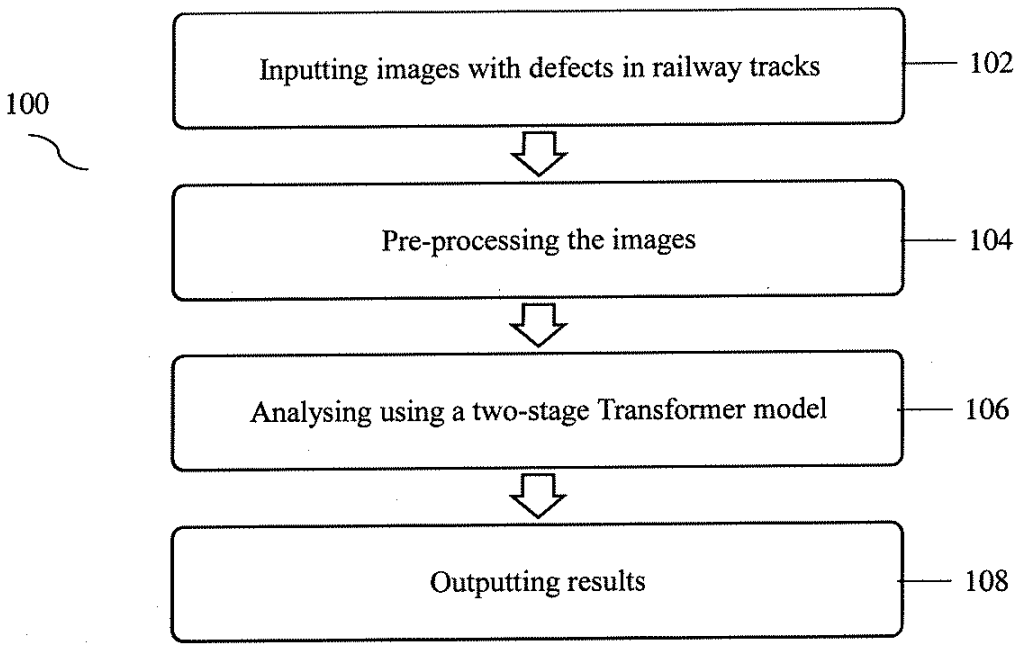
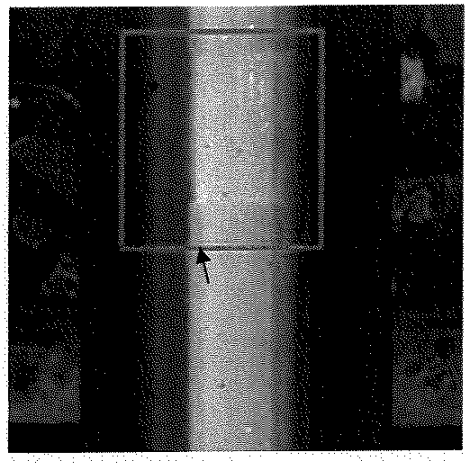
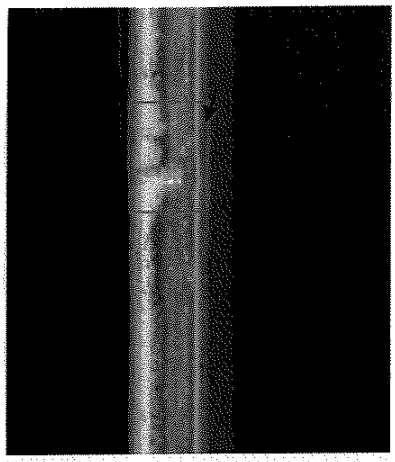

100
Inputting images with defects in railway tracks
102
Pre-processing the images
104
Analysing using a two-stage Transformer model
106
Outputting results
108

Fig. 1



Welding



Squat

Shell

Head Checking

Fig. 2

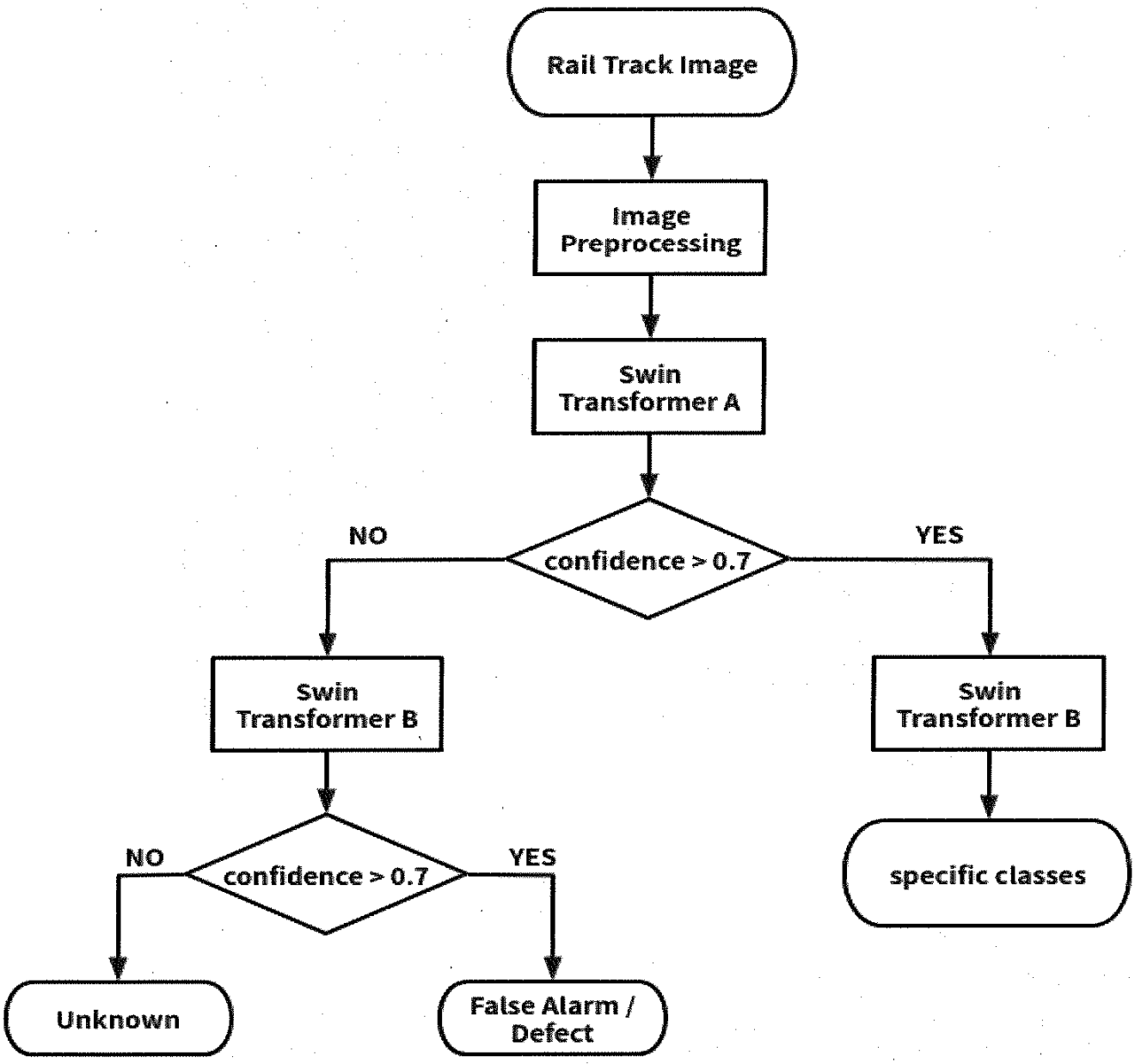

Rail Track Image
Image Preprocessing
Swin Transformer A
confidence > 0.7
NO
YES
Swin Transformer B
Swin Transformer B
confidence > 0.7
NO
YES
specific classes
Unknown
False Alarm / Defect

Fig. 3