This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3502438

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10

Re-DETR: Research on Fast Detection Technology for Railway Engineering Targets in the Dark Time Domain

ZHAOHUI ZHENG[1], TINGFENG JI[1], JIANPING JU[2], GUAN QING[1], SHUILONG ZOU[4,1], Qiang Zhang[3], Qing Zhou[4], and Yingjian He[1]

[1]School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology, Nanchang 330108 China,(e-mail: zhengzhaohui@sina.cn(Z. Z.), jtf544562426@gmail.com(T. J.), 752782167@qq.com(S. Z.), 309855851@qq.com(G. Q.), 926114924@qq.com(Y. H.)
[2]School of Artificial Intelligence, Hubei Business College, Wuhan 430070, China(e-mail:gjdxjjp@whu.edu.cn(J. J.))
[3]Wuhan Railway Vocational College of Technology, Wuhan 430200, China
[4]Nanchang Vocational College of Applied Technology, Nanchang 330108 China

Corresponding author: Tingfeng.Ji(e-mail: 544562426@qq.com), Jianping Ju(e-mail: gjdxjjp@whu.edu.cn).

ABSTRACT Fast detection of railway engineering targets under low light conditions has always been a challenging problem. Traditional target detection algorithms are limited by lighting conditions, resulting in a decrease in target visibility, which in turn affects detection accuracy. To address this issue, this study proposes a new target detection network for low-light environments (Re-DETR) that enhances the model's detection capability for targets under low light conditions by integrating an optimized RetinexNet image enhancement network and an improved transformer for the image recognition strategy (DETR). Re-DETR uses RetinexNet for image enhancement to improve image quality and visibility and then inputs the enhanced images into the DETR algorithm with an added global channel attention (GCA) module for target detection. The experimental results show that our method can quickly and accurately detect railway engineering targets in the dark time domain, which has significant advantages over traditional methods.

INDEX TERMS DETR, Retinexnet, Global Channel Attention (GCA), target detection, dark time domain, railway engineering

I. INTRODUCTION

In today's society, railway transportation, as an important means of traffic, plays a crucial role in connecting cities and promoting economic development. However, with the continuous expansion of railway networks and increasing demand for transportation, railway safety issues are becoming increasingly prominent. Among them, fast and accurate detection of railway engineering targets is crucial to ensuring railway transportation safety. However, in the dark time domain environment, insufficient lighting conditions limit the visibility of railway engineering targets, posing great challenges to target detection. In this case, the low accuracy of target detection affects the efficiency of railway engineering, leading to transportation delays and ultimately affecting railway safety. The traditional target detection methods based on manual features and deep learning have limitations in dealing with railway engineering target detection in dark time domain environments, making it difficult to effectively solve the current problem. Therefore, finding new solutions to improve the accuracy of railway engineering target detection is an urgent task.

To address this issue, this study proposes an RE-DETR network aimed at achieving fast detection of railway engineering targets in a dark time domain environment. The traditional DETR [1] achieves end-to-end target detection through a self-attention mechanism, which can effectively extract target features and perform correlation. However, its performance is relatively poor for railway engineering detection in the dark time domain. Our proposed RE-DETR network further combines DETR with RetinexNet Enhance-Net to improve the visibility of targets under insufficient lighting conditions, thereby improving the accuracy and robustness of target detection. In addition, the transformer is a nondiscriminative spatial attention mechanism, which means that while it weights target features, it also gives weight to noise features. Under low-light conditions, there is often considerable background noise, which can severely affect the detection performance of the model. Therefore, by using global channel attention (GCA) to increase the value of target features and decrease the value of noise features, the model can effectively distinguish between target and noise features, alleviating the problem of noise interference in the detection of buildings in high-resolution remote sensing imagery. Inspired by this, this paper introduces a global channel attention (GCA) module, which calculates the correlation between all channels and assigns different weights to each channel, enhancing the ability of the network to learn target features and reducing the interference of noise features.

The main contributions of this study include the following:

(1) For the dark time-domain conditions in railway maintenance, an improved image enhancement network, RetinexNet [2], was selected to process low-light images. By introducing RetinexNet for image enhancement, the improved data have better contrast and brightness when training the target detection model, thereby improving the accuracy and stability of target detection.

(2) By combining the enhanced images with an improved DETR strategy for low-light target detection, the original DETR model uses the GCA algorithm and a comprehensive loss function to increase its adaptability and accuracy in target detection tasks. Our proposed Re-DETR has achieved significant effects in low-light target detection tasks, improving detection accuracy and generalizability and providing an effective solution for target detection problems in complex scenarios.

II. RELATED WORK

With the continuous development of target detection technology, it has been applied in more and more scenarios and fields [3], among which target detection for railway engineering in dark time domain is a challenging scenario. In this scenario, the lighting conditions are poor, and the target object may be occluded or blurred, which brings additional difficulties and complexity to target detection.

In recent years, significant progress has been made in both target detection and image enhancement technologies. The target detection algorithm is continuously optimized and improved, evolving from traditional region-based methods to end-to-end detection models based on deep learning, such as Faster R-CNN, YOLO, SSD, etc. These models have achieved significant improvements in accuracy and speed, making target detection more widely used in various application scenarios. On the other hand, image enhancement technology is also constantly evolving. By enhancing the quality, contrast, clarity, and other aspects of the image, image enhancement technology can improve the visual effect of the image, which helps to improve the performance and accuracy of target detection algorithms.Image enhancement technology plays an important role in many fields, providing better input data for machine learning algorithms.

In terms of image enhancement, traditional image enhancement methods include techniques such as histogram equalization and filters. Mayathevar et al. [4] proposed the histogram equalization method was introduced for image enhancement. This histogram equalization can enhance the contrast and brightness of the image, but it can easily lead to excessive contrast enhancement, which may result in excessive sharpening and make the image look unnatural. Manjon et al. [5] proposed a Non-local Means Filter for image denoising. This method reduces noise by searching for similar blocks in the image and calculating the weighted average of these blocks, thereby preserving image details. However, its computational complexity is high, parameter selection is difficult, and edge blurring is handled. In contrast, image enhancement methods based on deep learning have more advantages than traditional methods. For example, Wei et al. [6] proposed a low light enhancement method based on deep Retinex decomposition. This method can effectively improve image brightness and contrast under low light conditions, while preserving detailed information. A multi-scale Retinex method was proposed in reference [7] by Huang et al, which is suitable for image enhancement at different scales. This method can better preserve image details and textures, and improve image quality when processing images of different scales. Lv et al. [8] introduced a deep dual Retinex network for low light image enhancement. This network not only enhances the brightness of low light images, but also effectively enhances the details and clarity of the images. Lee et al. [9] improved the deep Retinex decomposition method to enhance the quality of low light images. This improved method can produce clearer and more contrasting images under low light conditions. Zhang et al. [10] proposed an adaptive multi-scale Retinex method for image enhancement. This method can automatically select the appropriate scale based on the features of the image, improve the contrast and color balance of the image, and is particularly suitable for processing images under complex lighting conditions, thereby effectively improving the quality of the image. Yang et al. [11] proposed an image enhancement method based on Retinex, which uses adaptive gamma correction. This method can effectively enhance the brightness and contrast of the image under low light conditions while preserving the clarity of detailed information. Cai et al. [12] proposed a Retinexformer model based on ORF and IGT.

In target detection, traditional methods include Haar feature cascaded classifier and HOG+SVM. Zhu et al. [13] used Haar feature and cascaded AdaBoost classifier for target detection, which performs well in face detection but is sensitive to complex backgrounds and lighting changes, not suitable for complex scenarios, and requires manual feature design, which is not flexible enough. Llorca et al. [14] proposed a method for intelligent detection and classification of infrared images based on HOG features and SVM classifiers. However, it is sensitive to changes in target scale and pose, making it difficult to cope with occlusion and complex backgrounds, requiring manual adjustment of parameters and feature extraction methods. Therefore, we plan to use target detection networks from recent years. In recent years, many networks have been used, including the improved Fast R-CNN proposed by Maity et al. [15], which achieves more efficient target detection with targeted and diverse features. Bharati et al. [16] combined the Mask R-CNN model of target detection and fusion of different visual features, which can better predict the bounding boxes of targets. Wang et al. [17] proposed enhancing the performance of YOLOv2 by adjusting the detection layer of a single machine network, which can better perform real-time detection of key railway parts. The Single Shot MultiBox Detector (SSD) proposed in Kumar et al. [18] achieves fast target detection by predicting bounding boxes and categories at multiple scales. The Ghost RetinaNet proposed in reference [19] has good bounding box regression and localization accuracy, and can achieve fast detection. Chen et al. [20] proposed an improved Cascade-RCNN network that improves target detection accuracy by simultaneously introducing multi-scale training. The Transformer-based DETR model mentioned by Zhu et al. [21] achieves end-to-end target detection, using attention mechanisms to achieve target detection and category prediction. It has good plasticity and accuracy in target detection. Du et al. [22] mainly delves into state-of-the-art artificial intelligence (AI) technologies, with a special focus on pipeline parallelism, data parallelism, and multimodal learning.

Based on the above research results, we have chosen RetinexNet enhance-net and Transformer-based DETR model to design a RE-DETR method with high accuracy and recognition speed to meet the needs of railway operation and maintenance.

III. PROPOSED METHODS

The operation and maintenance of railways is characterized by dark time domain and a large number of tools, resulting in large differences in brightness and multiple overlapping forms in the railway tool dataset, making target detection difficult. To address this issue, this paper proposes a framework for railway tool recognition in the dark time domain. Firstly, the RetinexNet brightness enhance-net is used to enhance images under various lighting conditions. Subsequently, the enhanced image is input into the improved DETR target detection network to obtain accurate detection results. As is shown in Figure 1, this framework combines brightness enhancement and target detection techniques, aiming to effectively address the challenges of railway tool datasets, improve recognition accuracy and robustness.

A. IMAGE ENHANCEMENT MODULE

The maintenance and upkeep of railway equipment in the field usually requires constant attention, resulting in significant differences in brightness of the collected tool images. In response to this challenge, this article proposes an improved image enhancement framework based on RetinexNet for dark time domain environment. According to Retinex theory - human color perception modeling, an image S is considered as the product of the illumination component I and the reflection component R. The equation as follows:

$$S = I \circ R \tag{1}$$

The reflection component R is a constant part determined by the inherent properties of the target, while the illumination component I is the part affected by external lighting, $\circ$ represents multiplication. The purpose of image enhancement can be achieved by removing the influence of lighting or correcting the illumination component I.

The improved framework can be divided into four steps:

(1) Brightness judgment

Faced with the impact of brightness differences on enhancement, we classify images into three categories: dark images, medium brightness images, and bright images. By using statistical measures for threshold discrimination, the brightness type of the image is determined, and different strategies are used to enhance the image using the improved RetinexNet. Specifically, brightness adjustment and contrast enhancement are applied to dark images, color adjustment and noise removal are applied to medium brightness images, and exposure control and detail protection are applied to bright images. This method can better adapt to the brightness differences of railway tool images in the dark time domain, thereby improving the accuracy and robustness of target detection.

The brightness judgment formula is as follows:

$$T = (m_t - T_t)/T_t \tag{2}$$

where m_t is the average brightness of the image at time t, and T_t is the global average brightness of the expected normal image at time t. If $T < \tau_{t_1}$, judged as a dark image; If $T > \tau t_2$, judged as a bright image; If $\tau t_1 < T < \tau t_2$, it is determined as a medium brightness image. Among them, τt is the threshold used to determine image brightness. This article determines T_t and the threshold τt_1 and τt_2 through experiments. The most suitable values are 0.8, 0.5, and 0.2, respectively.Below are examples of three scenarios Figure 2.

(2) Layer Separation

In our railway equipment image processing framework, the image first passes through a model called Decom-Net, which consists of 5 convolutional layers with ReLU (excluding the first and last layers). This model takes low/normal lighting image pairs as inputs and shares network parameters to obtain the reflection component R_{low} and lighting component I_{low} of low lighting images, as well as the reflection component R_{normal} and lighting component I_{normal} of normal lighting images. To optimize this model, we utilized the constraint relationship between these four components and incorporated this constraint relationship into the objective function. Specifically, the loss function of the model consists of three parts: reconstruction loss, reflection component consistency loss, and lighting component smoothing loss:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_{ir}\mathcal{L}_{ir} + \lambda_{is}\mathcal{L}_{is} \tag{3}$$

Reconstructing losses yields:

$$\mathcal{L}_{recon} = \sum_{i=low,normal} \sum_{j=low,normal} \lambda_{ij} \left\| R_i \circ I_j - S_j \right\|_1 \tag{4}$$

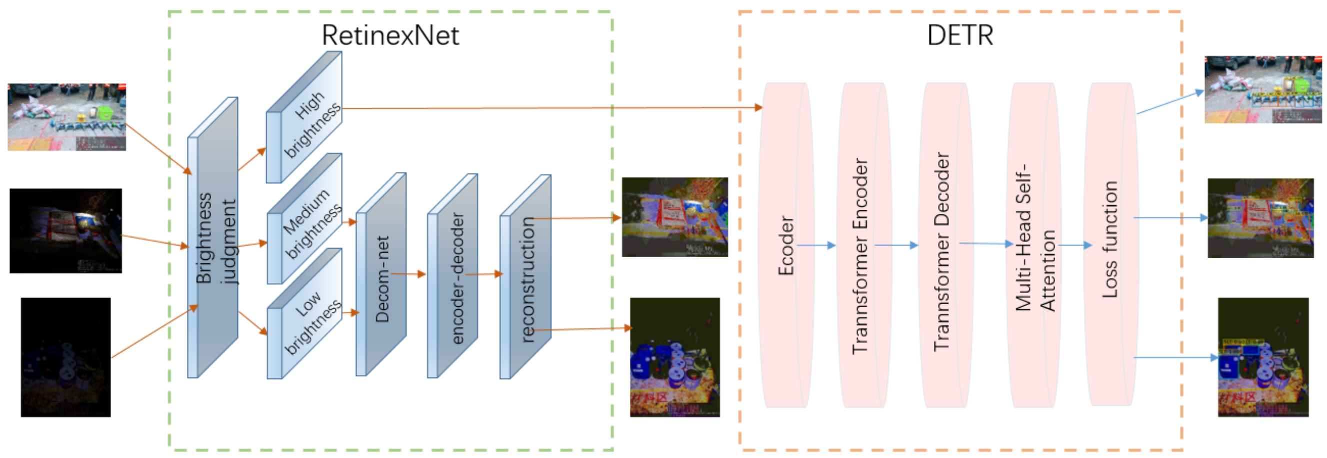


FIGURE 1: RE-DETR framework

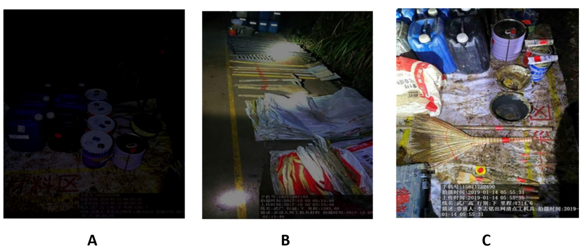


FIGURE 2: a is the tool image for dark images, b is the tool image for medium brightness images, and c is the tool image for bright images.

The main object of this item is to ensure that the reflection component R and lighting component I obtained from model decomposition can accurately reconstruct the details and features of the original image as much as possible, thereby improving the overall image reconstruction quality and fidelity.

The consistency loss of reflection components is expressed as:

$$\mathcal{L}_{ir} = \|R_{low} - R_{normal}\|_1 \tag{5}$$

According to Retinex image decomposition theory, the reflection component R is independent of lighting, so for paired low/normal lighting images, their reflection component R should be kept as consistent as possible.

The smoothing loss of lighting components is expressed as:

$$\mathcal{L}_{ir} = \sum_{i=low,normal} \|\nabla I_i \circ exp(-\lambda_g \nabla R_i)\| \tag{6}$$

In the RetinexNet paper, a hypothesis about the lighting component I was proposed, which is that the ideal lighting component should maintain smoothness in texture details and effectively preserve the overall structure. The implementation of this assumption is achieved by processing the gradient of the reflection component R, allocating the information of its gradient map to the illumination component I, to ensure that the smooth areas in the reflection component R correspond to the same smoothness in the illumination component I. The design of this loss function enables the model to better understand the texture details and overall structure of images during the learning process, thereby improving the quality of image decomposition and reconstruction.

(3) Adjusting the model

Regarding the adjustment of R_{low}: BM3D algorithm is used to suppress the amplified noise in R_{low}, and lighting related strategies are introduced to further optimize the quality of R_{low}.

Regarding the adjustment of I_{low}: Adopting the multi-scale lighting adjustment network of Enhance-Net, its overall structure is an encoder-decoder architecture, and multi-scale connections are introduced. This design enables the network to capture a wide range of lighting distribution contextual information, which helps improve its adaptive adjustment ability.

(4) Reconstruction

The final enhanced image can be obtained by multiplying the adjusted R_{low} and I_{low}.

Due to the possibility of overexposure for bright images during image enhancement, which may affect subsequent target detection, we only used RetinexNet brightness enhancement neural network for brightness enhancement and denoising for dark and medium brightness images, without performing bright image enhancement processing. Our processing flow includes three key steps: brightness separation, logarithmic transformation, and color restoration.

1. In the step of brightness separation, the image is decomposed into two parts: brightness and color, so that the brightness information and color information can be processed separately.

2.In the step of logarithmic transformation, the brightness image is processed to enhance the contrast of the image, making the image details clearer and more prominent.

3.In the step of color restoration, the processed brightness image and color image are resynthesized to generate the final enhanced image, making the overall effect of the image more vivid and natural.

Through the above processing flow, image quality and visual effects can be effectively improved.The renderings are shown in Figure 3.

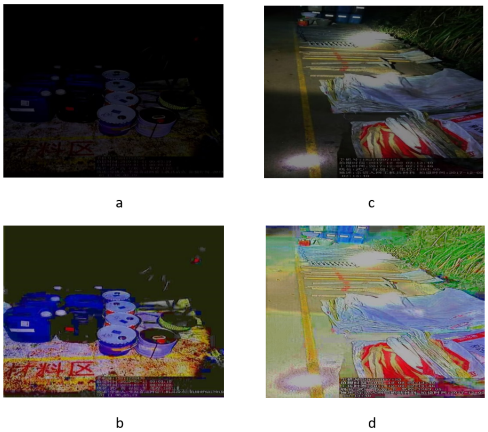


FIGURE 3: a is the tool image of the dark image, b is the result of enhanced dark image, c is the tool image of the medium brightness image, and d is the result of enhanced medium brightness image.

B. TARGET DETECTION MODULE

To facilitate rapid detection of railway maintenance targets, this paper has appropriately modified the DETR model to increase its performance and efficiency in detecting railway maintenance under dark time-domain conditions. The DETR model is mainly composed of four parts: the CNN backbone, the transformer's encoder, the transformer's decoder, and the prediction layer feed-forward network (FFN). To enhance the adaptability and accuracy of the original DETR model for target detection tasks, we improved it using the GCA algorithm and a comprehensive loss function. The self-attention mechanism A of the DETR model usually adopts a fixed fully connected weight matrix, as follows:

$$A = XW_q(W_k)^T \tag{7}$$

where X is the input feature matrix and W_q and W_k are the weight matrices of the query and key, respectively. However, this fixed weight matrix may not adapt well to the relationships and feature representations between different target objects. Therefore, this paper introduces global channel attention (GCA) to enhance the features of building targets and suppress background noise.

Multiscale feature fusion helps to enrich the semantic information in the spatial domain, thereby alleviating the problem of feature information loss. However, in complex scenarios of high-resolution remote sensing images, there is often a significant amount of background noise, which may affect the model's detection performance. To address this issue, this paper introduces a GCA mechanism to enhance the feature representation of building targets and suppress background noise. The design details of the GCA mechanism are shown in Figure 4, and the implementation process is as follows:

1. Input feature map: We denote the input feature map as $X \in \Omega^{H\times W\times C}$,and obtain a feature vector $P \in \Omega^{1\times C}$ through adaptive average pooling operation.

2. Feature flattening and relationship matrix: Flatten the input feature map X and adjust its shape to obtain $Y \in \Omega^{L\times C}, (L = H \times W)$. Then, the feature maps $T \in \Omega^{L\times C}$ and $S \in \Omega^{L\times C}$ are generated through a linear mapping layer, followed by matrix multiplication to obtain the relationship matrix $R \in \Omega^{C\times C}$. Finally, we exchange the last two dimensions of R to obtain R'.

3. Matrix concatenation and dimension adjustment: concatenate the feature vector $P \in \Omega^{1\times C}$, the relationship matrix $R \in \Omega^{C\times C}$, and $R^{'} \in \Omega^{C\times C}$ in the -2 dimension to obtain the matrix $Z \in \Omega^{(2C+1)\times C}$. Next, adjust its dimensions to form a new matrix $Z \in \Omega^{C\times (2C+1)}$.

4. Linear transformation and output generation: Process the obtained Z through a linear layer and apply a sigmoid activation function to generate $Z \in \Omega^{C\times (2C+1)}$. Finally, matrix multiply $Z'$with the input feature map X to output the enhanced feature map $X' \in \Omega^{H\times W\times C}$.

Through these steps, the Global Channel Attention mechanism effectively improves the extraction capability of building features and reduces the adverse impact of complex backgrounds on model performance. For the selection of hyperparameters, we set the number of channels compressed to C'=C/4, the learning rate to $1e^{-3}$, the batch size to 16, and the activation function type to Sigmoid.

C. IMPROVED LOSS FUNCTION

The bounding box loss function in the DETR model is calculated via a combination of the Generalized Intersection over Union ($GIoU$) and $L1$ loss, as shown in Equation 8:

$$L_{box}(b_i, \hat{b}_{s(i)}) = \lambda_{IoU} L_{GIoU}(b_i, \hat{b}_{s(i)}) + \lambda_{L1} L_{L1}(b_i, \hat{b}_{s(i)}) \tag{8}$$

In this formula, λ_{IoU} and λ_{L1} represent the weight coefficients for the $GIoU$ and $L1$ loss, respectively. The variable b_i denotes the coordinates of the ground truth bounding box for the i_{th} target to be detected, while $\hat{b}_{s(i)}$ refers to the coordinates of the predicted bounding box associated with the $s(i)_{th}$ prediction for the i_{th} target. L_{GIoU} and L_{L1} are the loss functions for $GIoU$ and $L1$, respectively, where the specific forms of the $GIoU$ loss are detailed in Equation 9 and Equation 10:

$$IoU(b, b^{gt}) = \frac{b \cap b^{gt}}{b \cup b^{gt}} \tag{9}$$

$$result = \begin{cases} IoU(b, b^{st}) - \dfrac{c - (b \cup b^{st})}{c}, IoU \neq 0 \\ -1 + \dfrac{(b \cup b^{st})}{c}, IoU = 0 \end{cases} \tag{10}$$

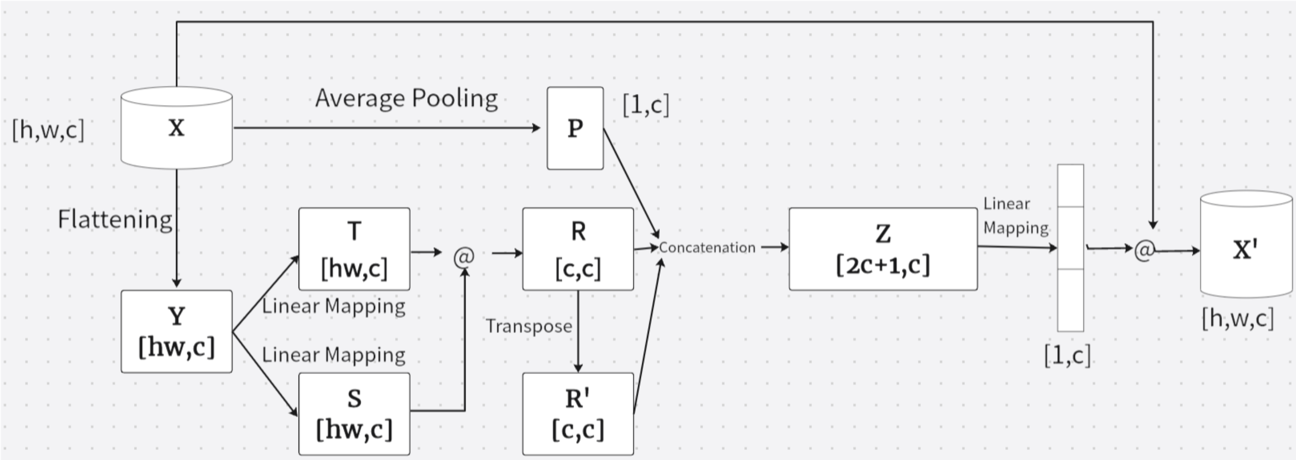


FIGURE 4: The design details of the global channel attention mechanism.

The *GIoU* is a distance metric used to evaluate the degree of overlap between bounding boxes, with a value range of (-1,1]. In this metric, c represents the area of the smallest enclosing area introduced due to attention to nonoverlapping regions, and $b \cup b^{st}$ represents the overlap between the predicted box and the true box. Although the *GIoU* can more accurately reflect the overlap between two objects than the *IoU* can, it also has several shortcomings: in special cases where there is a containment relationship between the predicted box and the true box, the loss values calculated by the *GIoU* and *IoU* are the same, making it difficult to determined their relative positional relationships effectively. This situation can slow the convergence of bounding box regression, thereby significantly extending the training time and failing to achieve effective bounding box regression.

To address this issue, the *CIoU* (complete intersection over union) introduces the ratio of the diagonal distance to the centre point distance, thereby improving the convergence problem of the *GIoU* when there is a containment relationship between the predict ed box and the true box. At the same time, the *CIoU* also considers the aspect ratio of the predicted box and the true box, making it more accurate in reflecting their overlap. The specific form of the *CIoU* loss function is shown in Equations 11 to 13:

$$v = \frac{4}{\pi^2}\left[\arctan\left(\frac{w^{gt}}{h^{gt}}\right) - \arctan(\frac{w}{h})\right]^2 \tag{11}$$

$$a = \frac{v}{(1 - IoU(b, b^{gt})) + v} \tag{12}$$

$$L_{CIoU} = 1 - IoU(b, b^{gt}) + \frac{m^2(b, b^{gt})}{c^2} + av \tag{13}$$

In the context of the CIoU loss function, w^{gt}, h^{gt}, w, h represent the width and height of the ground-truth bounding box, respectively, while $m(b, b^{gt})$ represents the Euclidean distance between the centers of the predicted bounding box and the target box. The improved bounding box loss function for the DETR model is shown in Equation 14.

$$L_{box}\left(b_i, \hat{b}_{s(i)}\right) = \lambda_{IoU} L_{CIoU}\left(b_i, \hat{b}_{s(i)}\right) + \lambda_{L1} L_{L1}\left(b_i, \hat{b}_{s(i)}\right) \tag{14}$$

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL PREPARATION

Given the scarcity of railway engineering datasets, this study constructed a construction tool dataset using the data collected by the railway system, which includes 351 images. To build the label, we used the LabelImage tool. The label covers a variety of construction tools, such as carts, motors, brooms, electric drills, wires, water pipes, woven bags, tool kits, buckets, blowers, tape measures, shovels, sand buckets, sanders, plastic buckets, and cement buckets, of which 16 are the most commonly used detection targets. The unique feature of this dataset is that all the images are sourced from onsite construction scenarios, reflecting real and complex data scenarios, so there is no need for further dataset expansion.

To effectively utilize this dataset, we divided the dataset images into training and testing sets at a 7:3 ratio and selected 30% of the images from the training set as the validation set. This partitioning method helps maintain the diversity and representativeness of the dataset while also ensuring that the model has sufficient generalization ability during training and testing. Through such data preparation work, we laid a solid foundation for subsequent model training and evaluation, with the aim of achieving satisfactory results in target detection tasks in railway engineering.

The experimental environment: The operating system is based on Windows 10 Professional Edition 64 bit (10.0, internal version 19045), the graphics card is NVIDIA GeForce RTX 3060, the system model is ASUS TUF Gaming A15 FA506QM_FA506QM, the processor is AMD Ryen 7 5800H with Radeon Graphics (16 CPUs), 3.2GHz, the memory is 16GB, and the deep learning framework based on Python is used.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3502438

B. EVALUATION INDICATORS

Our evaluation indicators were average precision (AP) and mean average precision (mAP). AP i is a commonly used evaluation indicator in target detection tasks and represents the average precision value at different intersection over union IoU thresholds; it can reflect the performance of the detector at different thresholds. The formula is as follows:

$$AP = \int_0^1 p(r)dr \tag{15}$$

The AP is the area under the accuracy curve at different recall rates, representing the average accuracy of the detector at different recall rates. The mAP is the average AP for multiple categories and is a commonly used comprehensive evaluation indicator in target detection tasks. The formula is as follows:

$$mAP = \frac{(AP_1 + AP_2 + \ldots + AP_n)}{n} \tag{16}$$

The mAP is an important evaluation indicator used to comprehensively evaluate the performance of target detectors in multiple categories and can comprehensively evaluate the accuracy and stability of target detection models in different categories. By comprehensively considering the AP and mAP, the performance of target detection models in different scenarios can be comprehensively evaluated. Usually, the larger the AP and mAP values are, the better, and they are important references for model optimization and improvement.

C. ANALYSIS OF EXPERIMENTAL RESULTS

During the training process of RE-DETR, the loss curve of the model is shown below.The loss curve tends to stabilize as the number of training rounds increases. When the number of epochs is approximately 80, the RE-DETR model gradually converges, and no fitting phenomenon occurs during this training process.

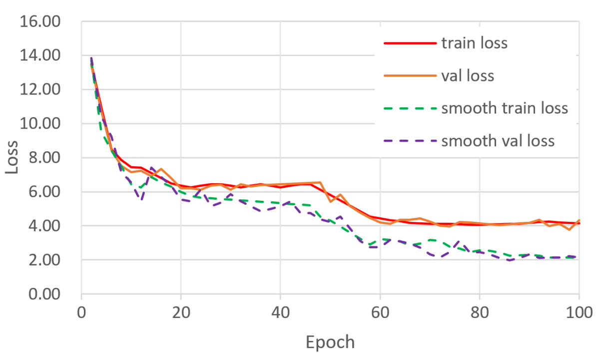


FIGURE 5: Changes in loss values of the RE-DETR model, where training loss refers to the error or loss value calculated during the training phase of a machine learning model. Smoothing training loss involves applying a smoothing technique to loss values over multiple training iterations or epochs to reduce fluctuations and provide a clearer trend of model performance over time.

To verify the performance of the RE-DETR model for dark time-domain tool detection, this paper designs a set of ablation experiments and comparative experiments for the model. We verify the impact of different improvements on network performance through ablation experiments and then conduct comparative experiments with current mainstream networks (DETR, YOLOv5, YOLOX, and RYOLO) through RE-DETR. On the basis of the experimental results, we comprehensively analyse the performance of the model.

1) Ablation experiment

To analyze the impact of the improvements made in this article on model performance, three sets of experiments are designed to analyze different improvements. Each experiment is tested on the same training parameters and different model contents. The performance test results of the model are shown in Table 1. Compared with the first and second rows in Table 1, the addition of an improved Retinexnet image enhancement module to the original DERT improved the model's detection ability for full time domain images, with a mAP increase of 0.92%. Comparing the experimental results in the second and third rows, after adding Retinexnet and modifying the loss function to CIoU, mAP increased by 1.84% again. Continuing to compare the experimental results in the third and fourth rows, adding a lightweight attention module can improve inter channel communication ability while weakening the impact of noise on deep networks, resulting in a 0.97% increase in mAP. This indicates the effectiveness and rationality of the improved model in this article. Figure 6 shows a schematic diagram of the detection effect of the DETR model before and after the improvement.

To analyse the impact of the improvements made in this study on model performance, three sets of experiments are designed to analyse different improvements. Each experiment is tested on the same training parameters and different model contents. The performance test results of the model are shown in Table 1. Compared with the first and second rows in Table 1, the addition of an improved RetinexNet image enhancement module to the original DERT improved the model's detection ability for full time domain images, with an mAP increase of 0.97%. Comparing the experimental results in the second and third rows, after adding RetinexNet and modifying the loss function to the $CIoU$, the mAP increased by 1.81%. While continuing to compare the experimental results in the third and fourth rows, adding a lightweight attention module can improve the interchannel communication ability while weakening the impact of noise on deep networks, resulting in a 0.97% increase in mAP. This indicates the effectiveness and rationality of the improved model in this study.

2) Model performance comparison experiment

To better test and verify the detection performance of the RE-DETR model, a comparative experiment is conducted with current mainstream detection models. The comparative test results of various railway engineering tool detection methods

TABLE 1: Experimental results of different improvement methods

method	mAP	FPS
DETR	73.61%	29.28
DETR+Retinexnet	74.53%	30.18
DETR+Retinexnet+Clou	76.37%	**33.62**
DETR+Retinexnet+CloU+Adaptive attention mechanism	**77.34**%	32.43

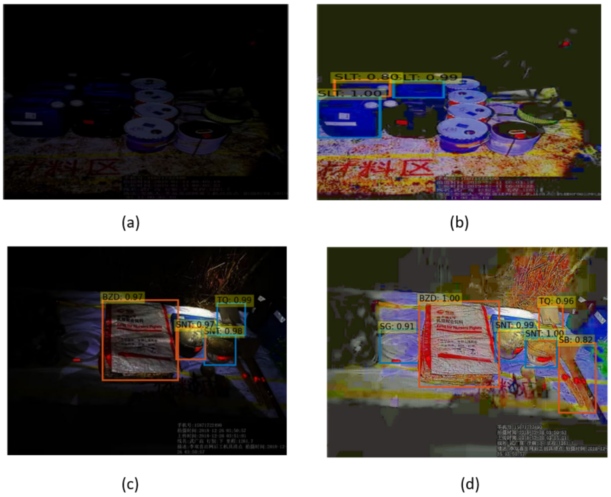


FIGURE 6: Schematic Diagram of Detection Performance Before and After Improvement in DETR Model. (a) Detection effect of the original DETR under low brightness conditions. (b) Detection effect of the Re-DETR under low brightness conditions. (c) Detection effect of the original DETR under medium brightness conditions. (d) Detection effect of the Re-DETR under medium brightness conditions.

are shown in Table 2. The mAP value of the RE-DETR model reaches 78.62%, which is 3.34% higher than that of the original DETR algorithm. According to the table, the AP values of plastic buckets, motors, electric drills, polishing machines, and woven bags improved to varying degrees compared with those of the original DETR algorithm, showing that the improved model achieves better detection performance than other mainstream target detection models (YOLOX, RetinaNet, YOLOV5). In particular, in terms of tool detection in the dark time domain, RE-DETR has significant advantages. While ensuring high-precision target detection, the FPS of the model itself does not significantly decrease, and it still has certain advantages in terms of detection speed compared with mainstream models.

TABLE 2: Performance comparison of mainstream target detection models

Model	AP(IoU=0.6)					mAP	FPS
	Plastic bucket	motor	Electric drill	Polishing machine	Woven bag		
Retinanet	0.97	0.89	0.67	0.14	0.55	71.02	23.67
YOLOV5	0.65	0.91	0.34	0.04	0.53	68.37	41.37
YOLOV8	0.88	0.91	0.66	0.53	0.45	74.16	**34.26**
RYOLO	0.99	0.92	0.75	0.61	0.57	77.26	32.25
DETR	0.81	0.91	0.72	0.61	0.49	73.61	29.28
RE-DETR	**0.99**	**0.95**	**0.84**	**0.73**	**0.67**	**77.34**	32.43

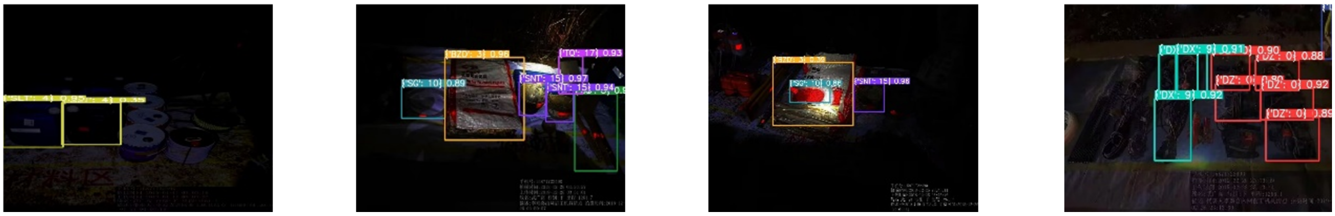
According to the data in Table 2, the improved DETR model significantly outperforms other traditional object detection models, such as the YOLO series and Faster-RCNN, in the multitarget detection scenario of railway maintenance sites. Moreover, compared with the original DETR model, the improved DETR model achieves higher detection accuracy. Figure 7 illustrates the detection effect of the improved DETR model. In Figures a to e, we compare the detection results of our network with those of YOLOV5, YOLOV8, RYOLO, and DETR under medium and low brightness environments. Row a presents the test results for YOLOV5, row b for YOLOV8, row c for RYOLO, row d for DETR, and row e for RE-DETR. Comparing the first column of images, we can see that the detection results of RE-DETR are significantly better than those of the original DETR and outperform the other three networks as well. In the second column, RE-DETR shows the best detection performance, successfully detecting six targets with high precision. While YOLOV5 detects the same number of targets as RE-DETR, RE-DETR achieves better accuracy. In the third column comparison, it is noted that only YOLOV5 and RE-DETR detect the same number of targets. The original low-light image processed by the RE-DETR network shows some distortion due to the excessively dark environment; however, in terms of both the number of detected targets and detection accuracy, it still surpasses the other four networks. In the fourth column, it is observed that under complex overlapping conditions, the detection counts for YOLOV5, YOLOV8, and RE-DETR are all eight, while RYOLO detects nine targets. In terms of accuracy, both YOLOV5 and RE-DETR perform better. Overall, RE-DETR demonstrates a higher detection capability in low-light environments compared to the other four networks.

V. CONCLUSION

This article proposes an RE-DETR tool detection model to address the inability of existing target detection models to detect railway engineering tools efficiently in the dark time domain. The model, which is based on the DETR framework, incorporates an improved RetinexNet image enhancement module, introduces a global channel attention mechanism in DETR, and utilizes a comprehensive loss function. The main goal in the future is to further improve the recognition accuracy of the model and further refine the classification ability of the dark time domain model.

ACKNOWLEDGMENT

This work was funded by the National Natural Science Foundation of China (No.62462046), the Science and Technology Research Project of Jiangxi (China)Provincial Department of Education (No.GJJ202516)(No.GJJ219002)(No.GJJ2203220), the Doctoral Initiation Fund Project of Nanchang Institute of Science & Technology (No.NGRCZX2008), 2024 Innovation and Entrepreneurship Training Program for College Students in Jiangxi Province(S202413421019).

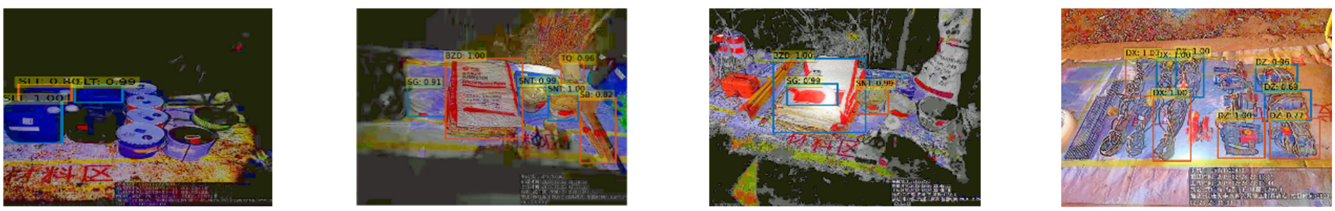


(a) YOLOV5

(b) YOLOV8

(c) RYOLO

(d) DETR



(e) RE-DETR

FIGURE 7: Visualization of the detection results of the dataset under different models.

REFERENCES

[1] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," 2020.

[2] W. Y. J. L. Chen Wei, Wenjing Wang, "Deep retinex decomposition for low-light enhancement," in *British Machine Vision Conference*, 2018.

[3] Z. Zheng, B. Xu, J. Ju, and et al., "Circumferential local ternary pattern: New and efficient feature descriptors for anti-counterfeiting pattern identification," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 970–981, 2022.

[4] K. Mayathevar, M. Veluchamy, and B. Subramani, "Fuzzy color histogram equalization with weighted distribution for image enhancement," *Optik*, vol. 216, p. 164927, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030402620307634

[5] J. V. Manjón and P. Coupé, "Mri denoising using deep learning and non-local averaging," *ArXiv*, vol. abs/1911.04798, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:207863583

[6] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.

[7] W. Huang, Y. Zhu, and R. Huang, "Low light image enhancement network with attention mechanism and retinex model," *IEEE Access*, vol. 8, pp. 74 306–74 314, 2020.

[8] X. Lv, Y. Sun, J. Zhang, F. Jiang, and S. Zhang, "Low-light image enhancement via deep retinex decomposition and bilateral learning," *Signal Processing: Image Communication*, vol. 99, p. 116466, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092359652100223X

[9] C.-H. Lee, J.-L. Shih, C.-C. Lien, and C.-C. Han, "Adaptive multiscale retinex for image contrast enhancement," in *2013 International Conference on Signal-Image Technology Internet-Based Systems*, 2013, pp. 43–50.

[10] X. Zhang and X. Wang, "Marn: Multi-scale attention retinex network for

low-light image enhancement," *IEEE Access*, vol. 9, pp. 50 939–50 948, 2021.

[11] J. Yang, Y. Xu, H. Yue, Z. Jiang, and K. Li, "Low-light image enhancement based on retinex decomposition and adaptive gamma correction," *IET image processing*, vol. 15, no. 5, pp. 1189–1202, 2021.

[12] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang, "Retinexformer: One-stage retinex-based transformer for low-light image enhancement," 2023. [Online]. Available: https://arxiv.org/abs/2303.06705

[13] J. Zhu and Z. Chen, "Real time face detection system using adaboost and haar-like features," in *2015 2nd International Conference on Information Science and Control Engineering*, 2015, pp. 404–407.

[14] D. F. Llorca, R. Arroyo, and M. A. Sotelo, "Vehicle logo recognition in traffic images using hog features and svm," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2013, pp. 2229–2234.

[15] M. Maity, S. Banerjee, and S. Sinha Chaudhuri, "Faster r-cnn and yolo based vehicle detection: A survey," in *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 2021, pp. 1442–1447.

[16] P. Bharati and A. Pramanik, "Deep learning techniques—r-cnn to mask r-cnn: a survey," *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, pp. 657–668, 2020.

[17] T. Wang, F. Yang, and K.-L. Tsui, "Real-time detection of railway track component via one-stage deep learning networks," *Sensors*, vol. 20, no. 15, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/15/4325

[18] A. Kumar, Z. J. Zhang, and H. Lyu, "Object detection in real time based on improved single shot multi-box detector algorithm," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, no. 1, p. 204, 2020.

[19] J. Wu, P. Fan, Y. Sun, and W. Gui, "Ghost-retinanet: Fast shadow detectionmethod for photovoltaic panels based on improved retinanet." *CMES-Computer Modeling in Engineering & Sciences*, vol. 134, no. 2, 2023.

[20] S. Cheng, J. Lu, M. Yang, S. Zhang, Y. Xu, D. Zhang, and H. Wang, "Wheel hub defect detection based on the ds-cascade rcnn," *Measurement*, vol. 206, p. 112208, 2023.

[21] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable detr: Deformable transformers for end-to-end object detection," *arXiv preprint arXiv:2010.04159*, 2020.

[22] J. Du, T. Lin, C. Jiang, Q. Yang, C. F. Bader, and Z. Han, "Distributed foundation models for multi-modal learning in 6g wireless networks," *IEEE Wireless Communications*, vol. 31, no. 3, pp. 20–30, 2024.

PLACE PHOTO HERE

ZHAOHUI ZHENG received the Ph.D. degree from Wuhan University. He is currently a Lecturer with the School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology. His research interests include machine learning, pattern recognition, robotic dynamics, and mathematical modeling.

PLACE PHOTO HERE

SHUILONG ZOU has a master's degree and is a professor and a provincial-level "advanced double-qualified teacher". He graduated from Nanchang University in 2010 with a major in computer technology and his research interests include the Internet of Things, big data, and artificial intelligence. He currently works at the School of Electronics and Information Engineering at Nanchang Vocational College of Applied Technology. His research interests include artificial intelligence, and the Internet of Things.

PLACE PHOTO HERE

TINGFENG JI is currently pursuing the B.S. degree with the School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology. His research interests include image processing, pattern recognition.

PLACE PHOTO HERE

JIANPING JU received the B.S. degree in optical information science and technology from the Huazhong University of Science and Technology and the M.S. degree in artificial intelligence and pattern recognition from the Wuhan Institute of Technology, Wuhan, China, in 2003 and 2010, respectively. He is currently a Professor with the School of Artificial Intelligence, Hubei Business College. His research interestsinclude machine learning and pattern recognition.

PLACE PHOTO HERE

GUAN QING is currently a Professor with the School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology. His research interests include consist of signal processing, and mechanical fault diagnosis.

PLACE PHOTO HERE

QIANG ZHANG received the master's degree in signal and informationprocessing from Nanchang Hangkong University. She is currently a Lecturerwith the Department of Public Courses, Wuhan Railway Vocational Collegeof Technology. Her research interests include image processing, and patternrecognition.

PLACE PHOTO HERE

QING ZHOU is currently pursuing the B.S. degree with the Nanchang Normal College of Applied Technology. Her research interests include image processing, pattern recognition.

PLACE PHOTO HERE

YINGJIAN HE is currently pursuing the B.S. degree with the School of Information and Artificial Intelligence, Nanchang Institute of Science & Technology. His research interests include image processing, pattern recognition.

• • •