



ARTICLE

Unsupervised Anomaly Detection System for High-Speed Railway Noise Barrier Using UAV Imagery

Jing Cui¹, Yong Qin^{2,*}, Yixuan Geng³, Miao Guo^{4,*}, Xue Yang⁴ and Wanyin Shi⁵

¹China Railway Signal & Communication Corp., Beijing, China

²The State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, China

³Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China

⁴Faculty of Transportation Engineering, Kunming University of Science and Technology, Kunming, China

⁵Yunnan Institute of Economics and Management, Kunming, China

*Corresponding Authors: Yong Qin. Email: yqin@bjtu.edu.cn; Miao Guo. Email: guomiao@kust.edu.cn

Received: 27 February 2026; Accepted: 09 May 2026; Published: 24 August 2026

ABSTRACT: Noise barriers (NBs) play a significant role in reducing railway noise and preventing foreign-object intrusion. However, surface damage, corrosion, rust, missing components, and local deformation may gradually reduce their structural reliability and threaten railway operation safety. Because NB anomalies are diverse and defect samples are limited, it remains difficult to build a general detector using conventional supervised learning. To address this problem, this study proposes an unsupervised anomaly detection system for railway NBs using UAV imagery. First, a color-prior-based NB localization algorithm is developed in the HSV color space to extract NB regions without cumbersome pixel-level labeling or localization-network training. Second, a teacher-student-autoencoder framework, denoted UADNet, is designed for NB anomaly detection. In this framework, a GhostConv-based student network learns normal local feature regression, while a skip connection autoencoder (SCAE) models normal global feature consistency. Anomalies are highlighted when both branches fail to reproduce the expected normal responses. Experiments on public anomaly-detection benchmarks and a customized UAV railway NB dataset demonstrate that the proposed system achieves high detection accuracy, accurate localization, and competitive inference speed in complex rail environments.

KEYWORDS: Anomaly detection; railway noise barrier; unmanned aerial vehicle (UAV); unsupervised learning

1 Introduction

As of 2023, China has the world's largest passenger network with 45,000 km high-speed railroad [1]. In recent years, the rapid expansion of high-speed rail network has brought great pressure to the inspection and maintenance of railway infrastructure. As shown in Fig. 1, the noise barriers, as crucial devices along railways, are primarily employed to mitigate the impact of train noise on nearby residents and prevent intrusion of foreign objects. Unfortunately, the complicated natural environment and the intense air pressure generated by high-speed train operations would progressively damage noise barriers (NBs), including but not limited to, loose nuts, missing components, and corroded parts. These anomalies might result in the collapse of NBs onto the track areas, potentially leading to serious accidents. Therefore, it is of great importance to conduct regular inspection of the NBs.

In the past decades, anomaly detection for NBs has mainly relied on manual inspection, involving on-site track patrols or reviews of videos captured by inspection vehicles. These inspections are labor intensive

and usually focus on the track-facing side, leading to limited coverage and low efficiency. Recently, UAV-based automatic detection methods have attracted increasing attention in railway engineering because of their high efficiency, low cost, and independence from terrain constraints. In particular, the pioneering work SCYOLO [2] has demonstrated effective NB surface-defect detection in UAV imagery. Nevertheless, supervised learning still has clear limitations in this task. First, it is trained on predefined defect classes and therefore cannot naturally generalize to unknown anomalies outside the annotated categories. Second, the collection and annotation of sufficient defect samples remain difficult in practical railway inspection, which limits deployment at scale.

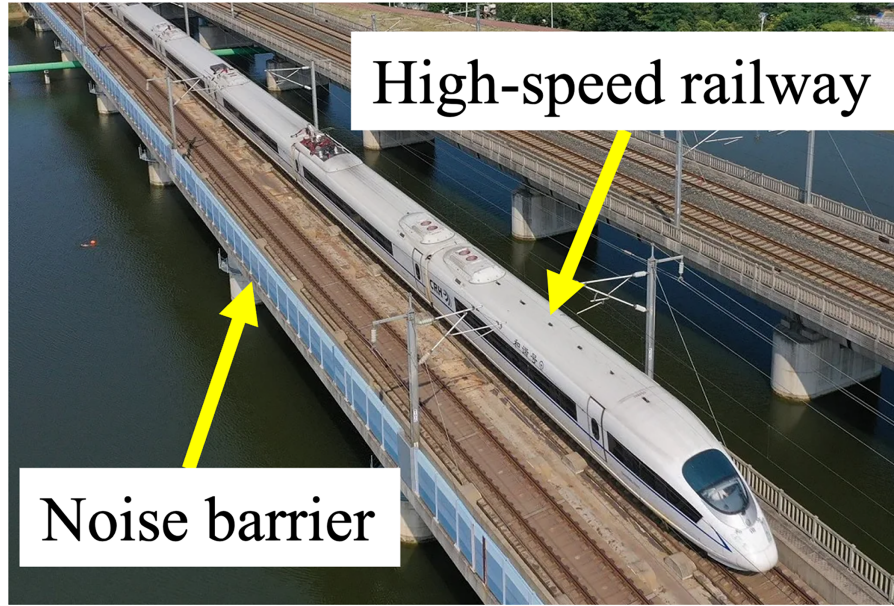


Figure 1: Noise barrier structure in high-speed railway system.

Recently, unsupervised learning models such as student-teacher (S-T), autoencoder (AE), and outlier-detection methods have made important progress in anomaly detection. In the S-T framework, the teacher is usually pretrained on large-scale natural images, while the student is trained only on anomaly-free samples; anomalies are then identified through feature-regression errors at inference time. This paradigm is attractive for NB inspection because it does not require defect labels and is naturally compatible with unknown anomaly categories. However, our investigation shows that a conventional S-T design still faces three difficulties in UAV-based NB inspection. First, many existing S-T architectures use identical or highly similar teacher and student structures, which may make the student overly capable of mimicking the teacher and thus weaken anomaly separation. Second, lightweight students are preferred for edge deployment, but an excessively weak student may lose local discrimination ability for small defects. Third, NB images contain strong structural repetition and large irrelevant backgrounds, which create redundant gradients during training and unnecessary false alarms during inference.

Compared with PatchCore [3] and EfficientAD [4], the proposed method is not intended as a generic replacement for all industrial anomaly detectors; rather, it is tailored to UAV-based railway NB inspection. PatchCore relies on memory-bank retrieval over patch embeddings and works well when normal patch distributions remain stable, whereas our system first removes large non-target regions and then fuses complementary local and global discrepancy cues. EfficientAD provides an efficient teacher-student-autoencoder baseline for generic anomaly detection, while our main contribution lies in adapting

this paradigm to repetitive NB textures and embedded deployment through an HSV-based localization stage, a GhostConv-constrained student branch, and a skip-connected autoencoder that strengthens global consistency modeling.

- (1) The system first integrates an NB localization algorithm based on HSV color statistics to remove irrelevant regions before anomaly detection, thereby reducing interference from sky, track, and surrounding background areas in UAV imagery.
- (2) UADNet adopts a teacher-student (S-T) structure in which the student branch incorporates GhostConv to introduce a controlled representational gap from the teacher, improving sensitivity to local abnormal patterns while keeping the model lightweight.
- (3) A skip connection autoencoder (SCAE) is further incorporated into UADNet to complement the S-T branch from a global perspective. The SCAE improves normal-structure reconstruction through multi-scale feature fusion and provides complementary anomaly cues for globally inconsistent regions.
- (4) Extensive experiments on public datasets and the customized UAV-based NB dataset demonstrate that the proposed system achieves high detection accuracy, robust localization, and competitive real-time performance in complex railway environments.

The remainder of this paper is organized as follows. [Section 2](#) reviews related work. [Section 3](#) describes the proposed NB localization algorithm and the UADNet architecture. [Section 4](#) reports the experimental results, additional analysis, and deployment discussion. Finally, [Section 5](#) concludes the paper and outlines future work.

2 Related Work

Over the past decade, convolutional neural networks (CNNs) have achieved remarkable success in computer vision because of their strong representation and computation capabilities. CNN-based object-detection algorithms can generally be divided into one-stage [\[5,6\]](#) and two-stage detectors [\[7,8\]](#). One-stage methods such as the YOLO series [\[5\]](#) and RetinaNet [\[6\]](#) complete feature extraction and prediction in a single pipeline, while two-stage methods such as Faster R-CNN [\[8\]](#) first generate candidate regions and then refine them. Because of this additional proposal stage, two-stage detectors usually have lower processing speed than one-stage models.

In recent years, with the increasing pressure on railway infrastructure inspection, UAV-based vision methods have received increasing attention and gradually become a research hotspot. For instance, Wu et al. [\[9\]](#) proposed an automatic detection system for railway track components using a novel all-in-one YOLO framework, and the experiments demonstrated both high detection accuracy and real-time speed. Tong et al. [\[10\]](#) introduced a fully decoupled residual CNN to parse railway scenes in UAV imagery and effectively capture local track details. Despite the success of UAV inspection in railway engineering, these methods cannot be directly transferred to NB anomaly detection because of the scene-specific characteristics of NB imagery, including large repetitive textures, oblique viewpoints, and small defect targets. Notably, Cui et al. [\[2\]](#) conducted defect detection on NBs using an improved YOLO framework and achieved 92.2 mAP with 78.7 FPS. However, supervised methods still require cumbersome labeling and sufficient defect samples, which constrains their applicability in large-scale field inspection.

To date, many effective unsupervised anomaly detection methods have been developed, including outlier detection, student-teacher learning, memory-bank matching, and generative reconstruction models. For example, Roth et al. [\[3\]](#) proposed PatchCore, which applies nearest-neighbor retrieval on a compact coreset of normal patch embeddings and achieves strong results on MVTec AD. Deng and Li [\[11\]](#) introduced reverse distillation to enhance anomaly sensitivity through teacher-student discrepancy learning.

Somepalli et al. [12] proposed an adversarial mirrored autoencoder for unsupervised anomaly detection. EfficientAD [4] further demonstrated that a carefully designed student-teacher-autoencoder framework can achieve both high accuracy and low latency. More recently, vision foundation models and vision-language models have also been explored for anomaly detection. Segment Anything provides strong general-purpose segmentation priors [13], while CLIP-based methods such as WinCLIP [14] and AnomalyCLIP [15] improve open-set anomaly understanding through image-text alignment. These methods are promising, but they typically require higher computation, prompt design, or large pretrained models, which are less suitable for lightweight deployment on UAV edge devices. This motivates the present work on a compact, scene-specific unsupervised solution for railway NB inspection.

3 Methodology

3.1 Overview of Research Methodology

This section presents the overall workflow of the proposed noise barrier (NB) anomaly detection system. As sketched in Fig. 2, the system mainly includes two stages: (1) an NB localization algorithm that calculates the NB color distributions of UAV images in the HSV color space to accurately localize the NB area; and (2) an unsupervised anomaly detection network, UADNet, which combines a Ghost student-teacher branch and a skip connection autoencoder (SCAE) to detect anomaly regions on the NB surface.

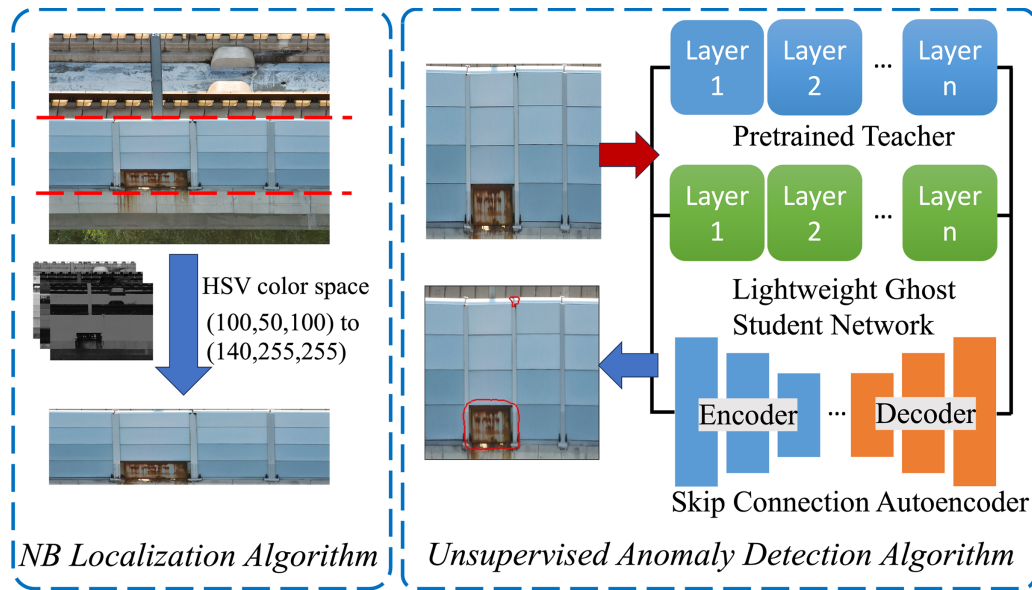


Figure 2: Overview of proposed system in this study.

3.2 Color-Prior-Based NB Localization Algorithm

There are numerous irrelevant regions in the upper and lower parts of the NB images collected by UAV. These background interferences are prone to causing excessive false alarms. Therefore, NB region extraction is the first key step for NB anomaly detection, because it determines whether the subsequent UADNet analysis is focused on meaningful structural content.

In image color extraction, the HSV color space is preferred over the RGB color space because of its more intuitive representation [16]. The average HSV histograms of the NB dataset are presented in Fig. 3, where each row corresponds to a different region in Fig. 2. It is evident that the NB structure predominantly exhibits

a blue hue, with a notable distribution difference from the other two regions. Accordingly, we constrain the HSV range for image extraction from (100, 50, 100) to (140, 255, 255). The localization procedure mainly involves three steps: (1) connected-component analysis is performed to identify the blue regions; (2) the bottommost position of the dominant blue region is used to determine the lower boundary of the NB area; and (3) an upward scan is carried out to confirm the upper boundary of the NB area.

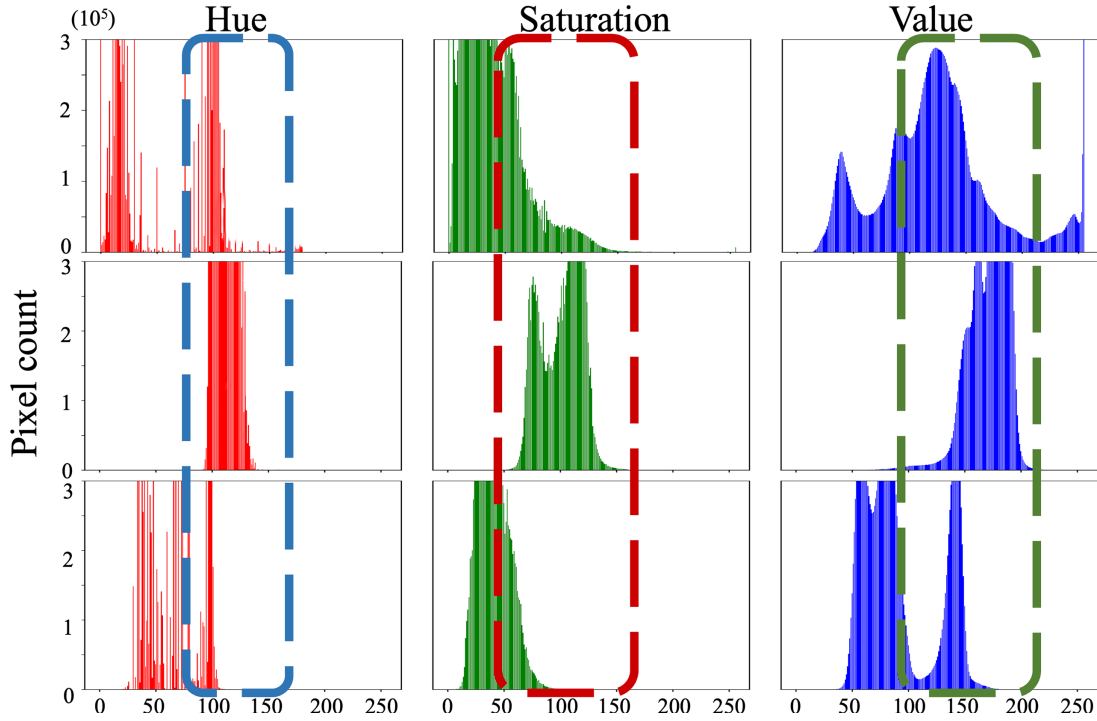


Figure 3: Average HSV histograms of the noise barrier images.

To improve robustness under illumination and appearance changes, the HSV prior is used only as a coarse region-proposal cue rather than a final defect detector. In the revised validation under sunny, cloudy, backlit, and light-rain conditions, the same localization procedure maintained stage-1 precision above 96%, because the hue constraint is combined with connected-component connectivity and boundary-scanning rules. For barriers with different dominant colors, the same pipeline can be retained by re-estimating the dominant hue interval from a small set of normal frames collected on the target line; no defect annotation or localization-network retraining is required.

3.3 Lightweight Ghost Student–Teacher

The student-teacher (S-T) model uses a teacher network pretrained on natural images for broad visual representation and a student network trained only on anomaly-free samples. During inference, the student's limited generalization outside the normal distribution leads to inaccurate predictions on abnormal regions. Anomaly scores are therefore obtained from the student's regression error relative to the teacher. In this study, we use the pretrained Patch Description Network (PDN) [4] with WideResNet-101 as the teacher network in UADNet, which reduces training cost and preserves strong normal-feature priors.

Fig. 4 illustrates the specific network architecture of the proposed UADNet, which comprises three main branches: the teacher network (PDN), the student network (GPDN), and the skip connection autoencoder. Table 1 provides the detailed configuration of the S-T network. The GhostConv structure is integrated

into the student network to avoid building a student that is too similar to the teacher while still keeping the model lightweight. As shown in Table 2, GhostConv combines intrinsic convolutions with inexpensive linear feature generation, which eliminates a large amount of redundancy in the intermediate feature map [17]. The local anomaly map of the S-T branch is obtained by calculating the element-wise squared difference between the two sets of output feature maps (“Diff1” in Fig. 4).

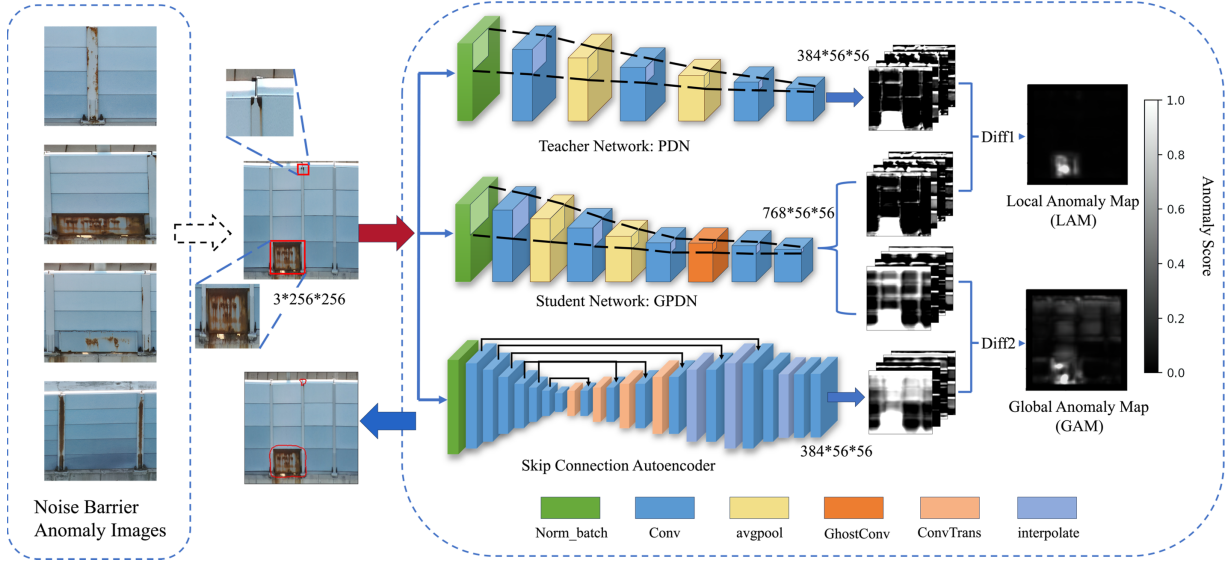


Figure 4: Specific network architecture of UADNet.

Table 1: Configurations of the S-T network (Teacher/Student).

Layer Name	Stride	Kernel Size	Numbers
Conv (T/S)	1×1	$4 \times 4/3 \times 3$	128
Pool (T/S)	2×2	2×2	128
Conv (T/S)	1×1	$4 \times 4/3 \times 3$	256
Pool (T/S)	2×2	2×2	256
Conv (S)	1×1	3×3	256
GhostConv (S)	–	–	256
Conv (T/S)	1×1	3×3	256
Conv (T/S)	1×1	$4 \times 4/3 \times 3$	384/768

Table 2: Structure parameters of the GhostConv module. “From –1” indicates that the input of this layer is that of the previous layer.

Number	From	Module	Args
1	–1	Conv2d	[128, 1, 1, 0]
2	–1	ReLU	–
3	–1	Conv2d	[128, 3, 1, 1]
4	–1	ReLU	–
5	[–1, 2]	Concat	–

The role of GhostConv here is not merely parameter reduction. Because part of the student's feature maps are generated from cheap linear transformations of intrinsic features, the student branch is intentionally prevented from becoming an exact structural replica of the teacher. This controlled representational asymmetry is important in repetitive NB scenes: it suppresses redundant gradient propagation from repeated background textures and makes the residual error more sensitive to local deviations such as rust, cracks, missing fasteners, and damaged panel boundaries. In other words, GhostConv keeps the student sufficiently expressive for normal-pattern regression while reducing the risk that the student will overfit and reproduce abnormal teacher responses.

Backpropagation reveals that the receptive fields of the teacher and student networks are 33×33 and 42×42 , respectively. Therefore, the S-T branch focuses mainly on local abnormal regions in the image. The visual feature maps at different stages of the S-T network are depicted in Fig. 5. With prior knowledge of natural-image characteristics, the deep layers in the teacher network are better at highlighting anomalous regions than the corresponding deep layers in the student network. For instance, local anomalies can be separated by subtracting the teacher's Conv4 feature map from the first map in the student's Conv5, followed by threshold segmentation. Therefore, the S-T branch effectively detects and localizes anomalies by comparing the output differences between teacher and student during inference.

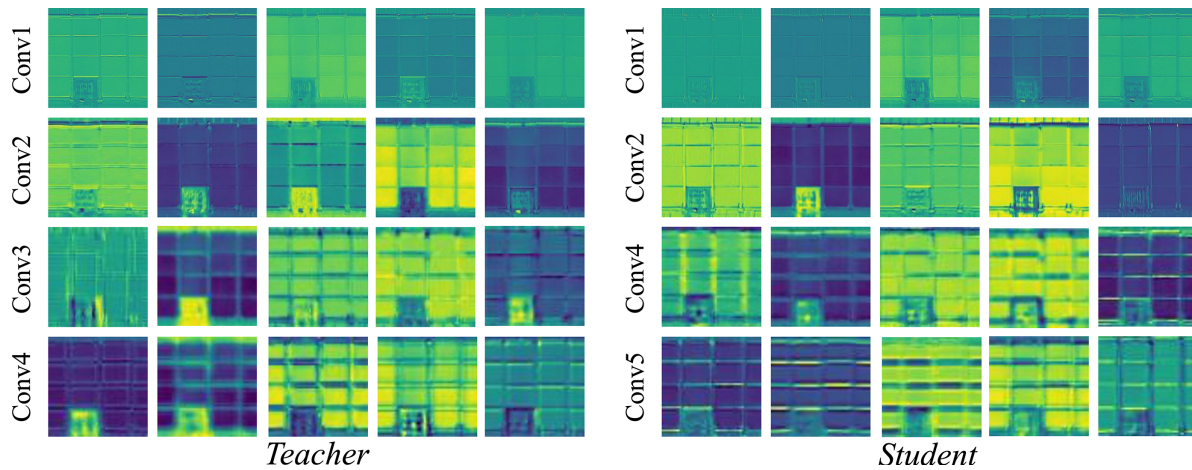


Figure 5: Visual feature maps of different stages in the S-T network.

3.4 Skip Connection Autoencoder Network

Recently, autoencoders (AEs) have become important tools in anomaly detection by using reconstruction errors to identify defects. Given the effectiveness of multi-scale feature fusion, we build a skip connection autoencoder (SCAE) in UADNet to complement the S-T branch from a global perspective. In the structure shown in Fig. 4, the SCAE is trained to predict the teacher's output. Since the AE struggles to reconstruct fine details in the teacher's feature space [4], we double the final output channel of the student network from 384 to 768. This adjustment allows the student network to predict the outputs of both the teacher and the SCAE.

As shown in Fig. 4, the SCAE encoder uses convolution operations to downscale the intermediate feature map. In the decoder, upsampling is performed using transposed convolution and bilinear interpolation. Additionally, feature maps of the same scale from the encoder and decoder are concatenated. The skip connection structure preserves high-resolution information, improves image reconstruction accuracy, and alleviates gradient vanishing as well as blurry reconstruction. The hyperparameters of the SCAE are listed in Table 3. Unlike the student branch, which focuses on small patches, the SCAE learns broader structural

consistency. This difference makes the student and SCAE outputs suitable for computing a complementary global anomaly map.

Table 3: Configurations of the skip connection autoencoder.

Layer	Size	Number	Repeat
EncConv	3×3	32	2
EncConv	3×3	64	4
EncConv	4×4	64	1
DecConvT	4×4	64	3
DecConv	3×3	64	
DecConvT	4×4	64	1
DecConv	3×3	32	
Bilinear		Resize the input features maps from 32 to 64	
DecConv	3×3	32	1
Bilinear		Resize the input features maps from 64 to 128	
DecConv	3×3	64	2
Bilinear		Resize the input features maps from 64 to 56	
DecConv	3×3	64	1
DecConv	1×1	384	1

The role of SCAE is therefore not redundant with that of the student branch. The S-T branch is more responsive to small local deviations, while the SCAE emphasizes global structural consistency and provides additional evidence for larger or spatially extended defects. This local-global complementarity is especially useful in UAV NB imagery, where a defect may appear either as a small corrosion point or as a broader damaged panel region.

The global map is obtained by calculating the pixel-wise squared differences between the feature maps of the student output and the SCAE output (“Diff2” in Fig. 4). The final anomaly output is obtained by fusing the global map and the local map. We visualize the anomaly maps of UADNet in Fig. 6. It can be observed that the local map primarily focuses on the most prominent anomalous regions in the image, while the global map generates reconstruction errors based on broader structural consistency. The final fused mask achieves accurate anomaly detection when compared with the ground truth.

3.5 Loss Function and Anomaly Map Normalization

In UADNet, local and global anomaly maps are computed based on the squared output feature maps of the teacher and student networks, as well as the squared output feature maps of the student network and autoencoder. Therefore, the loss function mainly comprises three parts: L_{ST} , L_{AE} and L_{STAE} . L_{ST} represents the loss between teacher and student network, as shown in Eq. (1).

$$L_{ST} = (T(I)_{c,w,h} - S(I)_{c,w,h})_{\text{Phard}}^2 + (CWH)^{-1} \sum_c \|S(P)_c\|_F^2 \quad (1)$$

where I represents the training image, T and S denote the teacher and student networks, respectively, and CWH corresponds to the channel number, width and height of the feature map. The hard quantile is set to 0.999 as recommended, and only the average value of the portion exceeding the quantile value is calculated. This design prevents a large number of easy pixels from dominating the regression objective and makes

training focus on the most difficult normal responses, which is beneficial for highly repetitive NB textures. is a randomly selected image extracted from the pre-training dataset ImageNet, serving to prevent the student from generalizing its mimicry of the teacher to out-of-distribution images.

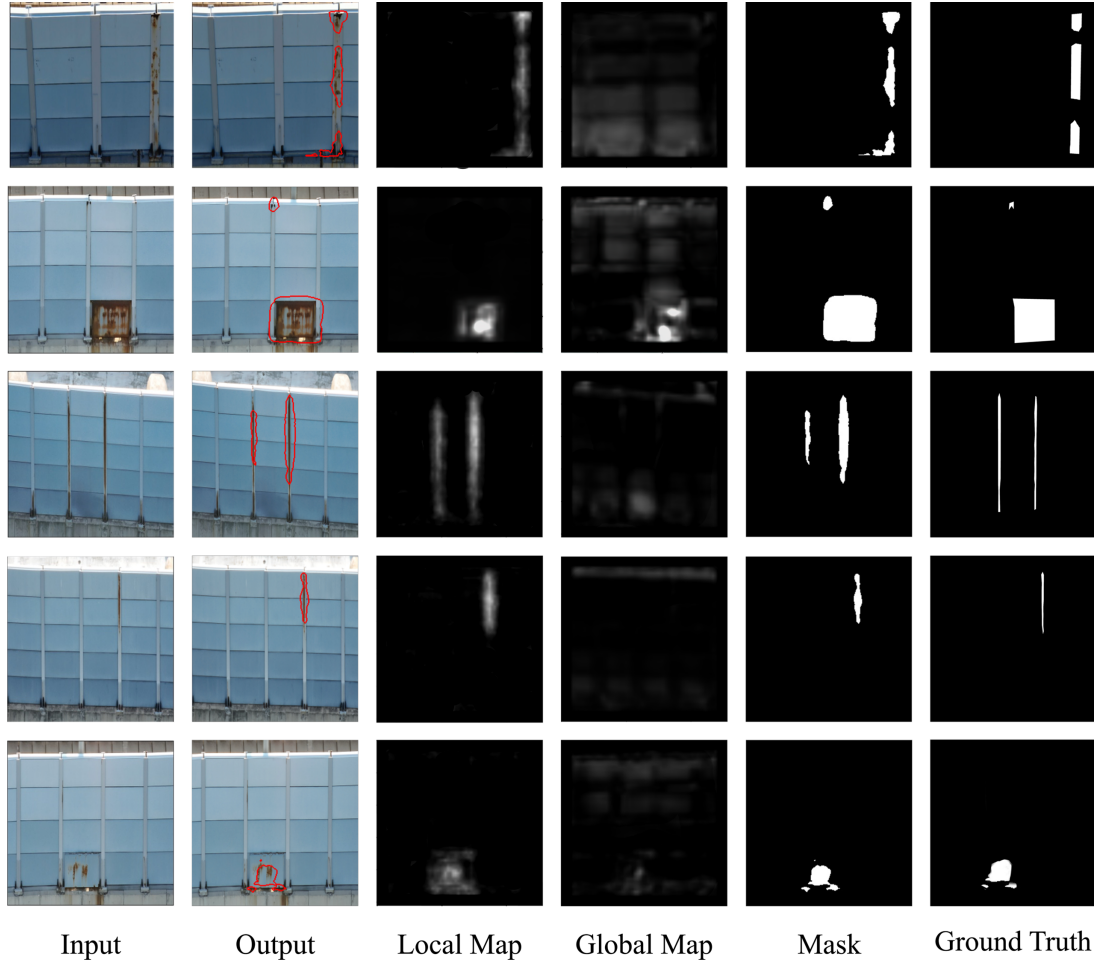


Figure 6: Anomaly map visualization of noise barrier images.

The formulas of the other two parts of the loss function, L_{AE} and L_{STAE} , are defined in (2) and (3):

$$L_{AE} = (CWH)^{-1} \sum_c \|T(I)_c - A(I)_c\|_F^2 \quad (2)$$

$$L_{STAE} = (CWH)^{-1} \sum_c \|A(I)_c - S'(I)_c\|_F^2 \quad (3)$$

where A is the SCAE, and S' corresponds to the additional output channel of the student network. The total loss function is the sum of these three components.

Before combining the anomaly maps, it is essential to normalize them to a comparable scale. Otherwise, noise in one anomaly map may obscure accurate detection in the other, rendering the merged map less reliable. The scale factors of the anomaly maps are determined through the ablation experiment in Section 4.3.

4 Experiments and Results

This section presents the ablation experiments on UADNet's structure and anomaly-map scale factors using the UAV-based noise barrier (NB) dataset. Comparison experiments with other methods are then conducted on the NB dataset and two public anomaly-detection datasets to evaluate the superiority of UADNet. Finally, the real-time detection capability of UADNet and the performance of the overall system are discussed.

4.1 Dataset and Evaluation Standard

An NB dataset is built using UAV images captured by an HD camera mounted on a DJI M300 RTK drone flying 20–50 m above the track at 2–3 m/s. The H20T camera ensures that the NB occupies the main field of view. The original image resolution is 5472×3078 , and representative examples are shown in Fig. 7. In addition to the main split, we performed a follow-up evaluation on an expanded anomaly test set containing 54 real anomalous images collected from additional sorties, another railway section, and different illumination and weather conditions. This supplementary set was used only for validation and testing. Comparison experiments on the public MVTec AD [18] and VisA [19] datasets were also conducted to verify the effectiveness of UADNet.

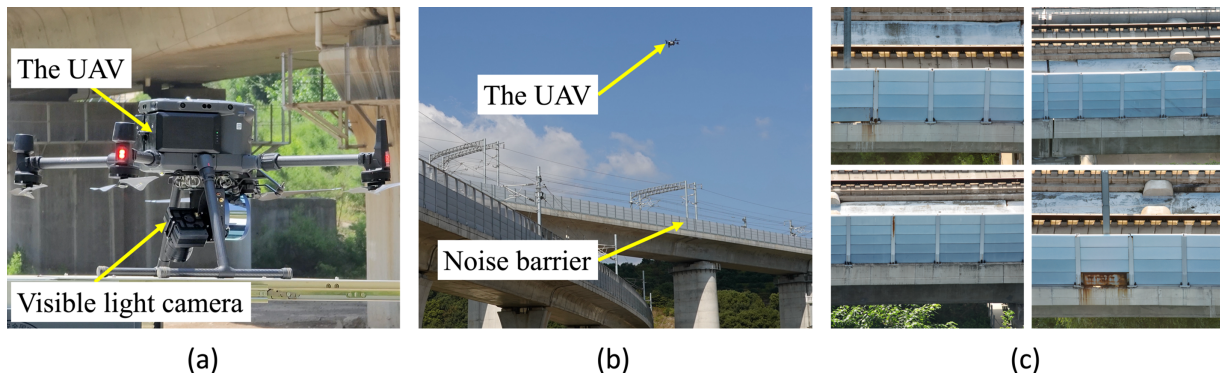


Figure 7: On-site data collection. (a) UAV equipment. (b) Flight experimental environment. (c) Samples of NB data.

To evaluate the anomaly-detection performance, image-level AU-ROC, pixel-level AU-ROC, and pixel-level AU-PRO are utilized. AU-ROC calculates the area under the receiver operating characteristic curve but may overestimate performance on highly imbalanced datasets [20]. Pixel-level AU-PRO, in contrast, weights each defect region more evenly by calculating the area under the per-region overlap curve, making it more suitable for small and sparse anomalies. In practical engineering inspection, missed detections may directly affect operational safety, whereas excessive false alarms increase the burden of manual re-checking. For this reason, AU-PRO is emphasized in our analysis because it better reflects region-wise localization quality on real defect areas, while AU-ROC is reported for completeness. Metric definitions are provided in (4) to (8).

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (5)$$

$$\text{AU-ROC} = \int_0^1 \text{TPRd}(\text{FPR}) \quad (6)$$

$$\text{PRO} = \frac{1}{N} \sum_i \sum_k \frac{|P_i \cap C_{i,k}|}{|C_{i,k}|} \quad (7)$$

$$\text{AU} - \text{PRO} = \int_0^{0.3} \text{PRO} \, d(\text{FPR}) \quad (8)$$

where an image or a pixel can be classified as either a true positive (TP), false positive (FP), true negative (TN), or false negative (FN). $C_{i,k}$ represents the set of pixels identified as anomalous for a connected component k in the ground truth image i . P_i denotes the set of pixels predicted as anomalous for a threshold t . N represents the total number of ground truth components in the evaluated dataset.

4.2 Training Details

The NB dataset, consisting of 1000 images, is randomly divided into training, validation, and test sets in a 7:2:1 ratio. Notably, the training set contains only normal NB images, while the validation and test sets include both normal and anomalous images in a 7:3 ratio. The MVTec AD and VisA datasets follow the same proportional allocation.

In the ablation experiments, all parameters are set identically for a fair comparison. Input images are resized to 256×256 , with 250 training iterations and batch size of 2. The learning rate is set at 0.0001. In the original benchmark comparison, all models use the default structure in ANOMALIB [21]. To further examine fairness, additional comparison runs were also conducted under a unified configuration with identical input resolution (256×256), preprocessing, optimizer type, and training budget for all methods.

4.3 Ablation Study

In this section, a series of ablation experiments with the classic student-teacher (S-T) structure of PDN are conducted on the customized NB dataset to validate the superiority of the UADNet framework equipped with GPDN (ghost patch description network) and SCAE (skip connection autoencoder). The quantitative metrics include image-level AU-ROC, pixel-level AU-PRO, pixel-level AU-ROC, GFLOPs and parameters. In addition to the structural-combination and weighting-factor studies, we also examine the individual contributions of GhostConv and SCAE in a more fine-grained manner.

- (1) **Ablation on UADNet Framework Structure:** Fig. 8 shows the results of the ablation experiments with different UADNet network structures. All UADNet variants use a teacher-student-autoencoder framework, with PDN structures pretrained on a large public dataset as the teacher networks. The PDN + GPDN + SCAE combination achieves the best performance in image-level AU-ROC, pixel-level AU-PRO, pixel-level AU-ROC, and training loss compared with the other frameworks. For instance, as shown by the green and orange lines in Fig. 8, PDN + GPDN + AE outperforms PDN + PDN + AE in pixel-level AU-PRO. Specifically, the GPDN equipped with the Ghost module improves AU-PRO from 80.0% to 83.6% (Table 4). On the basis of PDN and GPDN, the use of SCAE still increases AU-PRO from 83.6% to 86.7%, demonstrating the superiority of the proposed SCAE. Regarding FLOPs, PDN + GPDN + SCAE outperforms PDN + PDN + AE, PDN + GPDN + AE, and PDN + PDN + SCAE by 12.48, 10.2, and 2.26, respectively (Table 4). In terms of parameters, UADNet with GPDN and SCAE has the fewest parameters and the best performance.

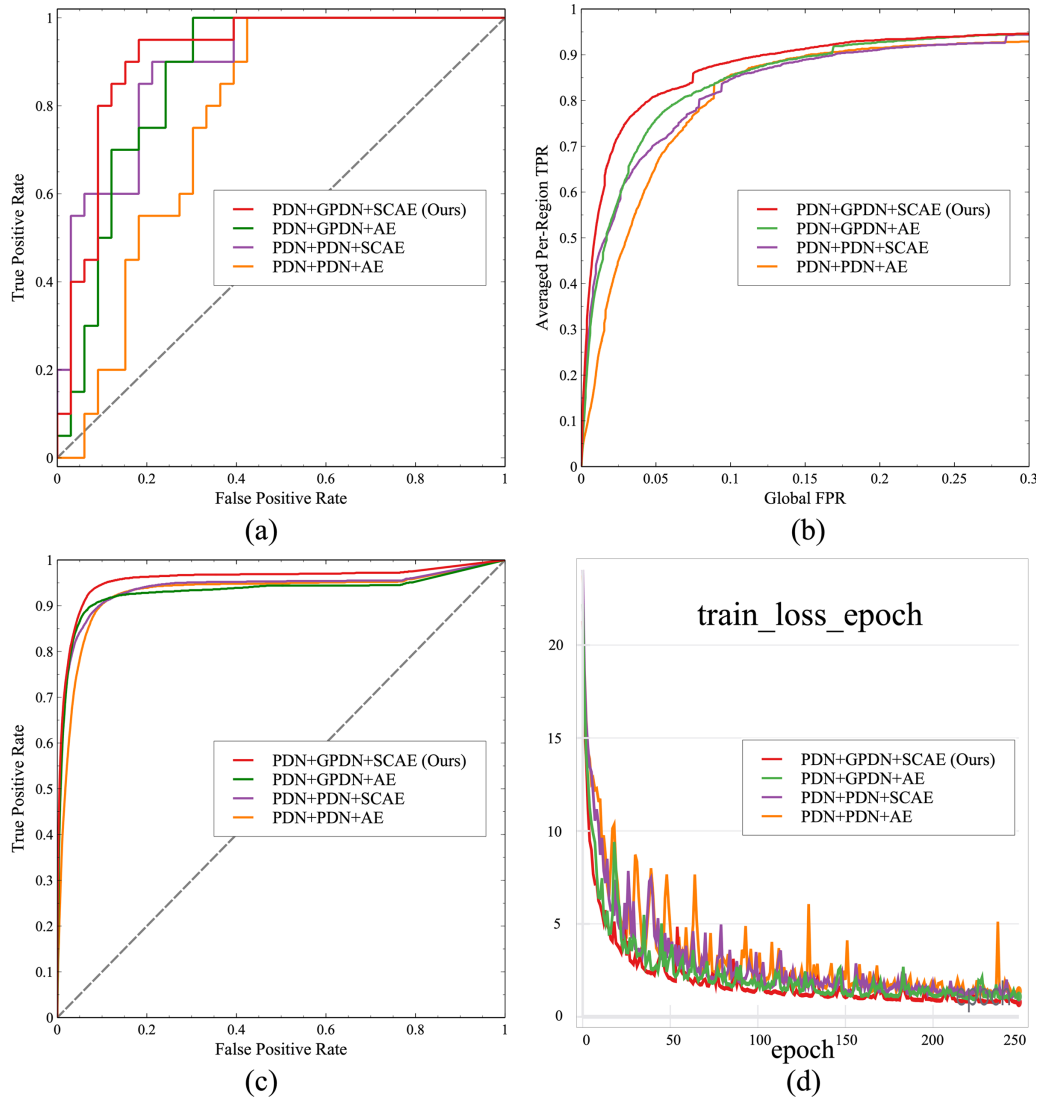


Figure 8: Score curves of different modules. (a) Image-level AU-ROC. (b) Pixel-level AU-PRO. (c) Pixel-level AU-ROC. (d) Training loss.

Table 4: Quantitative metrics of the models in the ablation study.

Structure Teacher + Student + AE	Image AU-ROC	Pixel AU-PRO	Pixel AU-ROC	FLOPs (109)	Params (106)
PDN + PDN + AE	77.2	80.0	92.7	76.12	8.06
PDN + PDN + SCAE	89.0	82.2	93.9	73.84	7.8
PDN + GPDN + AE	86.9	83.6	93.1	65.9	7.2
PDN + GPDN + SCAE	91.3	86.7	95.7	63.64	7.0

To further isolate the contribution of the key modules, we additionally tested three targeted variants: replacing GhostConv with standard convolution in the student, removing the SCAE branch, and using the teacher-student or SCAE branch independently. Replacing GhostConv reduced pixel-level AU-PRO

by 2.1 percentage points, removing SCAE reduced pixel-level AU-PRO by 3.1 points, and using either branch alone led to a larger performance drop. These results confirm that GhostConv and SCAE provide complementary benefits rather than acting as simple lightweight substitutions.

We also examined the importance of the stage-1 preprocessing step by applying the anomaly detector directly to the uncropped UAV frames. Without NB localization, pixel-level AU-PRO decreased by 5.4 percentage points and the anomaly maps produced substantially more false alarms in sky and track regions. This confirms that the color-prior localization stage is a practical and necessary component of the overall system in UAV inspection scenes.

- (2) **Ablation on Anomaly Map Scale Factors:** In UADNet, two anomaly maps are generated: the local anomaly map (LAM) from the disparity between the S-T network's feature maps, and the global anomaly map (GAM) from the dissimilarity between the SCAE and student network's feature maps. Additional experiments on the scale factors of these maps are necessary to optimize model effectiveness. As depicted in Table 5, experiments are conducted for various scale factors of LAM and GAM, including using only the S-T network branch (1, 0) and only the SCAE and student branch (0, 1). With scale factors at (0.5, 0.5), UADNet achieves the optimal results, establishing this as the final parameter choice.

Table 5: Experiments on LAM and GAM with different scale factors.

The Scale Factors of LAM and GAM	Image-Level AU-ROC (%)	Pixel-Level AU-PRO (%)	Pixel-Level AU-ROC (%)
(1, 0)	87.8	86.4	93.4
(0.9, 0.1)	91.1	86.3	93.6
(0.8, 0.2)	89.2	85.6	93.4
(0.7, 0.3)	87.5	86.4	94.6
(0.6, 0.4)	90.0	86.2	94.5
(0.5, 0.5)	91.3	86.7	95.7
(0.4, 0.6)	89.0	82.5	93.8
(0.3, 0.7)	86.0	81.1	92.5
(0.2, 0.8)	84.4	73.1	90.8
(0.1, 0.9)	76.5	58.3	87.2
(0, 1)	69.3	39.7	74.8

4.4 Comparative Study and Additional Discussion

UADNet is compared with STFPM [22], RevDist [11], DFM [23], PatchCore [3], CFA [24], CFLOW [25], and DRAEM [26] on the public and NB datasets to comprehensively evaluate its performance. In addition to the default-setting comparison, the revised study also verified the ranking under a unified training configuration, and the superiority of UADNet remained consistent. This indicates that the observed advantage is not caused by different default optimization settings alone.

A cross-condition evaluation was further conducted by training on the main normal set and testing on the supplementary set collected from another railway section and different weather conditions. Under this moderate line-shift and environment-shift setting, UADNet maintained 88.9% image-level AU-ROC, 82.1% pixel-level AU-PRO, and 94.6% pixel-level AU-ROC, indicating reasonable generalization beyond the original acquisition condition.

As shown in Table 6, the models almost achieve excellent results on the MVTec AD dataset [18]. UADNet achieves the highest pixel-level AU-PRO (95.1%) and pixel-level AU-ROC (98.4%). On the more challenging VisA dataset [19], only UADNet achieves a score above 90% on pixel-level AU-PRO, highlighting its advantage in detecting anomalies within small targets. Under the unified training configuration, UADNet still maintained the best pixel-level AU-PRO on the NB dataset, confirming the fairness of the overall comparison.

Table 6: Comparative analysis of anomaly detection performance between UADNet and SOTA models on public and NB datasets.

Model	Params (106)	FLOPs (109)	MVTec AD/VisA			Noise Barrier Dataset		
			Image-Level AU-ROC (%)	Pixel-Level AU-PRO (%)	Pixel-Level AU-ROC (%)	Image-Level AU-ROC (%)	Pixel-Level AU-PRO (%)	Pixel-Level AU-ROC (%)
STFPM	23.5	21.6	97.3/89.6	94.4/89.6	97.3/97.5	60.7	77.5	94.3
RevDist	80.6	67.9	88.2/83.2	94.8/89.9	98.3/97.0	56.2	70.5	95.3
DFM	8.54	8.7	93.2/85.6	88.2/80.0	97.1/96.4	92.1	47.4	81.0
PatchCore	24.9	24.1	97.4/82.3	92.4/73.6	98.1/94.7	82.8	72.8	95.2
CFA	31.3	50.5	93.1/82.5	88.6/81.7	97.0/96.8	69.2	64.3	92.6
CFLOW	235.7	230.2	95.7/85.3	92.1/83.1	98.0/95.4	82.2	47.2	86.0
DRAEM	97.4	397.2	71.8/66.4	49.1/50.0	75.5/71.9	84.5	67.2	84.6
EfficientAD	8.1	76.1	96.8/88.5	90.7/88.3	96.5/96.6	77.2	80.0	92.7
UADNet	7	63.6	97.1/89.0	95.1/90.5	98.4/97.1	91.3	86.7	95.7

Considering the more practically relevant NB dataset in railway engineering, models such as STFPM, PatchCore and EfficientAD, which perform well on public datasets, do not achieve equally satisfactory results. The main reason is that UAV-based NB imagery contains oblique viewpoints, motion blur, illumination variation, repeated textures, multi-scale defects, and large non-target backgrounds. Memory-bank or flow-based methods are more easily distracted by repetitive background patches and scale variation, whereas reconstruction-only methods may smooth out small corrosion or edge defects. For the stringent pixel-level AU-PRO metric, UADNet is the only model that scores above 80%, demonstrating its precise anomaly localization performance. Specifically, our model outperforms STFPM, RevDist, DFM, PatchCore, CFA, CFLOW, DRAEM, and EfficientAD by 9.2%, 16.2%, 39.3%, 13.9%, 22.4%, 39.5%, 19.5%, and 6.7%, respectively, in pixel-level AU-PRO.

The visual detection effects of SOTA models are shown in Fig. 9. For large anomalies in the MVTec AD dataset, most models exhibit effective recognition. However, for small anomalies in the VisA dataset, models like DFM, PatchCore, CFA, CFLOW, and DRAEM show severe missed detections. The NB dataset is even more challenging because the anomalies are often small, multi-scale, and embedded in cluttered UAV scenes. As shown in Fig. 9, the visual detection results of the proposed model are the closest to the ground truth. Compared with other SOTA models, UADNet provides more accurate anomaly localization and significantly reduces missed and false detections in practical railway applications. Additional qualitative gradient-based feature-response inspection on representative NB defect cases also indicated that the dominant responses were concentrated on corroded joints, damaged panel edges, missing fasteners, and other defect-related regions, whereas repeated sky/track backgrounds generally produced weak or scattered responses.

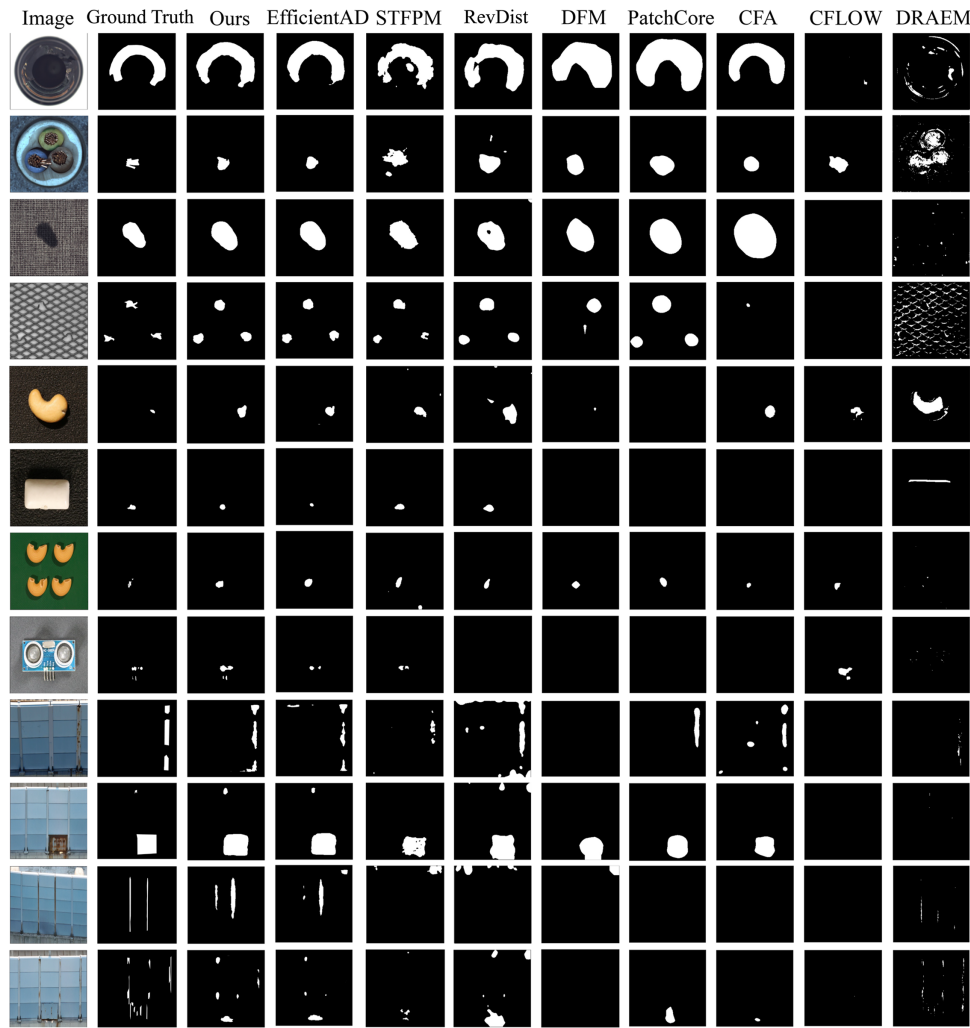


Figure 9: Visual detection effect of SOTA models on the MVTec AD (lines 1–4), VisA (lines 5–8) and noise barrier datasets (lines 9–12).

4.5 Inference Time on NVIDIA Embedded Board

To demonstrate the computational speed capabilities of our method, comparative experiments are conducted on a single NVIDIA TITAN Xp GPU and a single NVIDIA Jetson Orin NX embedded device. Table 7 presents the inference time of the networks on TITAN Xp and Orin NX, respectively. In this setup, we evaluate the inference speed of all models at various resolutions.

From the NVIDIA TITAN Xp experiment, only CFLOW and DRAEM exhibit slower inference speeds, with respective times of 13.5 FPS and 25.1 FPS at a resolution of 256×256 . UADNet achieves the highest real-time detection frame rates across all image resolutions, achieving 104.1 FPS at the default resolution of 256×256 .

In the NVIDIA Jetson Orin NX experiment, UADNet exhibits a significant advantage in inference at almost all low-resolution settings. Considering both detection accuracy and speed, UADNet has better performance than the other SOTA models. With an end-to-end speed of 10.8 FPS for the complete system on Orin NX, the method is suitable for practical UAV inspection pre-screening, in which suspicious frames are flagged online and then confirmed offline by maintenance personnel.

Table 7: Comparison of inference time on Titan Xp and Jetson Orin NX.

Model	TITAN Xp/Jetson Orin NX (fps)				
	256	320	512	640	768
STFPM	95.1/19.3	62.6/13.8	26.8/6.1	15.9/3.9	11.0/2.7
RevDist	67.7/17.1	48.4/12.7	24.9/6.5	17.0/4.3	12.5/2.9
DFM	81.3/19.9	53.3/14.5	29.2/6.6	21.1/4.5	12.0/3.1
PatchCore	74.5/16.0	50.0/11.3	26.5/5.2	17.9/3.3	8.6/2.4
CFA	71.8/13.2	47.5/9.2	23.9/4.1	15.8/2.6	7.4/1.9
CFLOW	13.5/2.7	8.7/2.2	3.6/0.8	2.4/0.5	1.7/0.4
DRAEM	25.1/5.4	15.5/3.5	6.6/1.7	4.3/0.8	3.0/0.7
EfficientAD	85.5/17.7	53.1/13.0	20.7/5.4	11.6/3.6	8.1/2.5
UADNet	104.1/21.4	65.5/14.9	34.5/7.3	23.0/4.4	16.0/3.1

4.6 Entire System Evaluation

As illustrated in Table 8, to assess the entire system performance, we calculate the inference speed and quantitative metrics for both the NB localization algorithm (stage 1) and UADNet (stage 2). An accurate NB localization algorithm is crucial for detecting anomalies in noise barriers. $\bar{R}$ is utilized as the evaluation metric for stage 1.

$$\bar{R} = \frac{TP}{NP} \times 100\% \quad (9)$$

where TP and NP represent the number of images accurately extracted from the NB region and the total number of images extracted, respectively. From the perspective of the entire system, the NB localization algorithm and UADNet not only maintain high recognition accuracy but also demonstrate rapid detection speeds, showcasing significant potential for real-time detection in railway NBs. The main false positives arise from repair patches, strong specular reflection, or newly replaced components that differ from the training distribution but are still structurally acceptable. In practice, such cases can be mitigated by periodically updating the normal-data pool and by suppressing isolated single-frame alarms during video inspection.

Table 8: Performance evaluation of the entire system.

Evaluation Index		TITAN Xp GPU	Jetson Orin NX
Stage 1	$\bar{R}$	98.6	98.6
	FPS	50.0	21.7
Stage 2	AU-PRO	86.7	86.6
	FPS	104.1	21.4
Entire system		33.8	10.8

4.7 Failure Cases and Detection Boundary

The proposed system is designed primarily for visible surface anomalies, including corrosion, rust, cracks, missing fasteners, local deformation, and panel damage. It does not directly detect hidden structural

deterioration behind the panel, internal fatigue, or loosened rear-side components that are invisible in UAV imagery.

Typical failure cases include extremely small anomalies occupying only a few pixels, severe motion blur at long range, strong shadow boundaries, and novel-normal regions such as repair patches or newly replaced hardware. Even in these cases, the fused anomaly map is generally more stable than either branch alone, but occasional false negatives and false positives still occur. These scenarios define the current practical detection boundary of the proposed system.

5 Conclusion and Future Work

This paper presents an unsupervised system for detecting anomalies in high-speed railway noise barriers (NBs). A new NB localization algorithm using HSV color differences is introduced to extract NB regions and reduce false detection rates in subsequent anomaly detection. The NB anomaly detection algorithm, UADNet, is trained on normal data to detect unknown anomalies. UADNet consists of a student-teacher network for local anomaly detection and a skip connection autoencoder for global reconstruction error through multi-scale feature fusion. Tested on multiple datasets as well as embedded hardware, the system outperforms other SOTA models, highlighting its practical potential.

This study is an early attempt to use unsupervised anomaly detection for railway NB inspection. Future research will expand cross-line and cross-weather data collection, develop more adaptive color-prior localization for barriers with different appearances, and further strengthen onboard edge deployment and false-alarm suppression strategies.

Acknowledgement: Not Applicable.

Funding Statement: This work was supported by the National Natural Science Foundation of China (No. 52502422).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Jing Cui; methodology, Jing Cui; software, Jing Cui; formal analysis, Jing Cui, Yixuan Geng; investigation, Jing Cui; resources, Jing Cui, Xue Yang; data curation, Jing Cui; writing—original draft preparation, Jing Cui; writing—review and editing, Jing Cui, Xue Yang; visualization, Jing Cui, Wanyin Shi; supervision, Yong Qin, Miao Guo; project administration, Yong Qin; funding acquisition, Yong Qin. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fang S, Yang L, Tang J, Guo W, Zeng C, Shao P. Visual measurement of lateral relative displacement of wheel-rail of high-speed train under earthquake. *Eng Struct.* 2024;305:117736. doi:10.1016/j.engstruct.2024.117736.
2. Cui J, Qin Y, Wu Y, Shao C, Yang H. Skip connection YOLO architecture for noise barrier defect detection using UAV-based images in high-speed railway. *IEEE Trans Intell Transp Syst.* 2023;24(11):12180–95. doi:10.1109/TITS.2023.3292934.
3. Roth K, Pemula L, Zepeda J, Scholkopf B, Brox T, Gehler P. Towards total recall in industrial anomaly detection. In: *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2022 Jun 18–24; New Orleans, LA, USA. p. 14298–308. doi:10.1109/cvpr52688.2022.01392.

4. Batzner K, Heckler L, König R. EfficientAD: accurate visual anomaly detection at millisecond-level latencies. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV); 2024 Jan 3–8; Waikoloa, HI, USA. p. 127–37. doi:10.1109/WACV57701.2024.00020.
5. Redmon J, Farhadi A. YOLOv3: an incremental improvement. arXiv:1804.02767. 2018. doi:10.48550/arXiv.1804.02767.
6. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. arXiv:1708.02002. 2017.
7. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014 Jun 23–28; Columbus, OH, USA. p. 580–7. doi:10.1109/cvpr.2014.81.
8. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28:1–9. doi:10.1109/tpami.2016.2577031.
9. Wu Y, Chen P, Qin Y, Qian Y, Xu F, Jia L. Automatic railroad track components inspection using hybrid deep learning framework. IEEE Trans Instrum Meas. 2023;72:5011415. doi:10.1109/TIM.2023.3265636.
10. Tong L, Wang Z, Jia L, Qin Y, Wei Y, Yang H, et al. Fully decoupled residual ConvNet for real-time railway scene parsing of UAV aerial images. IEEE Trans Intell Transp Syst. 2022;23(9):14806–19. doi:10.1109/tits.2021.3134318.
11. Deng H, Li X. Anomaly detection *via* reverse distillation from one-class embedding. In: Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 18–24; New Orleans, LA, USA. p. 9727–36. doi:10.1109/CVPR52688.2022.00951.
12. Somepalli G, Wu Y, Balaji Y, Vinzamuri B, Feizi S. Unsupervised anomaly detection with adversarial mirrored autoencoders. In: Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence; 2021 Jul 27–30; Virtual. p. 365–75.
13. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. In: Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV); 2023 Oct 1–6; Paris, France. p. 3992–4003. doi:10.1109/iccv51070.2023.00371.
14. Jeong J, Zou Y, Kim T, Zhang D, Ravichandran A, Dabeer O, et al. Zero-/few-shot anomaly classification and segmentation. In: Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2023 Jun 17–24; Vancouver, BC, Canada. p. 19606–16. doi:10.1109/CVPR52729.2023.01878.
15. Zhou Q, Pang G, Tian Y, He S, Chen J. AnomalyCLIP: object-agnostic prompt learning for zero-shot anomaly detection. In: Proceedings of the International Conference on Learning Representations; 2024 May 7–11; Vienna Austria.
16. Ganesan P, Rajini V, Sathish BS, Shaik KB. HSV color space based segmentation of region of interest in satellite images. In: Proceedings of the 2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT); 2014 Jul 10–11; Kanyakumari, India. p. 101–5.
17. Han K, Wang Y, Tian Q, Guo J, Xu C. GhostNet: more features from cheap operations. arXiv:1911.11907. 2020. doi:10.48550/arXiv.1911.11907.
18. Bergmann P, Fauser M, Sattlegger D, Steger C. MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. p. 9584–92. doi:10.1109/cvpr.2019.00982.
19. Zou Y, Jeong J, Pemula L, Zhang D, Dabeer O. SPot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: Computer vision—ECCV 2022; 2022 Oct 23–7; Tel Aviv, Israel. p. 392–408. doi:10.1007/978-3-031-20056-4_23.
20. Cook J, Ramadas V. When to consult precision-recall curves. Stata J Promot Commun Stat Stata. 2020;20(1):131–48. doi:10.1177/1536867x20909693.
21. Akcay S, Ameln D, Vaidya A, Lakshmanan B, Ahuja N, Genc U. Anomalib: a deep learning library for anomaly detection. In: Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP); 2022 Oct 16–19; Bordeaux, France. p. 1706–10.
22. Wang G, Han S, Ding E, Huang D. Student-teacher feature pyramid matching for anomaly detection. arXiv:2103.04257. 2021.

23. Ahuja NA, Ndiour I, Kalyanpur T, Tickoo O. Probabilistic modeling of deep features for out-of-distribution and adversarial detection. *arXiv:1909.11786*. 2019.
24. Lee S, Lee S, Song BC. CFA: coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*. 2022;10:78446–54. doi:10.1109/ACCESS.2022.3193699.
25. Gudovskiy D, Ishizaka S, Kozuka K. CFLOW-AD: real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; 2022 Jan 3–8; Waikoloa, HI, USA. p. 1819–28. doi:10.1109/WACV51458.2022.00188.
26. Zavrtanik V, Kristan M, Skocaj D. DRÆM—a discriminatively trained reconstruction embedding for surface anomaly detection. In: *Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*; 2021 Oct 10–17; Montreal, QC, Canada. p. 8310–9. doi:10.1109/iccv48922.2021.00822.