ITC 3/54
Information Technology and Control
Vol. 54 / No. 3/ 2025
pp. 768-777
DOI 10.5755/j01.itc.54.3.41460

Anormal Target Detection for Power Transmission Cable Drone Images Employing Improved YOLO10 and ESRGAN

Received 2025/05/10 | Accepted after revision 2025/07/13

HOW TO CITE: Zhao H., Meng, L., Jiao, Y., Liu, Z. (2025). Anormal Target Detection for Power Transmission Cable Drone Images Employing Improved YOLO10 and ESRGAN. *Information Technology and Control*, 54(3), 768-777. https://doi.org/10.5755/j01.itc.54.3.41460

Anormal Target Detection for Power Transmission Cable Drone Images Employing Improved YOLO10 and ESRGAN

Haosen Zhao
School of Energy and Power Engineering, Northeast Electric Power University Jilin, Jilin 132000, China; e-mail: 50501336@ncepu.edu.cn

Linghui Meng
Mechanical engineering and automation, Shenyang Institute of Technology School, Fushun, LiaoNing,113000, China; e-mail: 13811330386@163.com

Yuyang Jiao*
State Grid Beijing Power Cable Company, Beijing, 100022, China; e-mail: ncepu_ashur0925@sina.com

Zesan Liu
Information Industry Research Institute, State Grid Info & Telecom Group Co., Beijing, 101000 China; e-mail: 58735981@qq.com

Corresponding author: ncepu_ashur0925@sina.com

The anomaly target detection accuracy and speed in transmission cable drone inspection batch images with low pixel size is not enough high to make operation decision. To address this problem by recon-struction of image enhancing pixel, a super-resolution transmission cable target detection method based on improved YOLO10 is proposed. In this paper, we do the following: (1) the Best-Buddy loss function is used instead of the commonly used one-to-one MSE/MAE loss function, allowing low pixel images to dynamically find the most suitable supervised image, providing more reasonable image details and reducing training difficulty; (2) a region-aware adversarial learning strategy is introduced, focusing on training texture-rich image areas, enriching the texture structure of images, making them more realistic, reducing artifacts, and improving their

visual effects; (3) the feature-driven super-resolution algorithm is applied to the object detection of power transmission lines, using YOLO10's backbone as a feature extractor and fine-tuning the super-resolution model to generate more machine-readable feature maps to improve detection accuracy and speed. The experimental results show that the proposed model improves detection accuracy by 6.62% compared to using only the YOLO10 object detection model and 1.22% compared to using ESRGAN (Enhanced Super-Resolution Generative Adversarial Networks) as image pre-processing before using YOLO7 for object detection, achieving a detection accuracy of 92.55%.

KEYWORDS: Feature-driven; Super-resolution algorithm; GAN; Power transmission cable; Machine learning

1. Introduction

Under severe weather conditions such as wind, snow, and ice, imaginary power transmission cables can cause frequent vibra-tions and movements, which can lead to ma-terial fatigue and accidents such as wire breakage or damage. Attaching vibration dampers to the suspension points of power transmission cable towers can absorb or reduce wire amplitude, change the frequency of line sway, prevent line vibration or movement, and are a key factor in extending the service life of power transmission cables. However, due to the long-term exposure of vibration dampers to harsh out-door conditions, they are prone to malfunctions, making it extremely important to identify abnormal dampers during power transmission cable inspections. Traditional manual inspections not only have low inspection efficiency but also cannot guarantee the safety of inspection personnel in complex environments. With the continuous development of artificial intelligence, un-manned aerial vehicle (UAV) inspections, represented by drones, have gradually re-placed traditional manual inspections, with higher efficiency and greater safety in the inspection process. However, due to the poor controllability of drones in special weather conditions, if they do not maintain a safe dis-tance from the line equipment during flight, they are prone to accidentally collide with power lines, causing the drone to crash and power equipment to be damaged, or even causing a power outage accident. Therefore, the power grid management agency re-quires that drones maintain a certain distance from high-voltage power equipment, which leads to lower image resolution and smaller targets, increasing the difficulty of detecting small targets. The analysis of the detection results of UAV inspection images shows that the main-stream target detectors currently have poor capability of detection small targets in busy backgrounds. On the other hand, because small targets have fewer pixels in the original image due to their distance or limited carrying equipment, they lack relevant appearance information for detection, and their resolution and feature information gradually weaken after multiple down-sampling, increasing the difficulty of detection [6]. Therefore, it is ur-gently necessary to propose a method for de-tecting small targets in power transmission cables under complex backgrounds to improve detection efficiency [14].

A direct method to enhance recognition per-formance for tiny targets in object detection is to use super-resolution algorithms for image reconstruction. Super-resolution reconstruction algorithms can be used to reconstruct the entire image or small target images cropped from the image, improving pixel density in the target and enabling the detection model to extract richer semantic information from the small target images in the backbone stage for better prediction accuracy. To mitigate the diffi-culty of detecting small targets in low-resolution images under complex back-grounds, this study integrates image su-per-resolution reconstruction with object de-tection. The method introduces a region-aware adversarial learning strategy to enhance tex-ture details and applies the Best-Buddy loss function to improve the training efficiency and reconstruction quality by enabling flexible supervision.

The feature-driven super-resolution recon-struction algorithm combined with the YO-LO10 [12] model is proposed to solve the aforementioned problems in glass insulator detection under self-explosion. Our main contributions are as follows: (1) using the Best-Buddy loss [9] function instead of the standard used one-to-one MSE/MAE loss to prevent excessive smoothing and reduce ar-tifacts in

the generated images. (2) adopting a region-aware adversarial learning strategy [9] to make the model focus more on texture-rich regions during training, leading to more re-alistic texture effects in the generated images. (3) Fine-tuning the super-resolution model based on feature-driven techniques to gener-ate feature maps that are more understandable to the machine, reducing the differences be-tween the feature maps extracted from the super-resolution algorithmically created im-ages and the ground truth images, and im-proving the accuracy of the object detector.

2. Related Works

Traditional super-resolution approaches pre-dominantly rely on interpolation methods (e.g., Lanczos, non-uniform, and bicubic in-terpolation) for image upscaling. Despite their computational efficiency and implementation simplicity, these techniques generally pro-duce limited reconstruction performance.

Rely on deep learning methods [4], the SRCNN [3] super-resolution reconstruction algorithm was employed to reduce noise and blur in the original images, thereby opti-mizing and enhancing the dataset. Due to the presence of subpixel structures such as thin cables and insulators in aerial images, con-ventional SR methods like SRCNN, which rely on fixed-size convolutional kernels, often fail to reconstruct these fine-grained details.

Subsequently, they employed the YOLOv4 [2] detection algorithm for detecting self-explosion faults in glass insulators. However, it was found that SRCNN did not yield satisfactory results for reconstructing complex images [10]. To address this, dy-namically selected different super-resolution networks for reconstruction based on the images' diverse features, thereby achieving improved detection results. Authors in [15] introduced a feature texture transfer module that combined deep semantic features of low-resolution images and shallow area tex-tures of high-resolution reference features to reconstruct small target features and improve the detection capability of small targets. Alt-hough the above-improved object detection models have improved detection accuracy to a certain extent, they did not consider that the reconstructed images only focus on the sensory experience of human vision, while ignoring the machine's understanding of feature maps, which may cause unforeseen impacts in sub-sequent detection tasks. Moreover, during the generator model training phase, one-to-one MSE/MAE losses were used to force the map-ping transformation between the given low-resolution and high-resolution image pairs, which is difficult to train and may lead to excessive smoothing and generation of pseudo-artifacts. Unlike traditional su-per-resolution and detection approaches that focus primarily on visual enhancement or isolated module optimization, our proposed method integrates a feature-driven su-per-resolution network with a task-aware de-sign. This ensures that the reconstructed im-ages are not only visually improved but also better aligned with the feature representation required by the detection model, leading to superior performance in small object detection under complex conditions.

3. Super-resolution Network

3.1. ESRGAN Super-resolution model

Enhanced Super-Resolution Generative Ad-versarial Network (ESRGAN) [13] inherits the characteristics of the Super-Resolution Gen-erative Adversarial Network (SRGAN) [8] model. It is a deep neural network structure consisting of two networks, one opposing the other. One neural network generates its cor-responding high-resolution image based on the input LR (low-resolution image), which is called the generator. The other neural network is used to evaluate the authenticity of the input image, determining whether each High-definition imagery it reviews belongs to the actual training dataset, which is called the discriminator. The two neural networks are iteratively trained until the discriminator can no longer distinguish whether the input im-age is generated or real, which is considered as Nash equilibrium [11]. Finally, the generator model can generate high-resolution images that are indistinguishable from real ones.

ESRGAN uses Relativistic average GAN to learn "whether a picture is more real than others "rather than just "whether a picture is true or false" [1]. Ex-

perimental results demonstrate that the proposed enhancement enables the generator to reconstruct more authentic texture details.

3.1.1. Generator Model

Since the original BN (Batch Normalization) [7] layer relies on the mean and variance pa-rameters of the batch data when regularizing the features, but mainly relies on the esti-mated mean and variance parameters of the test set during the testing process, the BN layer introduces artifact interference and limits the generalization ability of the model when the data of both are widely different. Thus, the generator model of ESRGAN removes the BN layer to obtain a stable and consistent training effect. Based on the architectural design of SRGAN, ESRGAN introduces a deep residual dense connection [5] to propose RRDB (Re-sidual in Residual Dense Block) to replace the conventional fundamental block. It combines the multi-level residual network and dense connection with a deeper and more complex structure, which can ex-tract more effective features.

3.1.2. Discriminator Model

The network model of the ESGRAN dis-criminator uses VGG-19. The output value of SRGAN discriminator D is the probability that the reconstructed image is consistent with the real image, while ESGRAN is effectively improved by referring to Relativistic GAN, and the output value of discriminator D is the chance that the real image is more inclined to the real image than the false image. The probability of the discriminator in SRGAN is calculated as follows:

$$\begin{aligned} D(x_r) &= \sigma\left(C(x_r)\right) \to 1 \\ D(x_f) &= \sigma\left(C(x_f)\right) \to 0 \end{aligned}. \tag{1}$$

Further, the discriminator D probability of Relativistic GAN is calculated as follows:

$$D_{Ra}(x_r, x_f) = \sigma\left(C(x_r) - E_{x_f}\left[C(x_f)\right]\right) \to 1 \tag{2}$$

$$D_{Rb}(x_f, x_r) = \sigma\left(C(x_f) - E_{x_r}\left[C(x_r)\right]\right) \to 0, \tag{3}$$

where represents the indicate that the dis-criminated image is more real than the false image, and represents the indicate that the discriminated image is falser than the true image. From this, the original discrimination mechanism is supplanted by the relative av-erage discriminator RaD, which is expressed as in the formula. The standard discriminator in SRGAN can be expressed as D(x)=σ((x)), where represents the Sigmoid function, rep-resents the discriminator output of the end transformation, represents the average oper-ation of all spurious data in a small batch, and the discriminator The loss function of the identifier is defined as follows:

$$\begin{aligned} &L_G^{Ro} - E_{x_r}\left[\log\left(1 - D_{R_o}(x_r, x_f)\right)\right] \\ &- E_{x_f}\left[\log\left(D_{R_o}(x_f, x_r)\right)\right] \end{aligned}. \tag{4}$$

3.2. Improving ESRGAN

3.2.1. Best-Buddy Loss

A single image super-resolution reconstruc-tion is essentially an ill-posed problem with multiple solutions, so a low-resolution patch should correspond to multiple high-resolution super-vised patches while all current models use the common one-to-one MSE/MAE loss, which forces a mapping between a given low-resolution and high-resolution image pair during the training phase. This strictly limits the generated HR images, making it very difficult and un-reasonable to train the net-work, and the reconstructed HR images may lack several high-frequency structures. The Best-Buddy loss function improves the one-to-one MAE loss function by exploiting the self-similarity prevalent in natural images to find and use a suitable HR patch as a super-vised patch in high-definition images in a flexible manner patch.

For the patch to be evaluated, it is necessary to find the supervised patch (i. e., the best partner) corresponding to it in the current it-eration, which needs to satisfy two constraints: (1) must remain proximal to the prespecified real patch in the HR space, and it is likely to find an HR patch close to the real patch among the candidate patches of the real HR image. (2) To make the subsequent optimization easier, should be close to the 2nd term in Equation (7). Since the generator model uses a pre-trained model, is a reasonable prediction patch. If satisfying these two constraints, then is a reasonable HR supervised patch.

First, the real high-resolution images are down sampled using different scales:

$$I_{HR}^{r} = S(I_{HR}, r), r = \{2,4\}, \quad (5)$$

where is performing bicubic downsampling. Image downsampling results in a 3-layer image pyramid containing the original GT images at different resolutions. The HR image to be evaluated and the corresponding GT image pyramid are unfolded into n*n patches, and the GT image unfolded patches are used as the supervised candidate database G. For the i-th to be estimated.

The patch is noted as, which is not supervised by the predefined real patch and needs to find the best partner to supervise in the current iteration:

$$g_i^* = \arg\min_{g \epsilon G} \alpha ||g - g_i||_2^2 + \beta ||g - \hat{p}_i||_2^2, \quad (6)$$

where ≥ 0 and ≥ 0 are scaling parameters. The best buddy loss for this patch is expressed as:

$$L_{\mathrm{BB}}(\hat{p}_i - g_i^*) = ||\hat{p}_i - g_i^*||_1. \quad (7)$$

When ≪, the best partner loss function de-generates to a typical one-to-one MAE loss.

3.2.2. Region-aware Adversarial Learning Strategy

Images reconstructed using ESGRAN some-times produce artifacts, especially in smooth regions. Therefore, region-aware adversarial learning is used to distinguish texture-rich regions from smooth regions based on local pixel statistics, and only texture-rich content is provided to the discriminator since images in smooth regions are easy to train without ad-versarial training.

First, the real HR image (i. e., $\mathbf{I}_{\mathrm{HR}} \epsilon R^{H_S \times W_S}$) is expanded into block B($\mathbf{B} \epsilon R^{H_S \times W_S \times k^2}$) of size, and then the standard deviation of each block is calculated. Then, the binary masks are obtained according to Equation (9) as follows:

$$M_{i,j} = \begin{cases} 1, & \mathrm{std}(\mathbf{B}_{i,j}) \geq \delta \\ 0, & \mathrm{std}(\mathbf{B}_{i,j}) < \delta \end{cases}, \quad (8)$$

where is the predefined threshold and is the location of the patch. Texture-rich regions are marked as 1, while texture-smoothed regions are marked as 0. The evaluation results and ground truth are then multiplied with the same mask M to obtain and, which are then computed by the discriminator. Although more arithmetic power is consumed by using the complex training method, the results show that the model can generate more realistic texture effects.

4. Methods

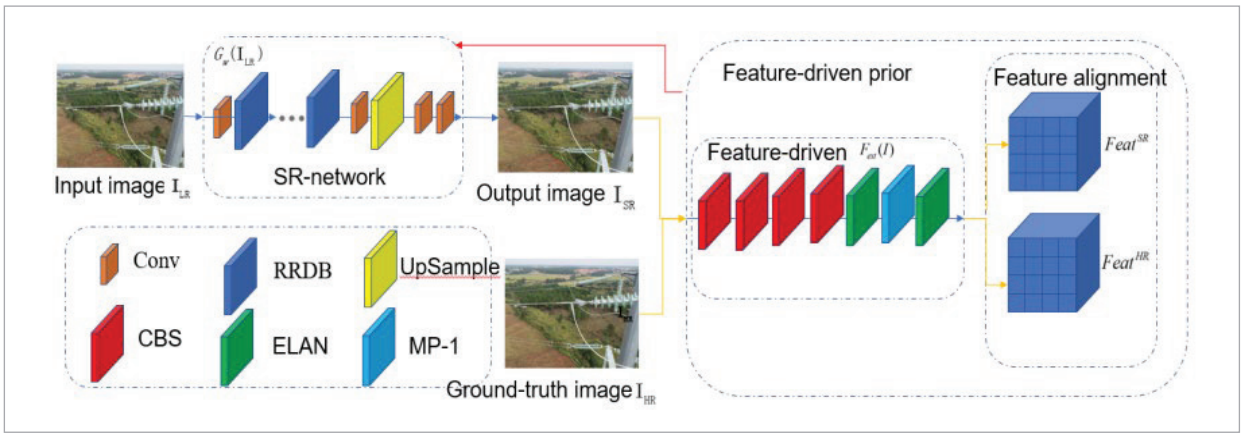
4.1. YOLO10 Network

The YOLO algorithm series constitutes the most characteristic implementation of sin-gle-stage object detection frameworks, em-ploying deep neural networks for simulta-neous object recognition and spatial localiza-tion. They run very fast and can be used in real-time systems. The network structure of YO-LO10 can be classified into three parts: Backbone, Neck, and Head. Backbone is the backbone network, which is the basic feature extraction layer responsible for extracting features of different scales in the image, with rich high-level semantic information and rich bottom-level structural positioning infor-mation. Neck is the feature fusion layer, which is responsible for fusing different scale feature maps together. Head is the prediction layer, which is responsible for processing the output of the Neck part and making predictions. YOLO10 first performs Mosaic data augmen-tation, adaptive anchor box calculation, and image scaling on the image, and then pre-processes the image into an RGB image of size 640*640 and inputs it into the Backbone to extract different sizes of features. The feature maps extracted at different stages are then input into PAFPN in the Neck part for full fusion of the top-level semantic features and bottom-level structural positioning features in the Backbone, and parameter aggregation of different detection layers in different back-bone layers. Finally, regression prediction is per-formed on targets of different sizes, namely large, medium, and small.

4.2. Feature-driven Su-per-resolution-based Recon-struction Network with Target Detector

Although the convolutional neural net-work-based super-resolution generative ad-versarial network

Figure 1
Feature-driven super-resolution network structure.



algorithms currently achieve good visual results in image recon-struction, these models are too focused on human visual perception and ignore the spe-cific needs of downstream vision tasks in-cluding object detection, image segmentation, and recognition. The generated feature maps of super-resolution reconstruction may differ from those of the real images after multiple convolutions, which makes it difficult for the model to understand and affects accuracy.

As shown in Figure 1, to balance the needs of subsequent tasks and improve detection performance, we use a trained object detection backbone network as a feature extractor to make the reconstructed HR image features with a similarity of possible to those of the real HR image features. The feature-driven su-per-resolution reconstruction network reduces the excessive coupling be-tween the feature extractor and the subsequent detection net-work, which helps enhance the practical utility of synthesized imagery in subsequent detec-tion networks. Mean squared error (MSE) is applied to evaluate the feature differences between the HR image extracted by the backbone net-work and the reconstructed image:

$$L_{MSE}^{Feat}=\frac{1}{N}\sum_{i=1}^{N}\left\| F_{ext}\left(G_w\left(I_i^{LR}\right)\right)-F_{ext}\left(I_i^{HR}\right)\right\|_F, \quad (9)$$

where $F_{ext}\left(G_w\left(I_i^{LR}\right)\right)$ is the feature of the reconstructed image, $F_{ext}\left(I_i^{HR}\right)$ is the char-acteristic attribute of the reference image, denotes the Frobenius parametrization, and is the index of the image.

4.3. Other Loss Functions

Perceptual loss: ESRGAN enhances the visual perceptual quality by minimizing the fea-ture-space discrepancy rather than pixel-level divergence. The perceptual loss is:

$$L_P=\sum_i \eta_i \left\| \varphi_i\left(\hat{I}_{HR}\right)-\varphi_i\left(I_{HR}\right)\right\|_1, \quad (10)$$

The loss function of the final generator net-work as a whole is:

$$L=\lambda_1 L_{BB}+\lambda_2 L_P+\lambda_3 L_{MSE}^{Feat}, \quad \text{where, } \lambda_1=0.1, \lambda_2=1.0, \lambda_3=1.0 \quad (11)$$

5. Experiments

The experimental environment is Windows 10 operating system, Python 3.7, Pytorch 1.12.1, and the GPU is NVIDIA GeForce RTX 2080TI, and the detailed configuration is shown in Table 1.

5.1.1. Model Evaluation Indicators

The effect of model improvement was evalu-ated by comparing several metrics before and after model improvement in the same ex-perimental setting. The performance of the target detection model was verified using four metrics, precision (precision) P, recall (recall) R, average precision (average preci-

Table 1
Experimental platform configuration.

CPU	Memory/GB	GPU	System	CUDA
Intel Xeon Gold 5120	16	NVIDIA GeForce RTX 2080Ti	Win 11	CUDA-11

sion, AP), and frames per second (FPS) F, which are expressed as follows:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\% \tag{12}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\% \tag{13}$$

$$A_{AP} = \int_0^1 P(R)\mathrm{d}R, \tag{14}$$

where N_{TP} indicates the number of accurately detected hammers; N_{FP} indicates the number of unlabeled but detected error samples; N_{FN} indicates the number of labeled hammers not identified, i.e. the number of missed detec-tions; the accuracy P indicates the probability of correct prediction among the detected hammers; and the recall R indicates the probability that a certain hammer is detected. In the P-R curve, the area enclosed by the P-R curve and the coordinate axis is equal to the size of the AP value. The mAP is obtained by averaging the AP values of all categories. The mAP is used in this experiment to evaluate the detection performance of the whole target detection network model.

5.1.2. Datasets

In this paper, we use a homemade power transmission cable dataset for the detection of anti-vibration hammers, which includes an-ti-vibration hammers of different sizes and angles. The power transmission cable dataset consists of 1184 power transmission cable in-spection images, and the image input scale is set to 640×640 during the training process. The clear and complete anti-vibration hammers in the images are labeled by using labelImg software, and the labeled dataset is randomly classified into a training set, test set, and validation set according to the ratio of 6:2:2.

5.1.3. Model Training

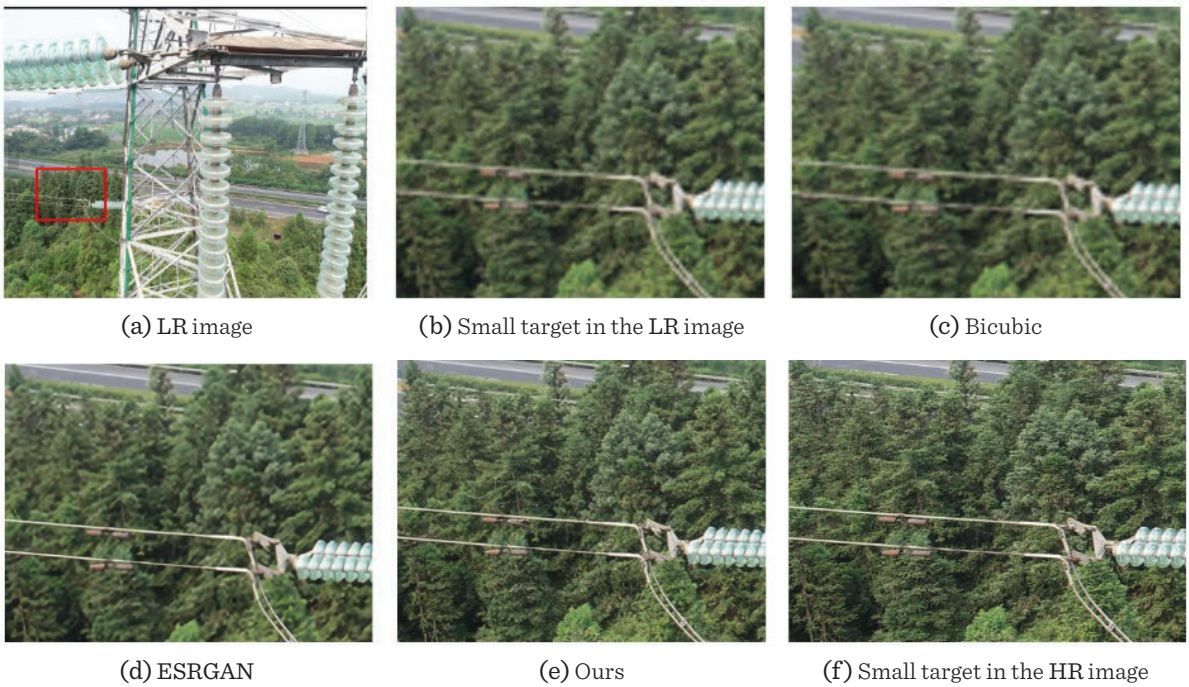
The training of this model is mainly divided into two parts: super-resolution (SR) network and feature extractor of the detection archi-tecture. In the stage of training the su-per-resolution model, the public pre-trained RRDB model is used. During the object de-tector training phase, we use a part of the YOLO10 backbone network as the feature extractor. First, we train the entire YOLO10 network to generate effective feature maps by the feature extractor, then we train the fea-ture-driven super-resolution network to gen-erate super-resolution feature maps that can pass off as real images, and finally, the sub-sequent YOLO10 object detection network is used to complete the detection of the shock absorber.

5.1.4. Analysis of Experimental Results

Five sets of experimental methods were used with the same test set and different model set-tings. Model a uses the ESRGAN su-per-resolution algorithm for image recon-struction; Model b adds region-aware adver-sarial learning strategy based on the ESRGAN model, supplanting the loss function with the Best-buddy loss function (generating su-per-resolution images instead of feature maps). Model c uses the standard YOLO10 object detector; Model d uses the ESRGAN super-resolution algorithm to improve image pixels before object detection; Model e uses our proposed feature-driven super-resolution al-gorithm for object detection. Traditional manual or low-accuracy detection methods often result in missed or false detections, leading to repeated inspections, increased labor costs, and potential operational risks. By significantly improving detection accura-cy—particularly for small and low-resolution targets—the proposed method reduces the frequency of manual verification and the likelihood of undetected faults.

The detection results of experiments a and b are shown in Figure 2. Although the shape features of the generated seismic hammer have not fully restored the structural features of the real image, the generated image has achieved good visual effects compared to the low-resolution image. In addition, the texture

Figure 2
(a) is a low-resolution image, (b) is the small target of the low-resolution image, (c) is the result of bicubic interpolation (d) is the image generated by the ESRGAN super-resolution model, (e) is the image generated by the improved super-resolution mod-el, and (f) is the real image.



(a) LR image (b) Small target in the LR image (c) Bicubic

(d) ESRGAN (e) Ours (f) Small target in the HR image

structure of the trees in the generated image is very close to that of the real image, and there is no unreasonable pseudo-shadow structure, which confirms the effectiveness of the re-gion-aware adversarial learning strategy. The results of experiments a and b show that by improving ESRGAN, the super-resolution model can generate images that are likely to real images, which confirms the effectiveness of the super-resolution model and provides a guarantee for subsequent feature-driven su-per-resolution object detection experiment.

The experimental results of super-resolution reconstruction are shown in Table 2:

Table 2
Experimental results of super-resolution reconstruction.

Method	PSNR (dB)	SSIM
Bicubic	22.36	0.6023
ESRGAN	23.78	0.6312
Ours	24.49	0.6948

The detection results of the target detection comparison experiment part are shown in Table 3 and Figure 3. The use of the su-per-resolution algorithm can enhance the detection accuracy of the target detector for small targets. It can be seen that the mAP value of model c is improved by 6.2% compared to model b without the su-per-resolution algorithm, and the target de-tection algorithm model based on the fea-ture-driven super-resolution algorithm com-bined with YOLO10 proposed in this paper has

Table 3
Target Detection Experiment Results.

Model	mAP@0.5 (%)	Precision (%)	Recall (%)
Model c (YOLO10)	85.53	86.83	85.36
Model d (ESRGAN + YOLO10)	91.33	90.76	91.56
Ours (Fea-ture-driven SR + YOLO10)	92.55	92.87	91.93

Figure 3

Comparison of detection results. As shown in Figure 3, (a) shows the target detection effect using only the YOLOv7 model; (b) shows the detection effect of the model in this paper. In (a), two anti-vibration hammers are not detected because they occupy too few pixels in the image, while in (b), the results demonstrate that the present model has good detection results for the small-scale anti-vibration hammers that were not detected before due to insufficient pixels. Most of the confidence levels exceed 80%.

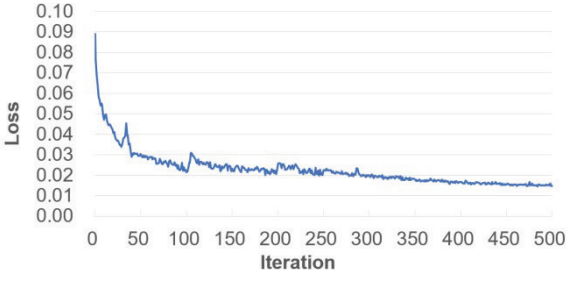
(a) (b)

an improved mAP value of 1.22% com-pared to the ESRGAN combined with YO-LO10 algorithm model.

The loss curve of the YOLO10 model is shown in Figure 4, where the horizontal axis repre-sents the number of iterations, and the vertical axis represents the loss value of the training process, which stabilizes at around 0.015 after 400 iterations.

Figure 4

YOLOv7 network loss curve.



6. Conclusions

A feature-driven super-resolution reconstruc-tion target detection algorithm is proposed to resolve the inefficiency of detection caused by small image pixels in UAV inspection. Com-bined with the experimental results, the analysis reveals are drawn: (1) Compared with the commonly used one-to-one MSE/MAE (Mean Squared Error/Mean Absolute Error) loss function, the one-to-many Best-Buddy loss function adopted in this paper can prevent the generated images from being over-smoothed and reduce artifacts. (2) The region-aware adversarial learning strategy allows the model to focus more on the texture-rich parts of the image and generate more realistic texture effects. (3) The use of an improved fea-ture-driven super-resolution target detection network can generate feature maps that are easy to understand by the machine and in-crease the predictive accuracy of the target detector. This paper verifies the usability of feature-driven super-resolution-based target detection in power transmission cable target detection, improves algorithm performance with high accuracy, and provides a new re-search method for future power transmission cable small target detection and defect detec-tion.

In future work, we aim to extend our approach to multi-frame super-resolution, which can leverage temporal redundancy to further enhance reconstruction quality. Additionally, exploring transformer-based attention mechanisms could help the model better focus on relevant small target regions.

Data Availability Statement

The data that support the findings of this study are available in power transmission cables at 10.6084/m9.figshare.25164839.

Conflicts of Interest

The authors declared no potential conflicts of interest with respect to the research, au-thor-ship, and publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Alexia, J.-M. The Relativistic Discriminator: A Key Element Missing from Standard GAN. arXiv Preprint, 2018, arXiv:1807.00734. doi:10.48550/arXiv.1807.00734.
2. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv Preprint, 2020, arXiv:2004.10934. doi:10.48550/arXiv.2004.10934.
3. Dong, C., Loy, C. C., He, K., Tang, X. Image Su-per-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(2), 295-307. doi:10.48550/arXiv.1501.00092. https://doi.org/10.1109/TPAMI.2015.2439281
4. Hao, S., Ma, R. Z., Zhao, X. S., Ma, X., Wen, H., An, B. Y. Self-Explosion Fault Detection Algorithm for Glass Insulator Based on Super-Resolution Deep Re-sidual Network. High Voltage Engineer-ing, 2022, 48(5), 1817-1825. doi:10.13336/j.1003-6520.hve.20210415.
5. He, K. M., Zhang, X. Y., Ren, S. Q., Sun, J. Deep Residual Learning for Image Recognition. Pro-ceed-ings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, 770-778. doi:10.48550/arXiv.1512.03385. https://doi.org/10.1109/CVPR.2016.90
6. Huang, Y. H., Liu, H. C., Qing, C., Chen, Z. Y., Zhang, J. R., Yang, C. R. Transmission Line In-sulator Fault Detection Method Based on USRNet and Im-proved YOLOv5x. High Voltage Engi-neering, 2022, 48(9), 3437-3446. doi:10.13336/j.1003-6520.hve.20220314.
7. Ioffe, S., Szegedy, C. Batch Normalization: Ac-celerating Deep Network Training by Reducing In-ternal Covariate Shift. Proceedings of the In-terna-tional Conference on Machine Learning, 2015, PMLR, 448-456. doi:10.48550/arXiv.1502.03167.
8. Ledig, C., Theis, L., Huszár, F., Caballe, J., Cun-ningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z. H., Shi, W. Z. Photo-Realistic Single Image Super-Resolution Using a Generative Adver-sarial Network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, 4681-4690. doi:10.48550/arXiv.1609.04802. https://doi.org/10.1109/CVPR.2017.19
9. Li, W. B., Zhou, K., Qi, L., Lu, L. Y., Jiang, N. J., Lu, J. B., Jia, J. Y. Best-Buddy GANs for Highly Detailed Image Super-Resolution. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2), 1412-1420. doi:10.1609/aaai.v36i2.20030. https://doi.org/10.1609/aaai.v36i2.20030
10. Liu, R. Z., Yan, G. W., He, H., An, Y. B., Wang, T., Huang, X. Small Targets Detection for Trans-mission Tower Based on SRGAN and Faster R-CNN. Recent Advances in Electrical & Electronic Engineering, 2021, 14(8), 812-825. https://doi.org/10.2174/2352096514666211026143543
11. Liu, Y., Zhu, L., Lim, K., Li, Y. H., Wang, F. P., Lu, J. Review and Prospect of Image Su-per-Resolution Technology. Journal of Frontiers of Computer Science and Technology, 2020, 14(2), 181-199. doi:10.3778/j.issn.1673-9418.1909057.
12. Wang, C. Y., Bochkovskiy, A., Liao, H. Y. M. YOLO10: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv Preprint, 2022, arXiv:2207.02696. doi:10.48550/arXiv.2207.02696.
13. Wang, X. T., Yu, K., Wu, S. X., Gu, J. J., Liu, Y. H., Dong, C., Qiao, Y., Loy, C. C. ESRGAN: En-hanced Super-Resolution Generative Adversarial Networks. Proceedings of the European Confer-ence on Com-puter Vision Workshops, 2018, 0-0. doi:10.48550/arXiv.1809.00219. https://doi.org/10.1007/978-3-030-11021-5_5
14. Wang, Y. Y., Li, W. D., Zhang, W. N., Li, X. J. Small Target Detection Algorithm Based on Adap-tive Feature Fusion. Journal of the Hebei Academy of Sciences, 2022, 39(6), 9-16. doi:10.16191/j.cnki.hbkx.2022.06.006.
15. Wang, Z. Z. Research on Small Object De-tection Method Under Super-Resolution Model. Yangtze University, 2022. doi:10.26981/d.cnki.gjhsc.2022.000603.