Signal, Image and Video Processing (2025) 19:227
https://doi.org/10.1007/s11760-024-03742-9

ORIGINAL PAPER

Research on perceptual methods of external damage hazards for transmission corridors

He Su[1] · Jiaomin Liu[1] · Zhenzhou Wang[2] · Pingping Yu[2] · Yuting Yan[2]

Received: 8 May 2024 / Revised: 10 November 2024 / Accepted: 23 November 2024 / Published online: 17 January 2025

Abstract

In the domain of power transmission, external damage hazards within transmission corridors significantly compromise the secure operation of the power grid. Therefore, it is of great importance to perceive the status and behavior. Given the complex environment surrounding transmission corridors, the types and scales of external damage hazards pose significant detection challenges. Thus, this study combined the improved YOLOv5s detection model with the lightweight StrongSORT tracking model for real-time perception of external damages. In the target detection phase, we construct the YOLO-CS-ASFF model based on the YOLOv5s architecture which uses the ConNeXt module with the fused SimAM attention mechanism to extract crucial features in complex backgrounds. The ASFF module enhances damage perception across different scales by optimizing feature fusion networks. Additionally, we use the SIoU loss function to improve the precision of external damage detection. In the target tracking phase, we optimize the appearance branching network of StrongSORT using the full-scale network (OSNet), which enhances the capability of StrongSORT to meet the real-time inspection requirements. Experimental results show that the improved YOLO-CS-ASFF achieved a mean Average Precision (mAP) and Recall of 92.8% and 85.3%, respectively, with an improvement of 3.24% and 1.32%. The StrongSORT tracking model attained tracking accuracy and precision of 63.3% and 78.9%, respectively, with a detection speed increase of 5.5 frames per second. The model effectively addresses the ID switching problem of obscured hazard targets, and improve the robustness of external breakage hazard tracking. The proposed method provides a technical reference for real-time perception of external breakage hazards in actual transmission corridors.

Keywords Transmission corridors · External damage hazards · YOLOv5s · StrongSORT · ConNeXt · OSNet

1 Introduction

As an essential component of power systems, the secure operation of transmission lines guarantees the safe of power supply [1]. In recent years, transmission corridors have been continuously threatened by sudden external damages, such as illegal construction activities, foreign objects hanging lines, vegetation growth [2]. Therefore, real-time detection and tracking of external breakage hazards in these corridors is crucial to prevent damage to overhead transmission lines. Despite advancements in deep learning, particularly in target detection and tracking, challenges remain in detecting and tracking these diverse external hazards.

In the context of visual detection in transmission corridors, Faster R-CNN and the YOLO series are commonly used object detectors [3]. Among them, the YOLO series object detector offers low computational cost and fast inference speed, meeting the real-time tracking requirements of high-demand scenes. Literature [4] proposes an improved YOLOv7 algorithm for long-distance foreign object recognition, focusing on enhancing the feature extraction network. By introducing two attention mechanisms, namely CBAM attention module and Swin Transformer self-attention module, the improved model effectively enhances the ability to detection ability of long-distance objects. Literature [5] primarily uses the YOLOv5 network to address issues of

✉ Zhenzhou Wang
chinawzz@163.com

1 Provincial and Ministerial Co-construction Collaborative Innovation Center on Reliability Technology of Electrical Products, Hebei University of Technology, TianJin 300401, China

2 School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China

large parameter quantities and slow detection speed in current foreign object detection networks. Literature [6] proposes an improved foreign object detection model based on YOLOv7-Tiny, which combines the SPPF module and SimAM attention mechanism to increase detection speed while maintaining accuracy. However, existing improvements in the YOLO series often focus on detecting single types of hazards or stationary objects, such as bird nests or hanging objects. However, in real-world transmission corridors, the hazards are varied, dynamic, and of different scales, especially construction vehicles. For example, literature [7] improves YOLOv7 to identify construction vehicles and foreign objects. Literature [8] uses the YOLOv5 model with the ASFF and CA modules to enhance detection accuracy and generalization. These models, while effective in improving detection accuracy, struggle in environments with diverse hazards in type, scale, and movement, necessitating robust real-time tracking solutions.

Object tracking methods, include DeepSORT [9], ByteTrack [10], and StrongSORT [11], which have been applied in various domains but were developed later than object detection algorithms. Literature [12] integrates the YOLOv4 and DeepSORT tracking algorithms to propose a tracking strategy for high-speed vehicles, achieving vehicle tracking and counting. Literature [13] introduces a multi-object tracking method utilizing optimized YOLOv5s and DeepSORT tracking algorithm. This method effectively enhances the feature extraction capability of vehicles and strengthens the tracking performance in complex environments. However, in the presence of similar background occlusions, the DeepSORT tracking algorithm still tends to exhibit frequent ID switches, leading to missed detections and false alarms.

Literature [14] applies StrongSORT for multi-vehicle tracking and employs the improved YOLOv5 algorithm as its detector with the CBAM attention mechanism. This algorithm effectively resolves missed detection and false alarm caused by occlusions. However, it encounters challenges related to larger model sizes and extended detection times. Furthermore, literature [15] leverages an improved YOLOv5 algorithm and ByteTrack to estimate vehicles. Comparative experiments demonstrated significant improvements in detection speed compared to StrongSORT, but at the cost of considerable loss in detection accuracy.

The literature review highlights the need for significant improvements in both detection and tracking algorithms to effectively perceive external damage hazards within the challenging environments of transmission corridors. To further optimize the perception of external damage in practical applications, this paper contributes in both detection and tracking as follows: On one hand, by improving the backbone network, feature fusion network, and loss function of YOLOv5s, the detection accuracy of the YOLOv5s model for multi-scale, multi-category external damage targets in complex backgrounds is improved. On the other hand, a lightweight and efficient OSNet network is utilized to enhance StrongSORT's feature extraction network, allowing the algorithm to meet real-time inspection requirements on equipment platforms. Figure 1 illustrates the optimized perception network framework. The YOLO-CS-ASFF represents the optimized YOLOv5s model: "YOLO" signifies YOLOv5s as the foundational framework, "CS" denotes the ConNeXt module integrated with the parameter-free attention mechanism (SimAM), and "ASFF" refers to the adaptive spatial feature fusion network.

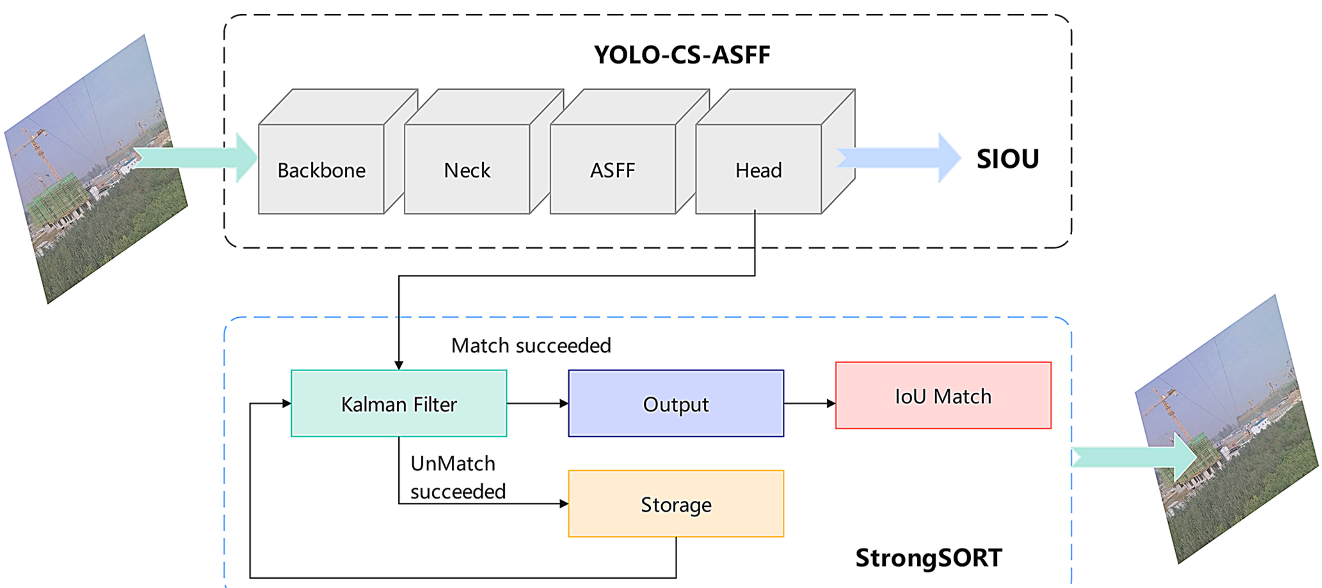

Fig. 1 Overall technology roadmap

2 YOLO-CS-ASFF

Selecting an appropriate detection model is a critical step in the task of dynamically inspecting transmission corridors using detection equipment. Given the stringent requirements for real-time performance and stability in transmission line scenarios, this paper adopts the lightweight YOLOv5s as the base network, and uses ConvNeXt to improve the object detection backbone network [16]. Additionally, it integrates the parameter-free attention mechanism (SimAM) [17] for enhancing the detection accuracy and speed of external break hazards. Considering the imbalanced distribution and large scale variations of external break targets, the paper proposes a multi-scale fusion approach through the ASFF module [18]. The optimized network model is shown in Fig. 2.

2.1 ConTeXt-SimAM

The original YOLOv5s network faces two challenges when directly applied to identifying external damage hazards in transmission line corridors: insufficient feature extraction for small-scale damage images and inadequate ability to extract discriminative yet inconspicuous details due to complex backgrounds and occlusions. In this regard, we propose an optimized ConTeXt module to improve the backbone network of the YOLOv5s model, enhancing feature extraction for small-scale targets while maintaining the same computational overhead. Figure 3 illustrates the optimized ConTeXt module.

The ConvNeXt module is a Depthwise Separable Convolution (DSC) that, based on ResNet, incorporates the design advantages of the Swin Transformer [16]. The introduction of ConvNeXt effectively addresses inadequate recognition of small-scale external damages. However, it is prone to miss inspections in areas with dense occluded where characteristics are not distinct. Therefore, we further embed the SimAM attention module into the ConvNeXt module, forming a new feature extraction module (ConvNeXt-SimAM) to better focus on extracting features in regions of interest. The SimAM attention mechanism calculates the weight of each neuron in the hidden layer using an energy function, and selects more discriminative neurons. Introducing the SimAM attention mechanism enables the ConvNeXt network to focus more closely on the crucial information required for hazard recognition tasks.

It is susceptible to feature redundancy or conflict problems due to the direct addition of different levels of semantic features in the FPN-PAN structure of the YOLOv5s network [19]. Therefore, we use the ASFF network to autonomously set different weight values for each feature map, which enhances the fusion of multi-scale features. Figure 4 illustrates the ASFF fusion approach.

After passing through the PANet network, three feature layers of different scales are generated, namely Level1, Level2, and Level3. ASFF adaptively adjusts the weights of the three-level feature maps. The specific algorithm is shown in the following equation.

$$y_{ij}^{l} = \alpha_{ij}^{l} x_{ij}^{1\to l} + \beta_{ij}^{l} x_{ij}^{2\to l} + \gamma_{ij}^{l} x_{ij}^{3\to l} \tag{1}$$

where, y_{ij}^{l} represents the output feature vector at position (i, j). $x_{ij}^{1\to l}$, $x_{ij}^{2\to l}$, $x_{ij}^{3\to l}$ represent the input features at position (i, j) from Level1, Level2, and Level3, respectively. $\alpha_{ij}^{l}, \beta_{ij}^{l}, \gamma_{ij}^{l}$ represent the weight parameters at

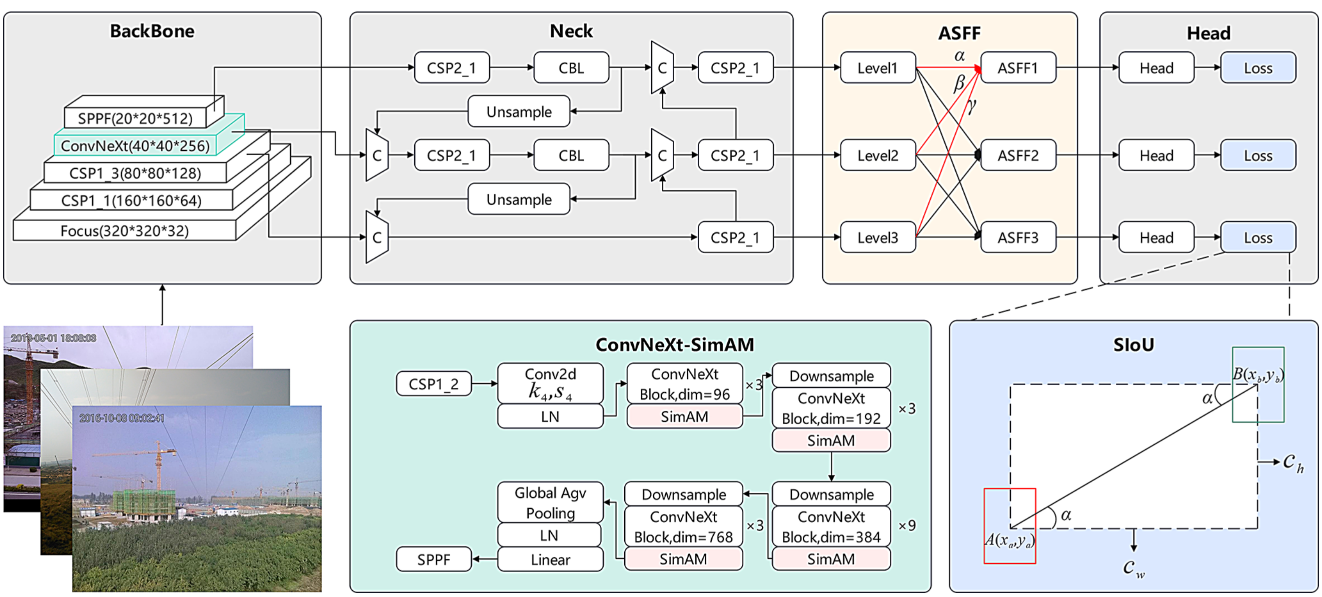

Fig. 2 YOLO-CS-ASFF network structure diagram

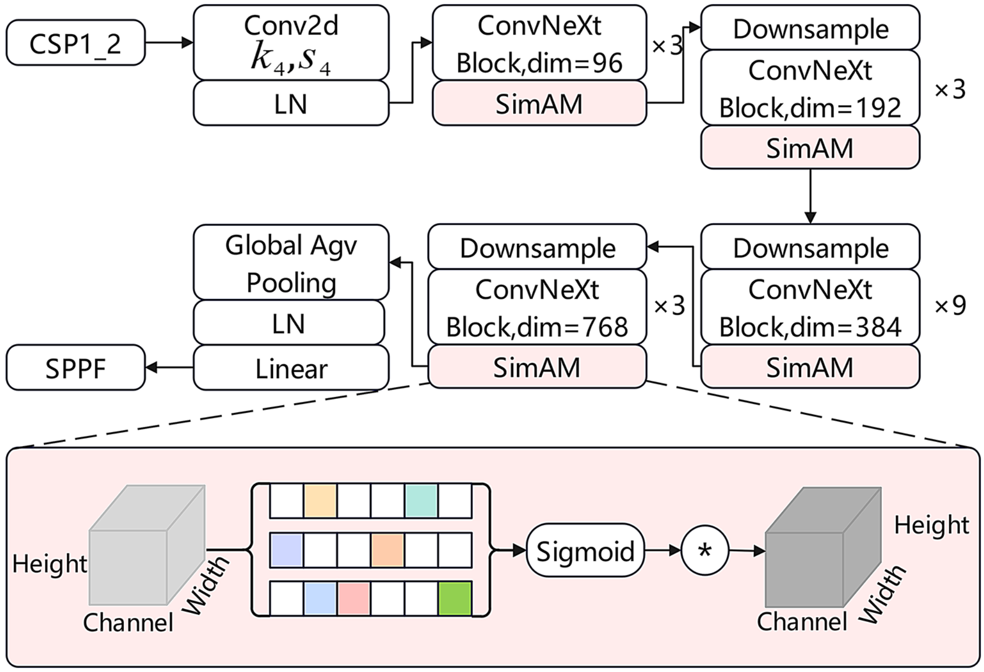

Fig. 3 Optimized ConvNeXt module

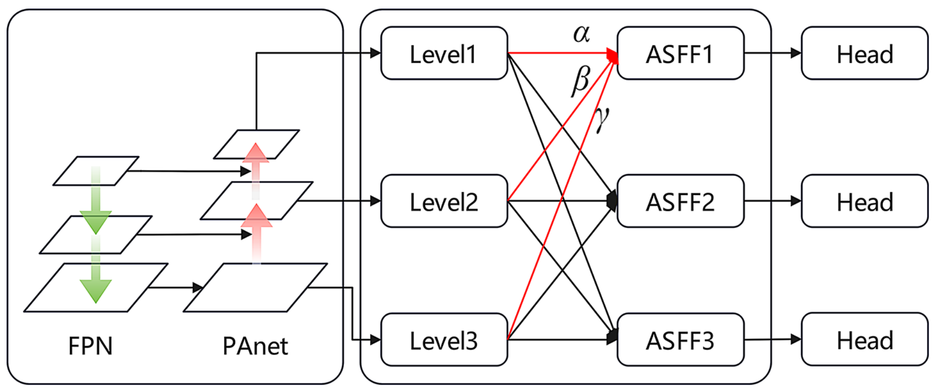

Fig. 4 Illustration of ASFF approach

position (i, j). Taking ASFF1 feature layer as an example, the output feature vector y_{ij}^{l} at position (i, j) is obtained by first multiplying the input features $x_{ij}^{1\to l}, x_{ij}^{2\to l}, x_{ij}^{3\to l}$ with the corresponding weight parameters $\alpha_{ij}^{l}, \beta_{ij}^{l}, \gamma_{ij}^{l}$, and then summing them up.

The adaptive feature weighting method of ASFF avoids insufficient feature representation caused by direct network concatenation and effectively strengthens the fusion of multi-scale features. It enhances the perception capability of different scale hazards.

2.2 Improvement of loss function based on SIoU

The YOLOv5s network adopts the Complete-IoU (CIoU) loss function, which considers three aspects: center distance, aspect ratio, and overlap area, but ignores the directional alignment between the predicted box and the truth box, resulting in slow convergence [20]. Especially in transmission line scenes, the scale of hazard targets changes as they move, making it difficult to accurately locate the truth box due to complex backgrounds or noise interference.

Therefore, we use the SIoU (SCYLLA-IoU) loss function instead of CIoU in our improved model. SIoU redefines the penalty mechanism and the angle between the vector of the predicted box and the truth box, which enhances the matching accuracy and improves the performance of the target detection model. Figure 5 illustrates the parameter diagram of the SIoU loss function.

3 Based on StrongSORT's tracking and prediction

To ensure accurate tracking of construction vehicles after occlusion, the YOLO-CS-ASFF is used as the object detector, which outputs information about detected targets. Subsequently, the StrongSORT object tracker is employed for

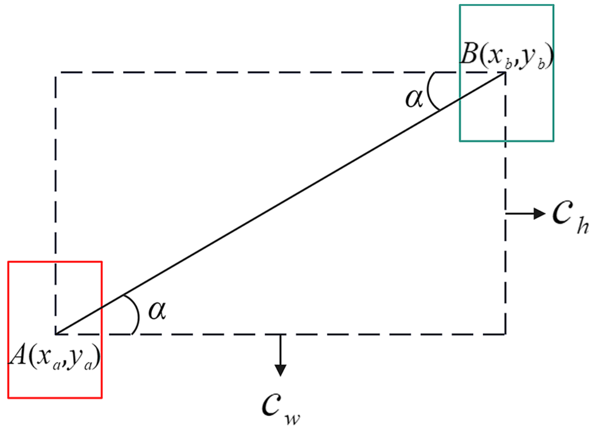

Fig. 5 Schematic diagram of SIoU border regression

re-identification, allowing for timely judgments about the future trends of construction vehicles. Specifically, the input image sequence of construction vehicles is processed by the YOLO-CS-ASFF detection model, which generates bounding boxes, class labels, and confidence scores. This detection model's efficiency and accuracy significantly decrease error rates in subsequent tracking, reducing missed and false detections. Once the detector identifies the targets' positions and classes, these results are input into the OSNet module of the StrongSORT tracker to extract their appearance features. Simultaneously, the NSA Kalman filter is used to predict the targets' motion features. The Hungarian algorithm optimally matches targets between consecutive frames, assigning unique identifiers and continuously tracking their trajectories. High-precision detection provides accurate inputs for tracking, reducing mismatch probability. Conversely, tracking the targets' long-term motion helps mitigate detection errors due to temporary occlusions or rapid changes.

However, the StrongSORT algorithm utilizes a BoT network based on ResNet improvements. It has a large number of parameters and model size, which result in increased computational resources requirements and processing time. When deployed on transmission line inspection equipment, it fails to meet the requirements of lightweight networks. In contrast, the lightweight OSNet network has demonstrated outstanding detection performance while providing higher real-time performance. Therefore, the feature extraction network is replaced by OSNet. The overall workflow is illustrated in Fig. 6, and the specific implementation steps are as follows:

Step 1: The YOLO-CS-ASFF network is utilized for real-time detection of video images to obtain the target detection boxes and corresponding features of the query frame. The recognition results are used as inputs to the StrongSORT tracking algorithm.

Step 2: The StrongSORT tracking algorithm generates appearance and motion information based on the target detection boxes of the query frame. In the appearance branch, the OSNet algorithm and Exponential Moving Average (EMA) update strategy are used to achieve more accurate associations between detection and tracking in complex scenarios. The calculation formula for the appearance state e_i^t is as follows.

$$e_i^t = \alpha e_i^{t-1} + (1-\alpha)\, f_i^t \tag{2}$$

where f_i^t is the appearance embedding of the current matched detection, and α is the momentum term, with a value of 0.9.

On the motion branch, the NAS Kalman filter algorithm is utilized to determine the state (Tracks) of the query frame based on the tracking results of the memory frame. The tracking state is divided into confirmed state (Confirmed) and unconfirmed state (Unconfirmed). The adaptive calculation of the noise covariance $\widetilde{R}_k$ given as follows.

$$\tilde{R}_k = (1 - C_k)\, R_k \tag{3}$$

where $\widetilde{R}_k$ is the pre-set fixed measurement noise covariance, and C_k is the detection confidence score under state k.

Step 3: Using the target's appearance information, the confirmed tracking results are associated with the detection results using the global linear matching of the Hungarian algorithm. The unconfirmed tracking results further use IoU matching. There are three cases in the matching results: the first one is detection matching failureand will be deleted. The second is that the trace match fails and its state will be evaluated. If it is an unconfirmed state, delete it; if it is a confirmed state, further determine the maximum number of target losses. If it exceeds the threshold, delete it; otherwise, track its trajectory in the next iteration. The third case is tracking matching success, which returns the NSA Kalman filter updates the state. The tracking results of the query frame are then input into a new tracking trajectory.

In the matching process, appearance information and motion information are combined for feature integration, and the cost matrix C is as follows.

$$C = \lambda A_a + (1-\lambda)\, A_m \tag{4}$$

Where A_aa represents the appearance cost, A_m represents the movement cost, and the weight parameter λ is set to 0.98.

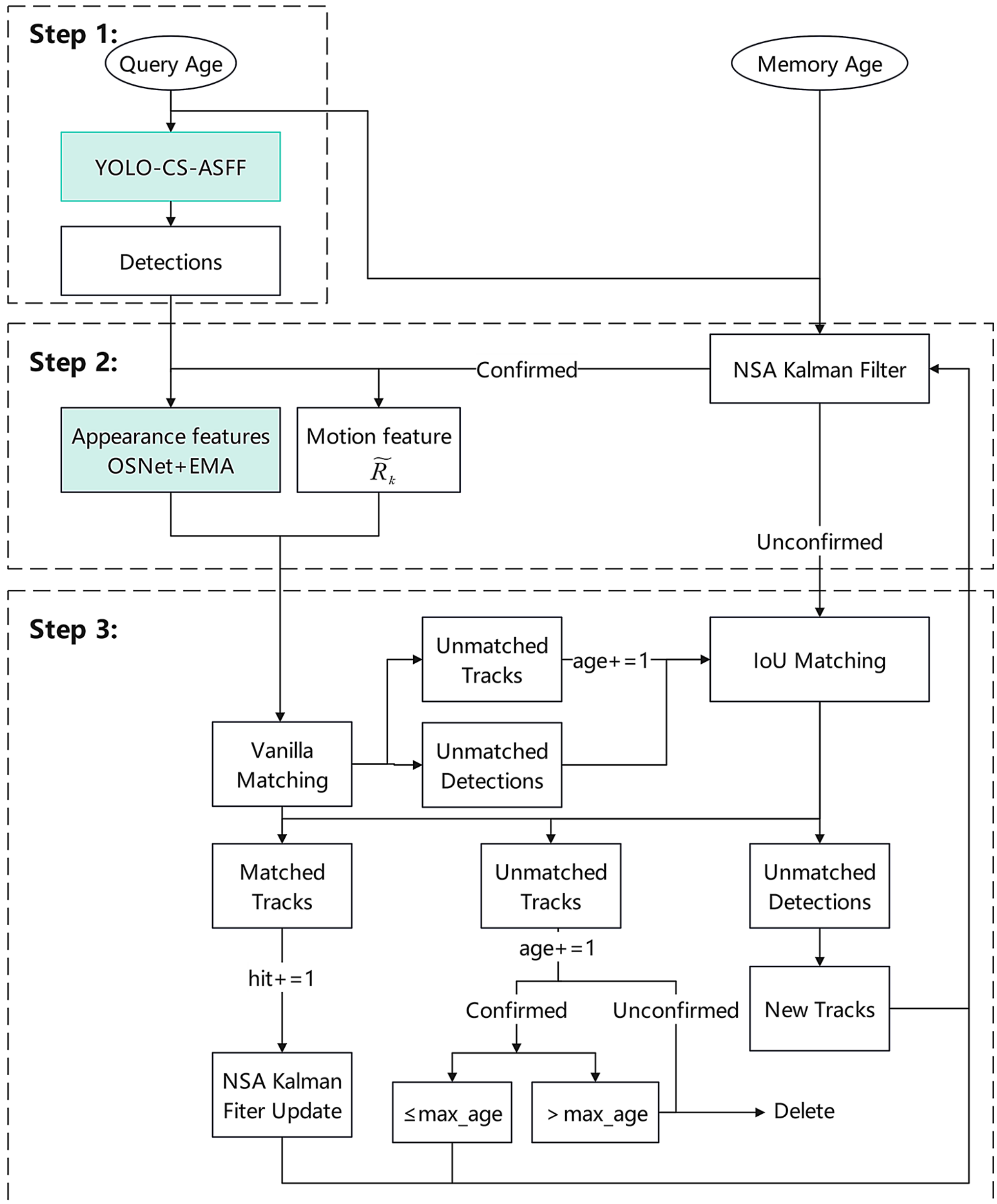

Fig. 6 StrongSORT tracking flowchart

4 Experiment and analysis

4.1 Dataset collection

Due to the lack of publicly available datasets of hidden danger images for transmission line corridors, we obtained image and video data of hidden dangers through online collection and manual methods from various angles and natural environments. We divided the hidden danger targets into two categories: foreign objects and construction vehicles. Foreign objects include kites, bird nests, plastic bags, and balloons, while construction vehicles include excavators, cranes, loaders, and tower cranes. However, there is a shortage of samples for kites, balloons, and loaders, resulting in an imbalance in sample proportions. To avoid overfitting, we employed image augmentation techniques such as noise introduction, brightness adjustment, blurring, and affine transformation [21], to expand the original 3,204 samples to 15,251 samples, as shown in Fig. 7.

To evaluate the effectiveness of the model before and after the improvement, we divided the original dataset and the augmented dataset into training, validation, and testing sets in the ratio of 8:1:1.

Additionally, for the subsequent tracking experiments, we selected four videos (comprising a total of 5,204 frames)

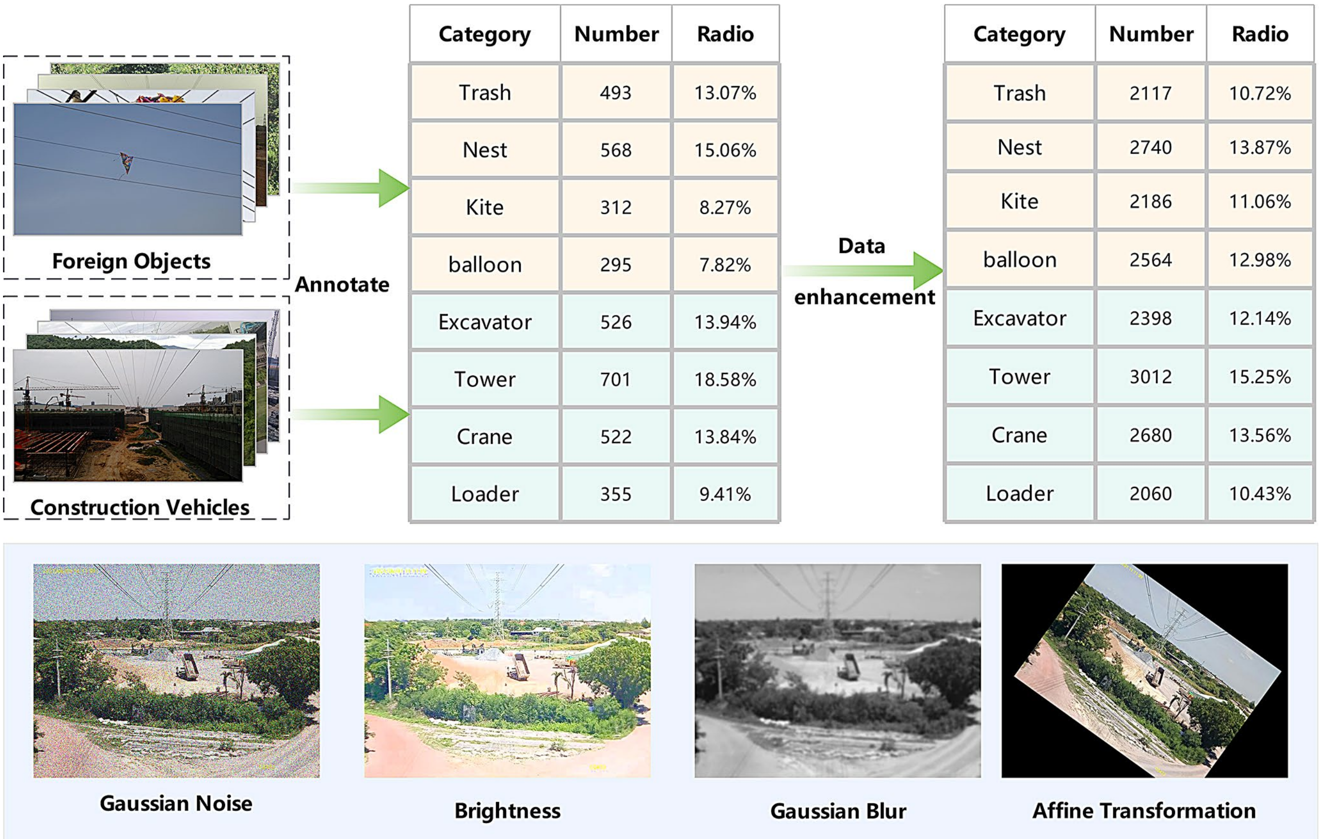

Fig. 7 Dataset construction process

to form an engineering vehicle dataset,. The dataset was divided into training, validation, and testing sets at a ratio of 8:1:1, where the training sets are 4,164 frames and test sets are 520 frames each.

All the experiments were performed on a workstation equipped with NVIDIA GeForce GTX 3090 GPU, the CUDA 10.1 parallel computing platform and running the Ubuntu 18.04 operating system with the PyTorch framework. The input image size was set to 640*640*3, and the network was trained the AdamW optimizer.

To ensure fairness in comparative experiments, we standardized the experimental conditions. The input image size was set to 640*640*3, and the network was trained the AdamW optimizer. The baseline YOLOv5s model has an initial learning rate of 0.001, a momentum of 0.937, a weight decay of 0.0005, 300 epochs, and a batch size of 8. Among these, the initial learning rate and number of epochs are critical parameters influencing the model's convergence speed and performance. Overly high or low parameter settings can adversely affect the model's convergence and final performance. The batch size determines the sample count per parameter update, typically set to 16, 32, or 64, depending on hardware capabilities and model update frequency. Momentum helps optimize the search process to avoid local optima and is usually set to 0.9, maintaining high training speed and smoothing gradient fluctuations. Weight decay is crucial for preventing overfitting. A well-chosen weight decay can enhance the model's generalization, as shown in similar tasks [22].

Table 1 The average detection rate of the model under different learning rates and iterations

Iterations	Learning-rates		
	0.01	0.001	0.0005
150	78.79	78.39	77.94
200	**80.64**	79.17	79.11
300	80.63	78.94	78.96

During training, we fine-tuned key parameters like initial learning rate, number of epochs, and batch size based on our custom dataset of transmission line corridor hazard images. We conducted experiments evaluating average detection rate and accuracy (detailed in Tables 1 and 2). Results indicate that the model performs optimally with a learning rate of 0.01 and 200 epochs. Additionally, a larger batch size does not always yield better results; we chose a batch size of 32 to balance convergence speed and accuracy.

To quantitatively analyze the performance of the YOLO-CS-ASFF object detector, typical evaluation metrics of object detection algorithms were employed, including

Table 2 The average precision of the model different batch-size

Batch-size	AP								
	Trash	Nest	Kite	Balloon	Excavator	Tower	Crane	Loader	AP@0.5
16	87.6	88.1	87.5	90.6	92.8	92.5	91.8	90.7	90.2
32	88.6	91.3	90.9	91.2	94.3	93.8	94	91.9	**92.0**
64	89.3	90.7	90.6	90.9	93.6	92.9	94	91.5	91.7

Table 3 Comparison experiment of object detectors

Model	mAP@0.5/%	Recall/%	Precision/%	Model size/MB
SSD	81.03	78.50	76.32	76.21
YOLOv4	87.03	81.80	80.65	59.19
YOLOv5s	89.56	83.98	94.54	54.55
YOLOv7	88.94	82.05	92.38	67.43
Faster-RCNN	90.94	84.25	85.86	177.66
YOLO-CS-ASFF	**92.80**	**85.30**	**95.06**	**58.15**

Table 4 Ablation experiment

Group	Improvemnt strategies	mAP/%	Recall/%	Size
A	YOLOv5s	89.56	83.98	54.55
B	A+ConvNeXt	91.35	84.20	56.58
C	B+SimAM	91.65	85.08	57.26
D	C+ASFF	92.18	85.16	58.15
E	D+SIoU	92.80	85.30	58.15

Accuracy (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP). Among these metrics, Precision measures the ratio of correctly identified targets within the predicted results, with higher values indicating more accurate recognition. Recall measures the ratio of correctly detected targets relative to all actual targets. The Average Precision (AP) reflects the relationship between Precision and Recall over the entire dataset. The mean Average Precision (mAP) is obtained by averaging the AP values across all target categories, which is commonly used to evaluate the overall detection performance of the entire network model.

To evaluate the performance of the improved StrongSORT tracker, the following metrics from the CLEAR MOT framework were utilized: Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and Frames Per Second (FPS). Among these metrics, MOTA reflects the occurrences of false positives, missed detections, and ID switches in the tracking process, with higher values indicating better model performance. MOTP measures the tracking error between the actual and tracked results and the ground truth annotations in the dataset, with values closer to 1 indicating better accuracy. FPS represents the model's recognition speed, with higher values indicating suitability for edge devices. The calculation of these evaluation metrics follows the methodology described in reference [23].

4.2 Detector selection and performance analysis

4.2.1 Object detection model selection

The detection performance of One-Stage and Two-Stage methods varies for different scenarios and objectives. To accurately select the target detector, we evaluated SSD, RetinaNet, YOLOv4, YOLOv5s, and YOLOv7 in One-Stage, as well as Faster R-CNN in Two-Stage, on a unified validation set.

As shown in Table 3, the Faster R-CNN algorithm in the Two-Stage approach performs well in terms of mAP@0.5 and recall. However, its relatively large model size and low detection accuracy make it unsuitable for detecting external breakage hazards. In the One-Stage approach, SSD's mAP@0.5 and recall are much lower than those of the YOLO series. Among the YOLO series, YOLOv5s outperforms other algorithms in mAP@0.5, recall, precision, and model size, reaching 89.56%, 83.98%, 94.54%, and 54.55 MB, respectively.

Furthermore, our model outperforms other models in mAP and recall. Among them, the mAP@0.5 of YOLO-CS-ASFF reaches 92.8%, representing a 3.24% improvement over the original YOLOv5s; the recall rate increases by 1.32%, while the false negative rate decreases. Despite an increase in parameter, it remains lower than that of other models. These results indicate that YOLO-CS-ASFF enhances the model's adaptability to targets at different scales, thereby improving detection accuracy and recall, through an improved feature fusion mechanism. Comparative experimental results show that the proposed method significantly outperforms Faster R-CNN, YOLOv5s, and YOLOv7 in detection accuracy and speed, effectively detecting targets of varying scales. It demonstrates that the proposed method exhibits excellent performance in quickly and accurately identifying and locating external damage risks within the transmission line corridor.

4.2.2 Fusion experiment

To further validate the performance of the improvement strategies in the YOLO-CS-ASFF model, we conducted ablation experiments using the constructed dataset of hazards in transmission corridors. Table 4 presents the results.

When we replaced the last CSP_3 module in the YOLOv5s backbone network with the ConvNeXt, the mAP@0.5 and

Table 5 Improvement effects of target tracking algorithms

Model	MOTA(%)	MOTP(%)	Frame rate	IDs
YOLOv5s + StrongSORT	49.7	71.3	29.8	36
YOLO-CS-ASFF + StrongSORT	36.6	70.8	30.1	24
YOLOv5s + StrongSORT(Improved)	59.6	72.7	32.3	17
Ours-SORT	63.3	78.9	37.8	11

Table 6 Target tracking algorithms performance comparison results

Model	MOTA(%)	MOTP(%)	Frame rate	IDs
DeepSORT	56.9	69.8	21.9	40
ByteTrack	60.1	78.6	31.8	32
Ours-SORT	63.3	78.9	37.8	11

recall rates improved to 91.35% and 84.2%, respectively. It indicates an increased ability to broaden the receptive field and enhance the detection accuracy of small targets. Integrating SimAM with ConvNeXt led to a significant increase in recall rates, which was further improved by 0.88% from Group B. The mAP was also slightly improved, indicating its effectiveness in reducing the influence of complex background. Although YOLOv5s shows slightly improved detection accuracy after incorporating ConvNeXt enhancements, YOLO-CS-ASFF, with the addition of the ASFF module, further enhances overall performance, increasing average precision and recall rates to 92.18% and 85.16%, respectively. It indicates that ASFF module better handles multi-scale fusion. In addition, the SIoU loss function has increased average precision by 3.24% points compared to CIoU, and the recall rate has also increased by 1.32%.

4.3 Tracking and visualization analysis

The experimental results indicate that the improved object detector model achieves high detection accuracy. To further validate the effectiveness of the proposed method in tracking and detecting construction vehicles in transmission corridor videos, we conducted experiments on a manually annotated dataset, as shown in Table 5. Additionally, we compared our algorithm with two representative multi-object tracking algorithms, DeepSORT [13] and ByteTrack [15], with the comparative results shown in Table 6.

As shown in Table 5, the optimized object detector (YOLO-CS-ASFF) combined with the original StrongSORT reduces IDs switch, decreasing by 12 instances compared to the baseline DeepSORT algorithm. However, while the improved detector enhances accuracy, the

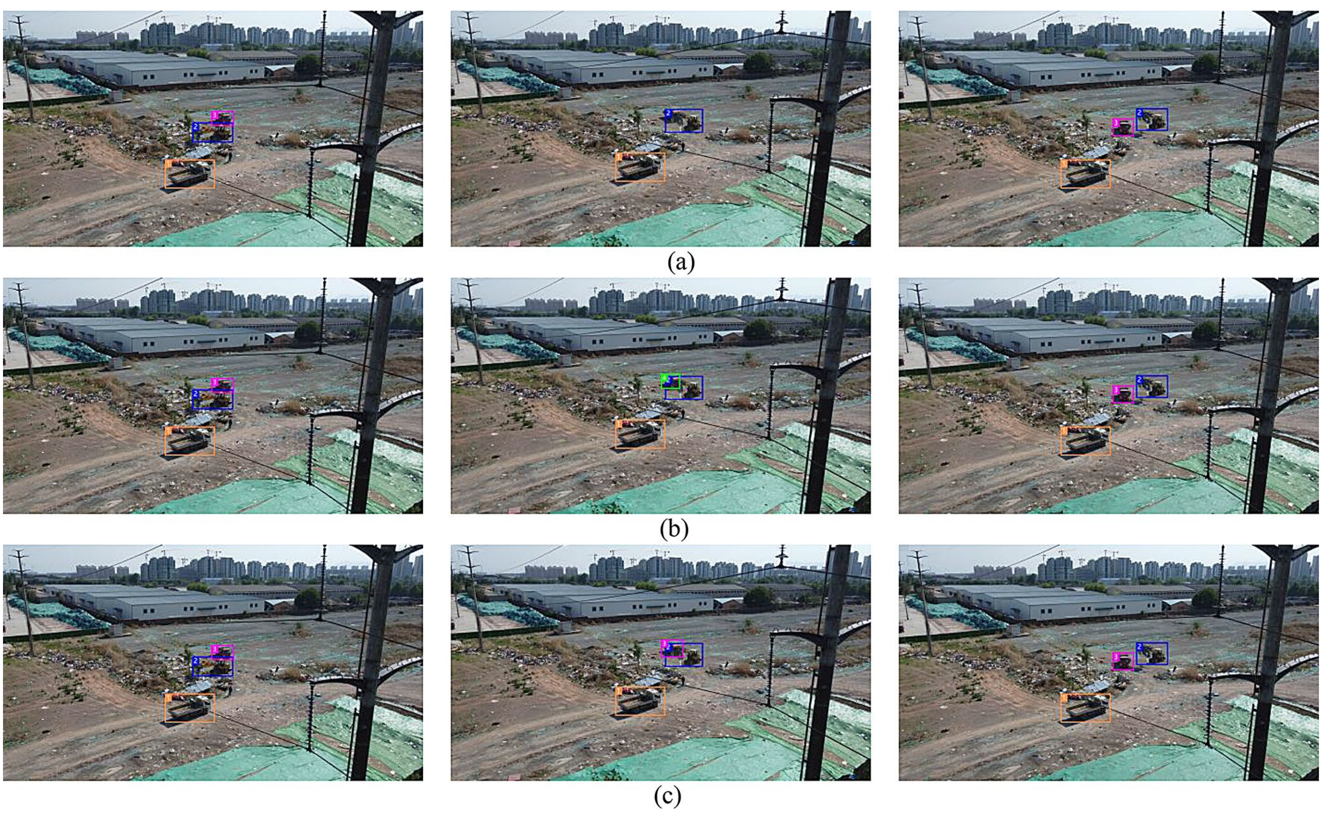

Fig. 8 Comparison experiment of two tracking algorithms. (**a**) Tracking results of Original DeepSORT. (**b**) Tracking results of Original StrongSORT. (**c**) Tracking results of improved StrongSORT

original tracker's limited feature extraction capability negatively impacts tracking precision, leading to a decline in overall performance. Combining YOLO-CS-ASFF with the improved StrongSORT (Ours-SORT), the MOTA and MOTP reach 63.3% and 78.9%, respectively, representing increases of 3.7% and 6.2% compared to the original algorithm (YOLOv5s + StrongSORT). Additionally, the video processing speed reaches 37.8 frames per second, meeting the real-time monitoring requirements for construction vehicles approaching transmission corridors. Additionally, combining the original detector with the optimized tracker also shows improvements in tracking performance. Although not as effective as the jointly optimized detection and tracking approach, this combination still enhances tracking accuracy and precision compared to the baseline. Table 6 shows a comparison of multi-object tracking algorithms, where DeepSORT and ByteTrack perform worse in tracking accuracy and drift control. In contrast, the proposed method outperforms others in terms of tracking accuracy, precision, and ID switch frequency.

Figure 8 shows the comparative experimental results of the improved StrongSORT algorithm, the original StrongSORT algorithm, and the original DeepSORT algorithm for tracking engineering vehicles in the same video sequence. The three images displayed are extracted from the video sequence in chronological order. In Fig. 8a, the original DeepSORT model fails to detect the engineering vehicle with ID number 3 after being obscured by another vehicle with ID number 2 and then reappears, resulting in a missed detection due to occlusion. In Fig. 8b, although the target is successfully re-identified after occlusion, its ID number changes to 4, leading to a false detection. Both scenarios illustrate tracking failures. In Fig. 8c, our algorithm maintains the same ID number for the target even under similar circumstances, demonstrating effective suppression of target ID switching while maintaining a high image processing speed.

5 Conclusion

This paper proposes an algorithm for perceiving external hazards in transmission line corridors, aiming to address the real-time detection and tracking challenges posed by diverse hazards of different scales and types in complex environments. To enhance detection performance, we leverage the YOLOv5s base network and optimize feature extraction and feature fusion networks to improve the detection of critical features across different scales of hazards. In addition, we introduce SIOU to accelerate model convergence. The improved YOLO-CS-ASFF target detection model is integrated into the tracking model, and the OSNet module is utilized to enhance the tracking capabilities of StrongSORT. Experimental results show that the proposed YOLO-CS-ASFF algorithm achieves a 3.24% increase in mean Average Precision (mAP), reaching 92.8%. The optimized StrongSORT exhibits faster processing speed with a frame rate of 37.8 frames per second (FPS), enabling real-time monitoring of foreign objects in transmission line corridors. This method not only accurately identifies potential external breakage target categories but also quickly and precisely locates potential hazard positions, guiding operation and maintenance personnel to promptly eliminate hazards and significantly enhancing efficiency. However, some issues remain in practical scenarios, such as the limited dataset of transmission line corridor hazards and the lack of multi-source information fusion. For these scenarios, it is necessary to further enrich the dataset and consider integrating different signal types, such as electrical signals and temperature, with visual data to establish an intelligent multi-modal perceptual model for transmission lines.

Acknowledgements We would like to thank the anonymous reviewers for their valuable and helpful comments, which substantially improved this paper. At last, we also would also like to thank all of the editors for their professional advice and help.

Author contributions All authors read and approved the final manuscript.

Funding No Financial support.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Zhang, W., Li, Y., Liu, A.: RCDAM-Net: A foreign object detection algorithm for transmission tower lines based on RevCol Network. Appl. Sci. **14**, 1152 (2024). https://doi.org/10.3390/app14031152
2. Yang, L., Liu, J.F.Y., Li, E., Peng, J., Liang, Z.: A review on state-of-the-art power line inspection techniques. Ieee T Instrum. Meas. **69**, 9350–9365 (2020). https://doi.org/10.1109/TIM.2020.3031194
3. Xuezhi Xiang, N., Lv, X., Guo, S., Wang, A.E., Saddik: Engineering vehicles Detection based on modified faster R-CNN for Power Grid Surveillance. Sensors. **18**, 2258 (2018). https://doi.org/10.3390/s18072258
4. Chunyang Liu, L., Ma, X., Sui, N., Guo, F., Yang, X., Yang, Y., Huang, X., Wang: YOLO-CSM-Based component defect and foreign object detection in overhead transmission lines. Electronics. **13**, 123 (2024). https://doi.org/10.3390/electronics13010123
5. Zhang, D., Zhang, Z., Zhao, N., Wang, Z.: A lightweight modified YOLOv5 Network using a swin transformer for transmission-line

foreign object detection. Electronics. **12**, 3904 (2023). https://doi.org/10.3390/electronics12183904
6. Shao, J., Li, Y.X., Liu, M., Li, L., Ding: DF-YOLO: Highly accurate transmission line foreign object detection algorithm. IEEE Access. **11**, 108398–108406 (2023). https://doi.org/10.1109/ACCESS.2023.3321385
7. Chenhui Yu, Y., Liu, W., Zhang, X., Zhang, Y., Zhang, X., Jiang: Foreign objects Identification of Transmission Line based on improved YOLOv7. IEEE Access. **11**, 51997–52008 (2023). https://doi.org/10.1109/ACCESS.2023.3277954
8. Zhao, A., Lv, H., Wang, Y., Li, H., Li, Y., Zou, G.: Research on External Damage Detection of Transmission Line In Complex Background. In: 2023 2nd International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC). pp. 171–174 (2023). https://doi.org/10.1109/RAIIC59453.2023.10280789
9. Wojke, N., Bewley, A., Paulus, D.: Simple online and realtime tracking with a deep association metric. In: 2017 IEEE International Conference on Image Processing (ICIP). pp. 3645–3649 (2017). https://doi.org/10.1109/ICIP.2017.8296962
10. Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., Wang, X.: ByteTrack: Multi-object Tracking by Associating Every Detection Box. In: Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.) Computer Vision – ECCV 2022. pp. 1–21. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-20047-2_1
11. Yunhao Du, Z., Zhao, Y., Song, Y., Zhao, F., Su, T., Gong: Hongying Meng: StrongSORT: Make DeepSORT Great again. Ieee T Multimedia. **25**, 8725–8737 (2023). https://doi.org/10.1109/TMM.2023.3240881
12. Kusumah, A.P., Djayusman, D., Setiadi, G.R., Nugraha, A.C., Hidayatullah, P.: Counting various vehicles using YOLOv4 and DeepSORT. J. Integr. Adv. Eng. (JIAE). **3**, 1–6 (2023). https://doi.org/10.51662/jiae.v3i1.68
13. Thioanh Bui, G., Wang, G., Wei, Q.Z.: Vehicle multi-object detection and tracking Algorithm based on Improved you only look once 5s Version and DeepSORT. Appl. Sci. **14**, 2690 (2024). https://doi.org/10.3390/app14072690
14. Zhang, Y., Zhang, T., Huang, Z.: Multiple vehicle detection and tracking using improved YOLOv5 and strong SORT. In: International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2022). pp. 592–596. SPIE (2023). https://doi.org/10.1117/12.2674583
15. Ng, J.J., Goh, K.O.M., Tee, C.: Traffic Impact Assessment System using Yolov5 and ByteTrack. J. Inf. Web Eng. **2**, 168–188 (2023). https://doi.org/10.33093/jiwe.2023.2.2.13
16. Zhuang Liu, H., Mao, C.-Y., Wu, C., Feichtenhofer, T., Darrell, S., Xie: A ConvNet for the 2020s. Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) June 1 (2022). https://doi.org/10.1109/CVPR52688.2022.01167
17. Lingxiao Yang, R.-Y., Zhang, L., Li, X., Xie: SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. In: Proceedings of the 38th International Conference on Machine Learning. pp. 11863–11874. ACM (2021)
18. Mulan Qiu, L., Huang, B.-H., Tang: ASFF-YOLOv5: Multielement Detection Method for Road Traffic in UAV images based on Multiscale Feature Fusion. Remote Sens-basel. **14**, 3498 (2022). https://doi.org/10.3390/rs14143498
19. Qiao, Y., Guo, Y., He, D.: Cattle body detection based on YOLOv5-ASFF for precision livestock farming. Comput. Electron. Agr. **204**, 107579 (2023). https://doi.org/10.1016/j.compag.2022.107579
20. Chen, F., Zhang, L., Kang, S., Chen, L., Dong, H., Li, D., Wu, X.: Soft-NMS-Enabled YOLOv5 with SIOU for Small Water Surface Floater detection in UAV-Captured images. Sustainability. **15**, 10751 (2023). https://doi.org/10.3390/su151410751
21. Hossain, M.D.Z., Sohel, F., Shiratuddin, M.F.: Hamid Laga: A Comprehensive Survey of Deep Learning for Image Captioning. Acm Comput. Surv. **51**, 118:1–118 (2019). https://doi.org/10.1145/3295748
22. Zheng, H., Hu, S., Liang, Y., Huang, J., Wang, T.: A hidden Danger object detection method for Transmission Line Corridor based on YOLO-2MCS. Diangong Jishu Xuebao/Transactions China Electrotechnical Soc. **39**, 4164–4175 (2024). https://doi.org/10.19595/j.cnki.1000-6753.tces.230666
23. Xinquan Ye, J., Pan, F., Shao, G., Liu, J., Lin, D., Xu, J., Liu: Exploring the potential of visual tracking and counting for trees infected with pine wilt disease based on improved YOLOv5 and StrongSORT algorithm. Comput. Electron. Agr. **218**, 108671 (2024). https://doi.org/10.1016/j.compag.2024.108671