

HCL-YOLOv8: Lightweight Transmission-Line Insulator Defect Detection via Hierarchical Consistency Learning and Task-Aligned Label Assignment

Lei Wang

Inner Mongolia Electric Power Digital Research Company

Jiamao Han

Inner Mongolia Electric Power Digital Research Company

Yaxing Wei

Inner Mongolia Electric Power Digital Research Company

Bin Zhao

Inner Mongolia Electric Power Digital Research Company

Jing Xu

Inner Mongolia Electric Power Digital Research Company

Lei Han

Inner Mongolia Electric Power Digital Research Company

Yu Zhang

Inner Mongolia Electric Power Digital Research Company

Yingchao Zhen

Inner Mongolia Electric Power Digital Research Company

E erdun

Inner Mongolia Electric Power Digital Research Company

Lai Man

Inner Mongolia Electric Power Digital Research Company

Jian Wei

Inner Mongolia Electric Power Digital Research Company

Liguo Qu

lgqu@ahnu.edu.cn

Anhui Normal University

Keywords: insulator defect detection, small object detection, hierarchical consistency learning, task-aligned label assignment, YOLOv8, data augmentation

Posted Date: August 7th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9822971/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

HCL-YOLOv8: Lightweight Transmission-Line Insulator Defect Detection via Hierarchical Consistency Learning and Task-Aligned Label Assignment

Lei Wang¹, Jiamao Han¹, Yaxing Wei¹, Bin Zhao¹, Jing Xu¹, Lei Han¹, Yu Zhang¹, Yingchao Zhen¹, E. Erdun¹, Lai Man¹, Jian Wei¹, Liguo Qu^{*2}

¹Inner Mongolia Electric Power Digital Research Company, Hohhot, Inner Mongolia 010000, China

²School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China;

Abstract

Accurate detection of transmission-line insulator defects is essential for power-system inspection, yet it remains challenging because of unstable localization of tiny defects, severe interference from complex backgrounds, and insufficient exploitation of hierarchical structural priors. To address these issues, this paper proposes HCL-YOLOv8, a lightweight detection framework enhanced by hierarchical consistency learning. Built upon YOLOv8n, the proposed method leverages the natural hierarchical relationship in which defect instances are typically located within or near insulator string regions. Specifically, a P2 high-resolution detection head is introduced to improve small-object visibility, and a Soft Boundary Prior Branch (SBP-Branch) is designed to guide the network toward defect-relevant parent regions. To incorporate hierarchical priors into training, a Hierarchy-Constrained Task-Aligned Label Assignment strategy is proposed to restrict positive-sample selection for defect classes to geometrically valid parent regions, thereby reducing noisy positives outside insulator strings. Furthermore, a Hierarchical Containment Regularization Loss is developed to explicitly enforce the containment consistency between defect boxes and their matched parent boxes during training. To improve regression stability for tiny defects, a Small-Object-aware Normalized Gaussian Wasserstein Distance loss is introduced for defect-class positive samples. At inference time, a Hierarchy-aware Soft Re-scoring strategy is applied after non-maximum suppression to suppress high-confidence defect predictions outside insulator strings while preserving recall. In addition, a defect-sensitive stratified offline augmentation strategy is adopted to alleviate data imbalance and improve robustness under complex imaging degradations. Experiments conducted on the public insulator defect dataset released by Zheng et al. in *Sensors* show that the proposed method achieves a Precision of 0.923, a Recall of 0.911, an mAP@0.5 of 0.923, and an mAP@0.5:0.95 of 0.599 with only 3.12M parameters. Compared with the YOLOv8n baseline, Recall is improved by 3.4 percentage points and mAP@0.5 by 1.5 percentage points. These results indicate that HCL-YOLOv8 improves the detection of small defects, particularly flashover defects, while maintaining a favorable balance between detection performance and lightweight deployment requirements.

Keywords: insulator defect detection; small object detection; hierarchical consistency learning; task-aligned label assignment; YOLOv8; data augmentation

Introduction

Insulators are critical components of transmission lines, and their defects, such as flashover and broken, may cause line trips and even large-scale blackouts. Traditional inspection methods rely heavily on manual examination or hand-crafted rule-based algorithms, which often lack robustness under complex backgrounds, viewpoint variations, illumination degradation, and scale changes. With the rapid development of UAV- and robot-assisted inspection as well as visual perception technologies, deep-learning-based object detection has become a major approach for the automatic identification of insulators and their defects [1]. From the perspective of detection paradigms, two-stage detectors, such as Faster R-CNN [2], generally provide strong localization accuracy but incur considerable inference overhead, which limits their suitability for real-time edge deployment. In contrast, one-stage detectors, including SSD [3], the YOLO family, and FCOS [4],

offer a more favorable trade-off between speed and accuracy and are therefore more appropriate for engineering-oriented inspection scenarios.

Despite these advances, insulator defect detection still faces three common challenges. First, defect instances are typically tiny relative to the full image and often coexist with visually similar structures, such as wires and tower components, resulting in a high false-positive rate under complex backgrounds. Second, both category distribution and scene distribution are highly imbalanced: defect samples are scarce, whereas environmental degradations are diverse, including haze, overexposure, noise, and occlusion, which easily lead to overfitting during training. Third, hierarchical structural information has not been fully exploited. In a large proportion of annotated insulator defect samples, defect targets are almost invariably located within insulator string regions. If this inherent *part-of* relationship is ignored, inconsistency arises between sample assignment and training objectives, which may in

turn produce false alarms outside insulator strings or missed detections within them.

To address these issues, a new detection framework, HCL-YOLOv8, is developed around the central idea of making hierarchical structural priors learnable. Built upon YOLOv8n, the proposed framework explicitly incorporates the natural hierarchy between defects and insulator strings into the learning process, so that the prior is no longer treated as implicit annotation knowledge but as an effective supervisory signal for detection. By integrating hierarchical consistency constraints into the detection pipeline, the method aims to improve the localization stability of tiny defects, suppress background-induced false detections, and strengthen the utilization of structural priors in insulator defect detection. Its effectiveness is further validated on a public dataset.

1 Related Work

1.1 Research on Insulator and Defect Detection

Early studies on insulator detection mainly relied on edge, texture, and morphological rules, or traditional classifiers, and generally showed limited generalization in complex scenes. In recent years, research has gradually shifted toward deep learning. For transmission-line inspection, frameworks such as Faster R-CNN [2], SSD [3], and CenterNet [6] have been applied to insulator or defect detection and have achieved encouraging results; however, they still exhibit limitations in either real-time performance or robustness to small objects. With the rapid development of the YOLO family, many studies have sought to improve insulator defect detection through backbone redesign, feature fusion, and loss optimization. For example, Zheng et al. proposed an improved YOLOv7-based algorithm for insulator defect detection and released a public dataset together with an augmentation strategy, thereby providing a reproducible experimental basis for subsequent studies [1]. In addition, for industrial and power-inspection scenarios, existing work has improved models such as YOLOv5, YOLOv8, and YOLOv7 from the perspectives of lightweight design, attention mechanisms, enhanced non-maximum suppression, and suppression of complex-background interference [7].

1.2 Evolution of One-Stage Detectors and Small Object Detection

One-stage detectors have evolved from the end-to-end regression paradigm of YOLOv1 to the multi-scale prediction framework of YOLOv3, and further to the engineering refinements introduced in YOLOv4, YOLOX, and YOLOv7, continuously improving the trade-off between detection accuracy and inference speed [8-11].

In parallel, feature pyramid architectures, including FPN [12], PANet [13], and BiFPN/EfficientDet [14], have been widely adopted to enhance multi-scale feature representation, particularly for small object detection. Nevertheless, small object detection remains challenging be-

cause of detail loss caused by repeated downsampling, severe imbalance between positive and negative samples, and unstable box regression. To address these issues, prior studies have explored finer-grained feature layers, more effective sample assignment strategies, and more robust regression losses [15-16].

1.3 Sample Assignment and Loss Functions

In dense object detection, one of the key factors affecting performance is the alignment between positive/negative sample assignment and training objectives. Methods such as ATSS [15], PAA [17], and OTA [18] improve sample assignment from the perspectives of adaptive thresholding, probabilistic modeling, and optimal transport, respectively. The task-aligned paradigm introduced by TOOD emphasizes the consistency between classification and localization, and uses joint scores to guide positive-sample selection, showing strong generalization across different detectors [5].

For box regression, the IoU family of losses, including GIoU, DIoU, and CIoU, has been widely adopted. To better align localization quality with classification confidence, GFL [19] introduced distributed box representations together with joint classification-localization learning, while VFNet [21] further incorporated IoU-aware classification scores to improve dense detection performance. For extremely small objects, however, IoU-based losses may be overly sensitive to slight boundary perturbations. To address this issue, NWD (Normalized Gaussian Wasserstein Distance) models bounding boxes as Gaussian distributions and measures their Wasserstein distance, exhibiting a metric property that is more suitable for tiny object regression [16].

1.4 Data Augmentation and Class Imbalance

In industrial inspection, data augmentation is an important means of improving robustness and alleviating overfitting. Methods such as MixUp [22], CutMix [23], Cutout [24], and AutoAugment/RandAugment [25] enhance generalization through sample mixing, local occlusion, and automatic policy search. In engineering practice, Albumentations provides efficient augmentation operators together with BBOX synchronization mechanisms, making it well suited for constructing offline augmentation pipelines while preserving annotation consistency [26].

Although existing studies have advanced insulator defect detection from multiple perspectives, current approaches still have difficulty jointly addressing the instability of tiny-object localization, strong background interference, and the underutilization of hierarchical structural priors. In particular, the hierarchical relationship between defect targets and insulator string regions has not been explicitly integrated into sample assignment, training regularization, and inference-time decision refinement in a unified manner. This gap provides the main motivation for the method developed in this work.

2 Dataset and Offline Data Augmentation

2.1 Dataset Source, Category Definition, and Hierarchical Structure

The experimental dataset used in this study is derived from the insulator defect detection dataset published by Zheng et al. in *Sensors* [1]. That work publicly released images and annotations of transmission-line insulators and defect targets, upon which the offline augmentation strategy in this paper is constructed

The dataset adopts the YOLO annotation format, in which each target is represented by a five-tuple, as shown in Eq. (1):

$$y = (c, x, y, w, h) \quad (1)$$

where $c \in \{0,1,2\}$ is the class label, (x,y) are the normalized coordinates of the bounding box center, and (w,h) denote the normalized width and height, respectively. The three classes are defined as follows: $c = 0$: flashover (defect subclass); $c = 1$: broken (defect subclass); $c = 2$: insulator string (parent class target).



Fig. 1. Examples from the insulator defect dataset

As illustrated in Fig. 1, this annotation system exhibits an explicit hierarchical relationship: each defect subclass target is typically located inside, or in the neighborhood of, an insulator string bounding box. This hierarchical structure provides the prior foundation for the subsequent method design.

2.2 Offline Data Augmentation

Insulator defect inspection presents a typical long-tail distribution and domain-shift problem. To improve model robustness under complex imaging conditions while avoiding uncontrollable disturbances to the validation and test distributions caused by online augmentation, this study adopts an offline augmentation strategy and strictly preserves the original train/val/test split. Specifically, the original images and labels are first copied in full, and augmented samples are then generated within the same split directories. The final training set is therefore composed of both original and augmented images, as shown in Fig. 2.

Considering the scarcity of defect samples and category imbalance, a defect-sensitive stratified offline augmentation strategy is proposed. According to the priority of scarce defect categories, samples are divided into three tiers, and different augmentation factors and augmentation intensities are assigned to each tier. This design enlarges

the training set while preserving the discriminative textures of defect targets.

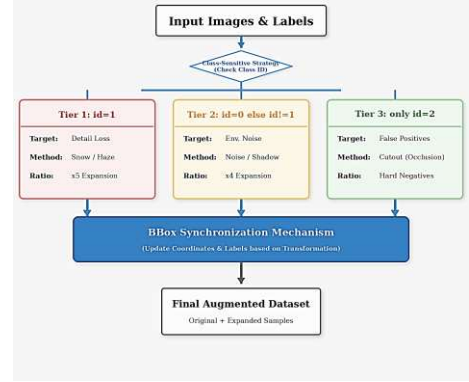


Fig. 2. Workflow for constructing the offline augmentation dataset

The augmentation pipeline is implemented using the Albumentations library. To simulate complex imaging conditions encountered in practical inspection, six types of augmentation operators are designed:

(1) snow glare/overexposure simulation, implemented using nonlinear tone curves, desaturation, gamma transformation, and CLAHE-based local contrast compensation;

(2) haze/transmission degradation, implemented by grayscale conversion, contrast reduction, JPEG compression artifacts, and Gaussian blur;

(3) shadow/backlight interference, implemented by random shadows, sun-flare simulation, and CLAHE;

(4) sensor noise, implemented using Gaussian noise;

(5) heavy occlusion, implemented by Cutout random block masking to generate hard negative samples; and

(6) amplitude-controlled geometric transformations, including horizontal flipping and small-range translation, scaling, and rotation, which are enabled only in designated templates to preserve defect textures.

The augmentation operators, parameter logic, and template/randomized combinations for different tiers are defined as follows:

(1) Tier 1 (broken, class_id=1): Five augmented images are generated from each original image using five deterministic templates: Snow Glare + horizontal flip; Haze + shift-scale-rotate; Shadow SunFlare + CLAHE; Sensor Noise; and an extreme combination of Snow Glare + Haze. Strong occlusion is not introduced in this tier in order to preserve broken textures.

(2) Tier 2 (flashover, class_id=0 and without broken defects): Four augmented images are generated from each original image using four deterministic templates: Snow Glare with CLAHE; Haze + CLAHE; Shadow SunFlare + CLAHE; and Sensor Noise. Strong occlusion is likewise avoided in this tier.

(3) Tier 3 (normal, only class_id=2): Forty augmented images are cyclically generated for each original image. In each cycle, one or two environmental degradation operators are randomly selected from Snow Glare, Haze,

Shadow SunFlare, and Sensor Noise and concatenated sequentially, while Coarse Dropout heavy occlusion is added with a probability of 0.9.

The visualized effects of the augmentation operators are shown in Fig. 3. With the BBOX synchronization mechanism, annotation consistency is preserved before

and after augmentation. After augmentation, the dataset reaches 10,621 images, among which Tier 1, Tier 2, and Tier 3 are expanded to 4,710, 3,820, and 2,091 images, respectively, thereby fulfilling the defect-priority design objective.

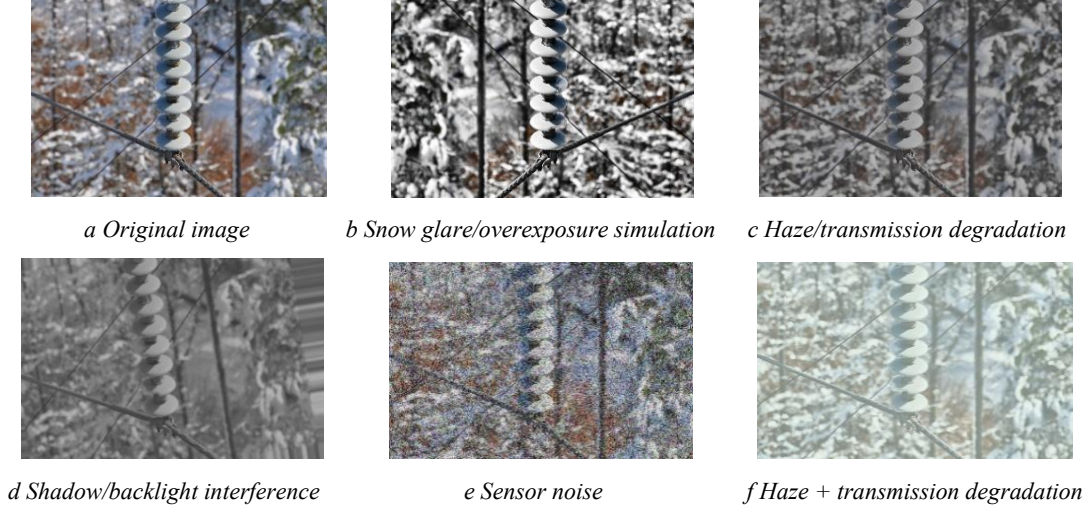


Fig. 3. Visualization of augmentation operator effects

3 Method: HCL-YOLOv8

3.1 Problem Definition

Given an input image $I \in \mathbb{R}^{H \times W \times 3}$ the objective of object detection is to predict a set of bounding boxes along with their associated class labels, as defined in Eq. (2).

$$\hat{Y} = \{(\hat{b}_i, \hat{c}_i, \hat{s}_i)\}_{i=1}^N \quad (2)$$

where H and W denote the input resolution, and N is the number of candidates. Each bounding box is represented as $\hat{b}_i = (\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$, corresponding to the normalized center coordinates and width–height representation (or equivalently the corner-based representation (x_1, y_1, x_2, y_2)). The class label $\hat{c}_i \in \{0, 1, 2\}$ denotes the object category (0: flashover, 1: broken, 2: insulator string). When $c_j \in C_d b_j^p$, and $\hat{s}_i \in [0, 1]$ represents the confidence score. For convenience, we denote the ground-truth set as $Y = \{(b_j, c_j)\}_{j=1}^M$, the corresponding parent object is defined as b_j^p . If multiple parent objects exist, the matching is determined based on inclusion relationships and area criteria. If no valid parent object is found, the defect sample is handled using the standard assignment strategy to preserve recall.

3.2 Overview of the Overall Framework

The overall network architecture is shown in Fig. 4. Using YOLOv8n as the baseline and without significantly increasing the parameter size, the proposed system introduces the following key mechanisms:

(1) Supporting modules: a P2 high-resolution detection head, a Soft Boundary Prior Branch (SBP-Branch),

and a Small-Object-aware Normalized Gaussian Wasserstein Distance loss (SO-NWD).

(2) Core optimizations

(i) Hierarchy-Constrained Task-Aligned Label Assignment (HC-TAL);

(ii) Hierarchical Containment Regularization Loss (HCR-Loss) and Hierarchy-aware Soft Re-scoring (HSR).

These components form a progressive closed loop: P2 and the SBP-Branch improve visibility, HC-TAL improves learnability, SO-NWD improves regressibility, HCR-Loss strengthens structural consistency, and HSR controls the risk of defects predicted outside the string during deployment.

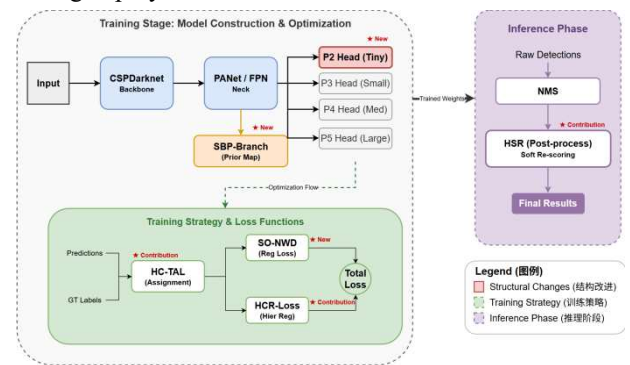


Fig. 4. Overall framework of HCL-YOLOv8

3.3 P2 Small-Object Detection Layer

The multi-scale detection head in YOLOv8n typically performs predictions on the feature levels $\{P3, P4, P5\}$. However, for insulator defects, which are characterized by very small pixel scales, repeated downsampling tends to dilute their responses in deeper feature maps. To enhance

the spatial resolution for small object detection, we introduce a higher-resolution $P2$ layer into the feature pyramid. As illustrated in Fig. 5, the detection head is extended to operate on $\{P2, P3, P4, P5\}$, thereby providing denser candidate locations and finer-grained localization gradients for defect detection.

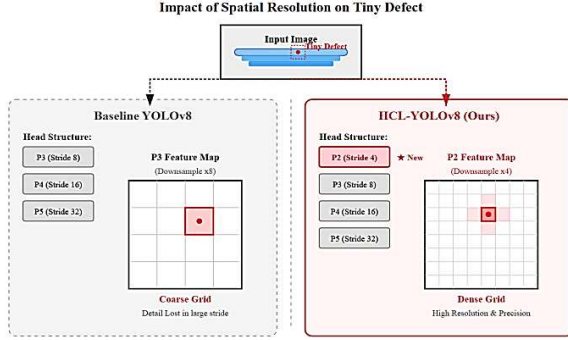


Fig. 5. Comparison of feature pyramid structures with the $P2$ detection layer

The $P2$ detection layer is a conventional enhancement for small-object detection. Its role here is to provide more reliable candidates for the subsequent hierarchical constraints, including HC-TAL and HCR-Loss/HSR. When feature resolution is insufficient, degraded candidate quality weakens the effect of structural priors.

3.4 Soft Boundary Prior Branch

Although the high-resolution detection layer helps capture small objects, the model may still produce false positives in complex backgrounds outside the insulator region. Therefore, this paper uses the annotated parent insulator string boxes to construct prior supervision and guide the network to focus on feature learning within insulator regions. Specifically, a lightweight prior branch is added beside the detection head to output a single-channel prior map P supervised by a soft mask G generated from parent target ground-truth boxes. The branch adopts a soft-boundary design: values in the central area inside the box approach 1 to strengthen supervision, while a buffer band is set near the boundary where supervision smoothly decays with the distance to the border, so as to tolerate annotation errors and instance deformation.

The prior branch loss $\mathcal{L}_{\text{prior}}$ is formulated as a weighted sum of the binary cross-entropy (BCE) loss and the Dice loss, as given in Eq. (3):

$$\mathcal{L}_{\text{prior}} = \mathcal{L}_{\text{bce}}(P, G) + \lambda_d \mathcal{L}_{\text{dice}}(P, G) \quad (3)$$

where:

$$\mathcal{L}_{\text{bce}}(P, G) = -\frac{1}{hw} \sum_{u=1}^h \sum_{v=1}^w (G_{uv} \log P_{uv} + (1 - G_{uv}) \log (1 - P_{uv})) \quad (4)$$

$$\mathcal{L}_{\text{dice}}(P, G) = 1 - \frac{2 \sum_{u,v} P_{uv} G_{uv} + \epsilon}{\sum_{u,v} P_{uv} + \sum_{u,v} G_{uv} + \epsilon} \quad (5)$$

Notation: h and w denote the spatial dimensions

of the prior map; λ_d is the weight of the Dice loss; and ϵ is a smoothing constant used to prevent division by zero. The SBP Branch serves only as an auxiliary supervision and does not affect the output of the main detection branch.

3.5 Hierarchy-Constrained Task-Aligned Label Assignment

To suppress false detections outside the insulator string caused by incorrect sample assignment, this study introduces a geometric containment prior on top of task-aligned label assignment. By filtering the candidate set through parent-target matching, positive sample selection is restricted to the insulator string region, thereby reducing pseudo-positive samples and training noise outside the string at the source.

(1) Parent Matching

For each defect ground-truth box b_j (with $c_j \in C_d$), its corresponding parent box b_j^p is selected from the set of parent ground-truth objects. When multiple parent candidates exist, priority is given to the one that contains the defect center and has the smallest area, in order to avoid incorrectly assigning the defect to a larger structural object. If no valid parent box is found, the defect sample follows the default assignment strategy.

(2) Margin Expansion

To account for annotation noise and prediction offsets, both the defect boxes and their corresponding parent boxes are slightly expanded. The expansion operator $\text{Expand}(\cdot)$ is defined as in Eq. (6). For a bounding box $b = (x_1, y_1, x_2, y_2)$, its width and height are given by $w = x_2 - x_1$ and $h = y_2 - y_1$, respectively. Let ρ denote the expansion ratio and δ the minimum pixel expansion:

$$\text{Expand}(b; \rho, \delta) = (x_1 - \Delta_x, y_1 - \Delta_y, x_2 + \Delta_x, y_2 + \Delta_y) \quad (6)$$

where $\Delta_x = \max(\rho w, \delta)$ and $\Delta_y = \max(\rho h, \delta)$. In this work, we set $\rho = 0.08$ for the expansion of parent boxes, and adopt a minimum expansion of $\delta = 2$ pixels at the $P2$ level to prevent degeneration of expansion for small boxes.

(3) Candidate Filtering

For defect classes, only candidate points that simultaneously fall inside both the expanded defect box and the expanded parent box are retained, as shown in Eq. (7).

$$\Omega_j = \{p_i \mid p_i \in \text{Expand}(b_j; \rho, \delta) \wedge p_i \in \text{Expand}(\bar{b}_j^p; \rho, \delta)\} \quad (7)$$

In addition, a geometric constraint based on the vertical projection of the parent box is introduced to further eliminate background points below the parent box, thereby suppressing interference from the background during defect learning.

(4) Task-Aligned Scoring and Positive Sample Selection

Within the filtered candidate set, top- k positive samples are still selected according to the task-aligned score,

as shown in Eq. (8).

$$t_{ij} = (p_{ij})^\alpha \cdot (\text{IoU}(\hat{b}_i, b_j))^\beta \quad (8)$$

where p_{ij} denotes the predicted probability of candidate point p_i belonging to class c_j ; $\hat{b}_i$ is the predicted bounding box regressed from this point; α and β are hyperparameters that balance the contributions of classification and localization; and top- k indicates selecting the k candidates with the highest scores as positive samples.

3.6 Small-Object-aware Normalized Wasserstein Distance loss

After completing structurally constrained label assignment, the instability of regressing tiny defect boxes still needs to be addressed. IoU-based losses are highly sensitive to slight boundary shifts in small-box scenarios and may cause gradient oscillation. To improve the localization stability of tiny defects, this study introduces a normalized Wasserstein distance regression term only for positive samples of defect classes.

The bounding box is modeled as a two-dimensional Gaussian distribution. For a box $b = (x, y, w, h)$, the mean is defined as $\mu = (x, y)$, and the covariance matrix is given by $\Sigma = \text{diag}(\sigma_x^2, \sigma_y^2)$, where $\sigma_x = w/2$ and $\sigma_y = h/2$. The second-order Wasserstein distance between the predicted box $\hat{b}$ and the ground-truth box b is defined as in Eq. (9).

$$W_2^2(\hat{b}, b) = |\hat{\mu} - \mu|_2^2 + |\hat{\Sigma}^{1/2} - \Sigma^{1/2}|_F^2 \quad (9)$$

The distance is further normalized and converted into a loss form:

$$\mathcal{L}_{nwd} = 1 - \exp\left(-\frac{D}{\tau}\right) \quad (10)$$

Where $D = \sqrt{W_2^2(\hat{b}, b)}$, and $\tau > 0$ is a scaling hyperparameter. In our implementation, SO-NWD is applied only to defect classes (class_id = 0/1), and its overall contribution is controlled by a weighting factor γ .

Considering that smaller defects are harder to regress, an inverse-area weight is introduced for defect samples:

$$w_{\text{small}} = \text{clip}\left(\frac{A_0}{A + \epsilon}, 1, w_{\text{max}}\right) \quad (11)$$

where $A = w \cdot h$ denotes the area of the ground-truth defect box, A_0 is a normalization constant (e.g., the batch-wise mean), and w_{max} is an upper bound. The final SO-NWD loss is defined as follows:

$$\mathcal{L}_{\text{sonwd}} = \gamma \cdot \text{mean}(w_{\text{small}} \cdot \mathcal{L}_{\text{nwd}}) \quad (12)$$

where $\text{mean}(\cdot)$ denotes the average over positive samples of the defect classes. In our implementation, $w_{\text{small}}[1:4]$ is clipped to 14 to prevent unstable gradients caused by extremely small bounding boxes.

3.7 Hierarchical Containment Regularization Loss

Although HC-TAL has reduced pseudo-positive samples outside the insulator string during label assignment,

the model may still generate out-of-bound predictions in the early training stage or under complex backgrounds. To explicitly reinforce the structural prior that a defect box should be contained within its parent box, this study introduces a Hierarchical Containment Regularization Loss (HCR-Loss), which is applied only to positive samples of defect classes with matched parent boxes.

Let the predicted defect bounding box be $\hat{b} = (\hat{x}_1, \hat{y}_1, \hat{x}_2, \hat{y}_2)$, and the expanded parent ground-truth box be $\tilde{b}^p = (\tilde{x}_1^p, \tilde{y}_1^p, \tilde{x}_2^p, \tilde{y}_2^p)$. We define four boundary violation terms as follows:

$$\Delta_l = \tilde{x}_1^p - \hat{x}_1, \Delta_t = \tilde{y}_1^p - \hat{y}_1, \Delta_r = \hat{x}_2 - \tilde{x}_2^p, \Delta_b = \hat{y}_2 - \tilde{y}_2^p \quad (13)$$

When $\hat{b}$ lies entirely inside $\tilde{b}^p$, the above boundary violation terms are non-positive (or close to zero); otherwise, they become positive when violations occur. To obtain smooth gradients and avoid the training instability caused by hard truncation, we adopt the Softplus function $\phi(z) = \ln(1 + \exp(z))$ to impose a smooth penalty on the violation terms, leading to the per-sample containment loss defined as follows:

$$\mathcal{L}_{\text{hcr}} = \phi(\Delta_l) + \phi(\Delta_t) + \phi(\Delta_r) + \phi(\Delta_b) \quad (14)$$

To mitigate the negative impact of mismatched parent boxes during training, the HCR-Loss introduces a quality gating coefficient $q \in [0, 1]$, which modulates the regularization strength based on the target quality score (e.g., target score) obtained from the label assignment stage:

$$\mathcal{L}_{\text{hcr}} \leftarrow q \cdot \mathcal{L}_{\text{hcr}} \quad (15)$$

Meanwhile, the weight of the HCR-Loss is scheduled using a quadratic ramp that gradually increases over training epochs, in order to prevent instability caused by noisy predictions in the early stages. Let the total number of training epochs be T , and the ramp start and end epochs be defined as $e_0 = 0.4T$ and $e_1 = 0.8T$, respectively. Let w_{hmax} denote the maximum weight. Then,

$$w_h(e) = \begin{cases} 0, & e < e_0 \\ w_{\text{hmax}} \cdot \left(\frac{e - e_0}{e_1 - e_0}\right)^2, & e_0 \leq e \leq e_1 \\ w_{\text{hmax}}, & e > e_1 \end{cases} \quad (16)$$

where e denotes the current training epoch. In our implementation, a quadratic ramp is adopted with $w_{\text{hmax}} = 0.6$.

3.8 Hierarchy-Aware Soft Re-scoring Post-processing

In inspection scenarios, high-confidence defects outside the insulator string usually indicate false alarms. However, directly deleting such defect boxes may mistakenly suppress real defects when the parent target is missed. To balance false-positive suppression and recall preservation, this study introduces Hierarchy-aware Soft Re-scoring (HSR) after Non-Maximum Suppression (NMS): for each defect box, it checks whether its center point falls within the expanded region of any predicted parent box; if not, its confidence score is scaled down rather than the box

being removed.

Specifically, let the set of predicted parent targets be

$$\widehat{\mathcal{B}}^p = \left\{ \widehat{b}_k^p \mid \widehat{c}_k = 2, \widehat{s}_k \geq \delta_s \right\} \quad (17)$$

where δ_s denotes the confidence threshold for parent objects. For each defect prediction $(\widehat{b}_i, \widehat{c}_i, \widehat{s}_i)$ with $\widehat{c}_i \in \mathcal{C}_d$, its center point is defined as $m_i = (\widehat{x}_i, \widehat{y}_i)$. If

$$m_i \notin \bigcup_k \text{Expand}(\widehat{b}_k^p; 0, m) \quad (18)$$

then

$$\widehat{s}_i \leftarrow \eta \cdot \widehat{s}_i \quad (19)$$

where m denotes the pixel-level expansion margin used for tolerance, and $\eta \in (0,1)$ is the score attenuation factor. In our implementation, we set $\delta_s = 0.15m = 8$ pixels, and $\eta = 0.4$. When no parent object is detected in an image, HSR is not applied to preserve recall.

3.9 Overall Loss Function

Combining the above improvements, the total loss in the training stage is

$$\mathcal{L} = \mathcal{L}_{yolo} + \lambda_p \mathcal{L}_{prior} + \lambda_n \mathcal{L}_{sonwd} + w_h(e) \mathcal{L}_{hcr} \quad (20)$$

where $\mathcal{L}_{yolo}$ denotes the YOLOv8n detection loss, including classification and regression/distribution terms, and λ_p and λ_n are hyperparameters. The function $w_h(e)$ represents the epoch-adaptive weight of the HCR-Loss. During inference, HSR does not modify the training architecture and is applied only to confidence refinement after NMS, thereby ensuring structural consistency and mitigating deployment-side risks.

4 Experimental Results and Analysis

4.1 Experimental Dataset

This paper is based on the insulator defect detection dataset published by Zheng et al. in Sensors, and expands the data scale using the defect-sensitive stratified offline augmentation strategy proposed in Section 2. After augmentation, the dataset contains 10,621 images while strictly preserving the original train/val/test split so as to avoid contamination of the validation distribution by the augmentation strategy. The dataset includes three categories: flashover, broken, and insulator string (parent class). This annotation system exhibits a clear hierarchical relationship: defect subclasses usually lie inside or near the parent insulator string region, providing prior knowledge for subsequent hierarchical consistency learning and inference-time risk suppression.

Fig. 6 shows the annotation distribution of the augmented training set. The results indicate that the bounding boxes of the two defect classes are significantly smaller than those of insulator strings, showing typical small-object characteristics. In addition, the spatial distribution of targets is relatively concentrated in the middle of the image, reflecting the practical difficulty that small defects

are prone to missed detection and interference under complex backgrounds.

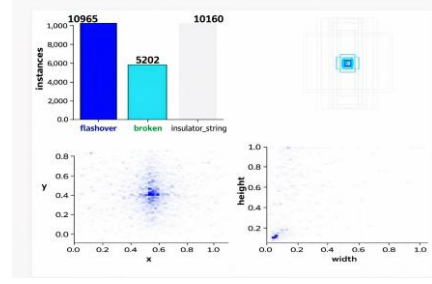


Fig. 6. Statistics of sample distribution in the dataset

4.2 Training Framework and Hyperparameters

The proposed method was implemented based on the Ultralytics YOLOv8n training framework and trained from scratch (pretrained=False). The input size was set to 800, the model was trained for 150 epochs, and the batch size was 8. Conventional online augmentation strategies were adopted, including Mosaic and MixUp, with `close_mosaic = 20` and `mixup = 0.1`. Both the ablation study and the baseline comparison experiments were conducted under the same dataset, input scale, and training-epoch settings to ensure experimental fairness and reproducibility.

4.3 Evaluation Metrics

This study adopted standard object detection metrics, including Precision, Recall, F1-score, and COCO-style AP/mAP. Let TP, FP, and FN denote the numbers of true positives, false positives, and false negatives, respectively:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (21)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (22)$$

$$F1 = 2 \cdot \text{Precision} \cdot \text{Recall} / (\text{Precision} + \text{Recall}) \quad (23)$$

AP is defined as the area under the Precision-Recall curve. mAP@0.5 denotes the mean AP of all categories at an IoU threshold of 0.5, whereas mAP@0.5:0.95 denotes the mean AP averaged over IoU thresholds from 0.5 to 0.95 with a step size of 0.05, which provides a more comprehensive evaluation of localization accuracy and stability.

4.4 Analysis of Training Convergence and Threshold Selection

4.4.1 Convergence and Stability

Fig. 7 shows the training curves of the proposed method. The loss decreases rapidly in the early stage and then gradually stabilizes. Precision, Recall, and mAP converge steadily in the middle and late stages, without obvious oscillation or overfitting, indicating that the training process is stable and reliable.

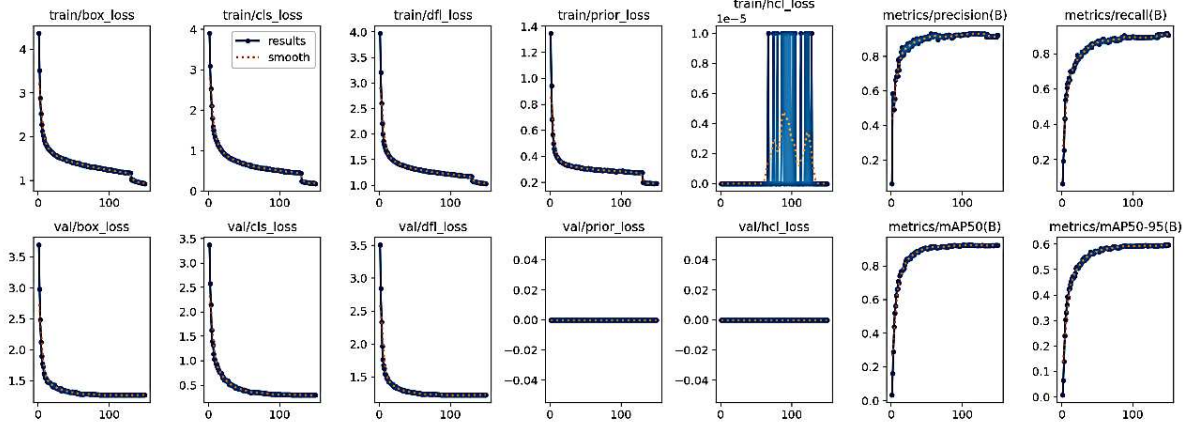


Fig. 7. Training curves of the proposed scheme

4.4.2 Confidence Threshold and Precision-Recall Trade-off

Fig. 8 presents the PR curve of the proposed method at IoU = 0.5, with an overall mAP of 0.923. The AP values of flashover, broken, and insulator string are 0.845, 0.928, and 0.995, respectively.

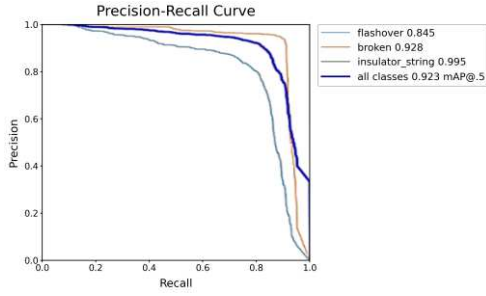


Fig. 8. PR curve

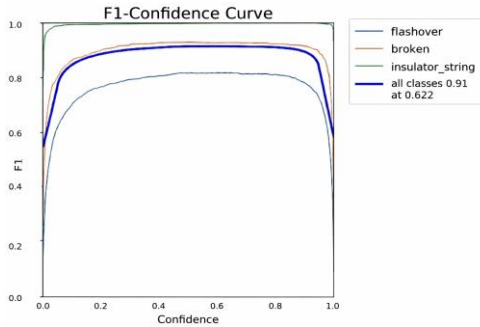


Fig. 9. F1-Confidence curve

The F1-Confidence curve in Fig. 9 shows that the best overall F1-score, approximately 0.91, is achieved at a confidence threshold of around 0.622. This value is therefore recommended as the default operating point for engineering deployment to balance false alarms and missed detections. Threshold selection should be aligned with application objectives. When higher Recall is preferred, the threshold may be appropriately lowered and used together with the soft score-reduction mechanism of HSR to control false positives outside the insulator string. When

higher Precision is emphasized, the threshold may be increased, while the structural prior imposed by HSR can still effectively suppress high-confidence false positives outside the insulator string. This threshold analysis provides a unified basis for subsequent evaluation and also supports the robustness of the HSR soft strategy in deployment scenarios.

4.5 Ablation Study

As shown in Table 1, from the perspective of parameter count and computational cost, introducing the P2 detection head in ABL_01 increases the number of parameters from 3.01M to 3.08M and GFLOPs from 8.1 to 12.5, representing the necessary computational overhead for improving small-object detection. After adding the SBP-Branch in ABL_02, the parameter count and GFLOPs further increase slightly to 3.12M and 12.6, respectively, indicating that the additional overhead remains controllable. The subsequent modules, namely HC-TAL, SO-NWD, HCR-Loss, and HSR, are all optimization strategies applied during training or post-processing and do not alter the inference architecture. Accordingly, the number of parameters and GFLOPs remain unchanged, reflecting the design principle of optimizing training without increasing deployment burden.

In terms of detection accuracy (mAP@0.5), P2 and the SBP-Branch progressively improve mAP@0.5 from 0.908 to 0.918. After HC-TAL is introduced, a slight fluctuation is observed (0.918 $\rightarrow$ 0.917), which is consistent with the expectation that stricter sample selection may temporarily affect Recall. Subsequently, SO-NWD and HCR-Loss raise mAP@0.5 to 0.922, and HSR further increases it to 0.923, resulting in an overall upward trend.

In terms of localization accuracy (mAP@0.5:0.95), P2 and the SBP-Branch each bring an improvement of 0.005, indicating that high-resolution features and regional priors contribute to finer localization. After HC-TAL is introduced, mAP@0.5:0.95 decreases from 0.606 to 0.594, a drop of 0.012. This change directly reflects the reduction in localization diversity caused by a tighter positive-sam-

ple region and is consistent with the expected trade-off between higher sample purity and lower localization richness. SO-NWD and HCR-Loss then gradually compensate for this degradation, and HSR further restores the

metric to 0.599, which is slightly above the baseline, suggesting a complementary effect among the modules.

Table 1. Ablation study

ID	Model Configuration	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Params(M)	GFLOPs
ABL_00	YOLOv8n	0.915	0.877	0.908	0.596	3.01	8.1
ABL_01	YOLOv8n+P2 Head	0.915	0.887	0.911	0.601	3.08	12.5
ABL_02	YOLOv8n+P2 Head+SBP-Branch	0.922	0.878	0.918	0.606	3.12	12.6
ABL_03	YOLOv8n+P2 Head+SBP-Branch+HC-TAL	0.926	0.887	0.917	0.594	3.12	12.6
ABL_04	YOLOv8n+P2 Head+SBP-Branch+HC-TAL+SO-NWD	0.917	0.900	0.918	0.596	3.12	12.6
ABL_05	YOLOv8n+P2 Head+SBP-Branch+HC-TAL+SO-NWD+HCR-Loss	0.919	0.910	0.922	0.595	3.12	12.6
ABL_06	YOLOv8n+P2 Head+SBP-Branch+HC-TAL+SO-NWD+HCR-Loss+HSR	0.923	0.911	0.923	0.599	3.12	12.6

In terms of Precision and Recall, introducing P2 in ABL_01 increases Recall from 0.877 to 0.887, indicating that high-resolution features are beneficial for detecting small objects. After adding the SBP-Branch in ABL_02, Precision increases to 0.922, whereas Recall decreases slightly from 0.887 to 0.878, suggesting that the regional prior suppresses background-induced false positives but may also filter a small number of marginal defects. After HC-TAL is introduced in ABL_03, Precision further rises to 0.926 and Recall recovers to 0.887, indicating that the hierarchy-constrained sample selection strategy reduces false detections without excessively sacrificing Recall. After SO-NWD is added in ABL_04, Recall improves markedly to 0.900, while Precision declines from 0.926 to 0.917, suggesting that the small-object-friendly regression loss improves recall at the cost of a limited increase in false positives. After adding HCR-Loss in ABL_05, Recall reaches 0.910, the highest value at this stage, while Precision remains relatively stable at 0.919, suggesting that structural-consistency regularization improves the

learning quality of positive samples and brings a clear gain in Recall. After HSR is introduced in ABL_06, Precision rises to 0.923 and Recall is adjusted to 0.911, ultimately yielding a more balanced trade-off between Precision and Recall, which is consistent with the design objective of softly suppressing false positives outside the insulator string during inference.

Overall, the ablation results exhibit a clear progressive logic. P2 and the SBP-Branch establish the feature and prior foundation; HC-TAL improves sample purity; SO-NWD and HCR-Loss compensate for Recall and localization performance from the perspectives of regression stability and structural consistency, respectively; and HSR completes the optimization of the Precision-Recall trade-off during inference. The observed fluctuations in all metrics remain within a reasonable range and collectively support the effectiveness and complementarity of the proposed modules within the hierarchical consistency learning framework.

Table 2. Comparison with mainstream detectors

ID	Model	Precision	Recall	mAP@0.5	mAP@0.5:0.95	Params(M)	GFLOPs
ABL_00	YOLOv8n	0.915	0.877	0.908	0.596	3.01	8.1
ABL_06	Ours	0.923	0.911	0.923	0.599	3.12	12.6
ABL_07	YOLOv5n	0.908	0.864	0.903	0.58	2.50	7.1
ABL_08	YOLOv10n	0.906	0.855	0.89	0.596	2.27	6.5
ABL_09	YOLOv11n	0.917	0.889	0.919	0.609	2.58	6.3
ABL_10	YOLOv8s	0.918	0.893	0.91	0.603	11.12	28.4
ABL_11	YOLOv11s	0.914	0.896	0.909	0.614	9.41	21.3

4.6 Comparison with Mainstream Lightweight Detectors

To evaluate the general competitiveness of the method, the proposed scheme is compared under the same training protocol with several mainstream lightweight detectors, including YOLOv5n, YOLOv10n, YOLOv11n, and the larger yet still lightweight YOLOv8s and YOLOv11s.

The comparison results in Table 2 show that, while maintaining a lightweight scale of 3.12M parameters and 12.6 GFLOPs, the proposed method achieves an mAP@0.5 of 0.923 and a Recall of 0.911, outperforming all comparison models in these two metrics and demonstrating strong capability in detecting small defects. Although its mAP@0.5:0.95 of 0.599 is slightly lower than that of YOLOv11n (0.609) and YOLOv11s (0.614), the

parameter count and computational cost are reduced by about 70% and 50%, respectively. Overall, the proposed method achieves a better engineering trade-off between defect detection capability and computational overhead.

Fig. 10 shows several visualization results of the improved algorithm in insulator defect detection. The results indicate that HCL-YOLOv8 accurately localizes insulator string regions and effectively detects the two tiny defect categories, flashover and broken, under complex backgrounds such as similar textures, illumination variation, and partial occlusion. The predicted boxes align well with the ground-truth targets. In the illustrated examples, few obvious missed detections or false positives outside the insulator string are observed.



Fig. 10. Detection results of the improved algorithm

Conclusion and Future Work

To address the challenges of unstable localization of tiny targets, strong interference from complex backgrounds, and insufficient utilization of hierarchical structural priors in transmission-line insulator defect detection, this study proposed HCL-YOLOv8, a lightweight detection framework enhanced by hierarchical consistency learning. Built upon YOLOv8n, the proposed method introduces hierarchical constraints into sample assignment, training loss, and inference post-processing around the natural hierarchical structure between defects and insulator strings. Specifically, HC-TAL restricts positive samples to the parent string region, thereby suppressing false positives outside the string at the source; HCR-Loss reinforces the geometric containment relation between defect boxes and parent boxes during training; and HSR softly

reduces the confidence of defect predictions outside the string during inference, so as to balance false-positive suppression and recall preservation. In addition, a P2 high-resolution detection head, the SBP-Branch, and SO-NWD are incorporated to further improve localization stability and detection capability for tiny defects.

On the public dataset published by Zheng et al. in *Sensors* and its augmented version, the proposed method achieves an mAP@0.5 of 0.923 and a Recall of 0.911 with a lightweight scale of 3.12M parameters and 12.6 GFLOPs. Compared with the YOLOv8n baseline, Recall improves by 3.4 percentage points and mAP@0.5 by 1.5 percentage points, with particularly notable gains for small flashover defects. The ablation study verifies the effectiveness and complementarity of the proposed modules, while comparison with mainstream lightweight detectors

indicates that the proposed method provides a more favorable engineering trade-off between defect detection capability and computational cost. Overall, the results suggest that the proposed framework offers stable advantages for small-object defect detection and shows practical potential for engineering deployment under lightweight constraints.

References

- [1] Zheng J, Wu H, Zhang H, Wang Z, Xu W. Insulator-defect detection algorithm based on improved YOLOv7. *Sensors* 2022; 22: 8801.
- [2] Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell* 2017; 39(6): 1137-1149.
- [3] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single shot multibox detector. In: *Computer Vision – ECCV 2016*. Springer; 2016: 21-37.
- [4] Tian Z, Shen C, Chen H, He T. FCOS: Fully convolutional one-stage object detection. In: *Proc ICCV*; 2019: 9627-9636.
- [5] Feng C, Zhong Y, Gao Y, Scott MR, Huang W. TOOD: Task-aligned one-stage object detection. In: *Proc ICCV*; 2021: 3490-3499.
- [6] Zhou X, Wang D, Krähenbühl P. Objects as points. *arXiv:1904.07850*.
- [7] Zhou M, Wang J, Li B. ARG-Mask RCNN: An infrared insulator fault-detection network based on improved Mask RCNN. *Sensors* 2022; 22: 4720.
- [8] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: *Proc CVPR*; 2016: 779-788.
- [9] Bochkovskiy A, Wang CY, Liao HY. YOLOv4: Optimal speed and accuracy of object detection. *arXiv:2004.10934*.
- [10] Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: Exceeding YOLO series in 2021. *arXiv:2107.08430*.
- [11] Wang CY, Bochkovskiy A, Liao HY. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proc CVPR*; 2023: 7464-7475.
- [12] Lin TY, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proc CVPR*; 2017: 936-944.
- [13] Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: *Proc CVPR*; 2018: 8759-8768.
- [14] Tan M, Pang R, Le QV. EfficientDet: Scalable and efficient object detection. In: *Proc CVPR*; 2020: 10778-10787.
- [15] Zhang S, Chi C, Yao Y, Lei Z, Li SZ. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In: *Proc CVPR*; 2020: 9756-9765.
- [16] Wang J, Chang X, Yang W, Yu L. A normalized Gaussian Wasserstein distance for tiny object detection. *arXiv:2110.13389*.
- [17] Kim K, Lee H. Probabilistic anchor assignment with IoU prediction for object detection. *arXiv:2007.08103*.
- [18] Ge Z, Liu S, Li Z, Yoshie O, Sun J. OTA: Optimal transport assignment for object detection. In: *Proc CVPR*; 2021: 303-312.
- [19] Li X, Wang W, Wu L, Chen S, Hu X, Li J, Tang J, Yang J. Generalized focal loss: Learning qualified and distributed bounding boxes for dense object detection. *arXiv:2006.04388*.
- [20] Zheng Z, Wang P, Liu W, Li J, Ye R, Ren D. Distance-IoU loss: Faster and better learning for bounding box regression. In: *Proc AAAI*; 2020: 12993-13000.
- [21] Zhang H, Wang Y, Dayoub F, Sünderhauf N. VarifocalNet: An IoU-aware dense object detector. In: *Proc CVPR*; 2021: 8510-8519.
- [22] Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. mixup: Beyond empirical risk minimization. *arXiv:1710.09412*.
- [23] Yun S, Han D, Chun S, Oh SJ, Yoo Y, Choe J. CutMix: Regularization strategy to train strong classifiers with localizable features. In: *Proc ICCV*; 2019: 6022-6031.
- [24] DeVries T, Taylor G. Improved regularization of convolutional neural networks with Cutout. *arXiv:1708.04552*.
- [25] Cubuk ED, Zoph B, Mane D, Vasudevan V, Le QV. AutoAugment: Learning augmentation strategies from data. In: *Proc CVPR*; 2019: 113-123.
- [26] Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA. Albumentations: Fast and flexible image augmentations. *Information* 2020; 11: 125.

Author's information

Lei Wang (b. 1991), Bachelor, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include artificial intelligent and data analysis. Email: moshoushijie237@163.com

Jiamao Han (b. 1993), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include artificial intelligent and data analysis. Email: jjamaohan@163.com

Yaxing Wei (b. 1996), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include heuristic optimization algorithm, artificial intelligent. Email: wyx179024@163.com

Bin Zhao (b. 1997), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include cyberspace Security, network anomaly detection. Email: 18804909045@163.com

Jing Xu (b. 1988), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include computer vision, Artificial Intelligent. Email: 724243882@qq.com

Lei Han (b. 1981), Bachelor, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include Big data, Unmanned Aerial Vehicle. Email: 1196594994@qq.com

Yu Zhang (b. 1980), Bachelor, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include Cloud Computing, Distributed system. Email: 0471zhangyu@163.com

Yingchao Zhen (b. 1989), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests include Internet of Things, Natural Language Processing. Email: 15247159089@163.com

E Erdun (b. 1984), Master, is an engineer at Inner Mongolia electric power digital research company (hohhot, China). Research interests

include artificial intelligence applications, differential privacy, and distribution automation. Email: eerdun8@163.com

Lai Man (b. 1982), Bachelor, is an engineer at Inner Mongolia Electric Power Digital Research Company (Hohhot, China). Research interests include network security, automated operation and maintenance. Email: stefanka@qq.com

Jian Wei (b. 1989), Bachelor, is an engineer at Inner Mongolia Electric Power Digital Research Company (Hohhot, China). Research interests include Pattern Recognition, Intelligent Perception. Email: 272643132@qq.com

Liguo Qu (b. 1979), PhD, is an Associate Professor at Anhui Normal University (Wuhu, China). Research interests include intelligent information processing, detection technology and automation devices. Email: lgqu@ahnu.edu.cn