remote sensing

MDPI

Article

Enhanced DINO for Cross-Domain Transmission Tower Detection Using Remote Sensing Images

Junjie Wang [1], Jiahe Tian [1], Lingli Zhao [1,*], Ningsheng Liao [2], Jie Yang [3] and Lei Shi [3]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; 2025206190034@whu.edu.cn (J.T.)
[2] School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401349, China
[3] State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China
* Correspondence: zhaolingli@whu.edu.cn

Highlights

What are the main findings?

- This study addresses cross-domain transmission tower detection under geographic domain shift, where models trained in one region suffer performance degradation when applied to remote sensing images from another region.
- An enhanced DINO framework that integrates Query-level Objectness Gating, MPDIoU regression and localisation-aware confidence calibration improves detection accuracy with limited target-domain annotation.

What are the implications of the main findings?

- The proposed framework improves the robustness of transmission tower detection across regions with different tower morphologies, background textures and imaging conditions.
- The results demonstrate that lightweight detector-level adaptations can enhance cross-domain remote sensing object detection without substantially increasing the inference burden.

Abstract

Transmission towers are fundamental components of electric power networks. However, their structure, scale and background textures vary substantially across remote sensing images acquired from different geographic regions. These discrepancies often reduce the detection accuracy of a model trained in one region when it is applied to another region. This paper proposes an enhanced DINO-based framework for cross-domain transmission tower detection that incorporates three lightweight optimisation modules. First, a Query-level Objectness Gating (QOG) module adaptively reweights decoder queries by estimating per-query objectness scores, thereby suppressing background-dominated queries. Second, MPDIoU regression is used to improve the localisation accuracy of elongated transmission tower targets. Third, a Quality-aware Scoring Module (QSM) calibrates classification confidence using predicted localisation-quality logits, thereby reducing high-confidence false detections caused by poor box alignment. Experiments are conducted on two remote sensing image datasets from different geographic regions. Under the 10% target-domain annotation setting, the proposed method achieves a precision of 0.8947, a recall of 0.8199, an F1-score of 0.8556 and an mAP@0.5 of 0.8684, outperforming the original DINO baseline and mainstream detectors including YOLOv8, YOLOv9 and YOLOv11. The results demonstrate that the proposed framework provides an effective solution for robust cross-domain detection of slender transmission tower targets in remote sensing images.

Check for updates

Academic Editor: Xiangtao Zheng

Received: 27 May 2026
Revised: 27 June 2026
Accepted: 29 June 2026
Published: 5 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: transmission tower detection; cross-domain target detection; query-level objectness gating; MPDIoU; DINO

1. Introduction

The safe and stable operation of electric power infrastructure is vital to national economic stability and social functioning, as it directly affects energy supply and public safety. As core structural elements of power transmission networks, transmission towers are widely distributed across diverse geographic environments, including mountains, farmlands, towns and cities. Efficient and accurate tower detection and monitoring provide a critical foundation for power-grid inspection, disaster early warning and digital infrastructure management. With the rapid development of high-resolution remote sensing satellites and unmanned aerial vehicle imaging technology, automated transmission tower inspection based on visible-light imagery has become an important research focus at the intersection of computer vision and remote sensing.

Existing methods for transmission tower detection in remote sensing images primarily include traditional template-matching approaches [1] and methods that combine manual feature extraction with shallow machine-learning classifiers [2]. These conventional approaches often rely on handcrafted features and therefore exhibit limited robustness and generalisation in complex environments. With the advancement of deep learning, deep neural network-based detection methods have become dominant in this field because of their strong automatic feature-extraction capability. Ma et al. [3] achieved instance segmentation of transmission towers in aerial images using an enhanced Solov2 network and introduced a path-aggregated feature pyramid (PaFPN) to improve multi-scale feature fusion. Related studies have also explored deep learning-based joint detection methods for transmission towers and power lines, optimising the loss function and model architecture to improve detection integrity [4]. Nevertheless, these methods generally assume that the training and test datasets follow the same distribution. In practice, however, the morphological structure, scale distribution, background texture and imaging conditions of transmission towers may vary substantially when models are deployed across different regions. Context bias between the training and test sets, specifically excessive model reliance on foreground–background correlations in the source domain, has been shown to significantly impair the cross-domain generalisation capability of target detectors [5]. Son et al. [5] found, through experiments involving background masking and feature perturbation, that convolutional neural networks learn statistical associations between foreground and background during training. These associations subsequently pose a substantial challenge to model generalisation in scenarios characterised by domain bias.

Cross-domain target detection aims to use richly labelled source domains to improve detection performance in unlabelled or weakly labelled target domains. It has therefore attracted considerable attention in recent years. Zhou and Wang [6] surveyed domain-adaptive object detection and categorised existing methods into feature alignment, adversarial training, reconstruction-based methods, knowledge distillation and other emerging paradigms. Among these methods, adversarial learning-based domain adaptation methods learn domain-invariant representations through adversarial training between domain discriminators and feature extractors, typically using a gradient reversal layer [7]. Pseudo-label-based self-training methods iteratively refine the model by generating target-domain pseudo-labels, and teacher–student frameworks have recently been used to improve pseudo-label reliability [8]. Image generation-based methods achieve

pixel-level adaptation by transforming source-domain images into the style of the target domain with generative adversarial networks [9].

Most existing studies on cross-domain detection focus on rigid targets, including aircraft, ships and vehicles. For Syn2Real object detection, Guo et al. [8] introduced a soft adaptive teacher method to improve adaptation from synthetic to real data. For maritime ship targets, a detection and recognition system has been deployed in real maritime surveillance scenarios [10]. In the broader context of general-purpose domain-generalised object detection, Danish et al. [11] proposed perturbing dominant low-level feature modes to simulate domain shifts in feature space and improve detector generalisation to unseen domains. However, the target types considered in these studies generally have relatively consistent morphological structures. For example, fuselage and wing configurations are fixed in aircraft, ship hulls share similar contours, and vehicle shapes are constrained by specific models. As a result, cross-domain variations are primarily reflected in background styles, lighting conditions and other low-level visual features.

In contrast to the aforementioned targets, transmission towers exhibit more complex morphological variations across different regions. Significant discrepancies exist in the design standards and structural forms of transmission towers across areas. Cat-head towers and wine-glass towers predominate in the eastern plains of China and are characterised by wide structures; dry-shaped and compact towers are primarily used in western regions and have longer, more slender designs; and notable differences also exist between tower types used in urban and rural power grids. In addition, the contexts in which transmission towers are distributed vary considerably, ranging from the farmland and lakes of Jiangnan water towns to the deserts and plateau mountains of north-western China. Variations in background texture and scene structure therefore far exceed those typically encountered in general cross-domain target detection. Furthermore, as slender targets, transmission towers impose higher requirements for bounding-box localisation accuracy than general-purpose targets. For elongated targets such as transmission towers, conventional IoU-based regression losses may provide insufficient geometric constraints under large aspect-ratio variations, motivating improved bounding-box regression losses such as MPDIoU [12].

To address these challenges, this paper uses DINO [13] as the baseline detector and introduces three lightweight improvements for cross-domain transmission tower detection: Query-level Objectness Gating (QOG) to suppress background-biassed decoder queries, MPDIoU to improve bounding-box regression for slender towers and the Quality-aware Scoring Module (QSM) to calibrate classification confidence with localisation quality. The proposed framework aims to improve cross-region generalisation while keeping the inference cost close to that of the original DINO model.

2. Materials and Methods

This section describes the materials, cross-domain dataset construction and enhanced DINO-based method used in this study. It first clarifies the public TT dataset, namely the power transmission and distribution infrastructure imagery dataset released through the Duke Energy Initiative Data+ project, and the self-annotated DNP dataset, namely the Jiaxing–Huzhou target-domain transmission tower dataset constructed from Google Earth imagery. It then introduces the overall framework, DINO baseline, Query-level Objectness Gating (QOG), MPDIoU regression loss, Quality-aware Scoring Module (QSM), implementation settings and evaluation metrics.

2.1. Overall Framework

The overall framework of the proposed method is shown in Figure 1. The main detection network is based on DINO. The framework follows DINO's contrastive denoising training strategy and adopts hybrid query selection within a two-stage inference paradigm. Three enhancements are incorporated into the framework. First, the QOG module is applied at the decoder-query level in the first two decoder layers to dynamically rescale query embeddings according to objectness-aware gating coefficients. Second, the MPDIoU loss function completely replaces the original GIoU loss to strengthen vertex-alignment constraints in bounding-box regression. Third, an inference scoring mechanism is introduced. Notably, these three enhancements are mutually compatible, are lightweight and introduce negligible additional inference overhead.

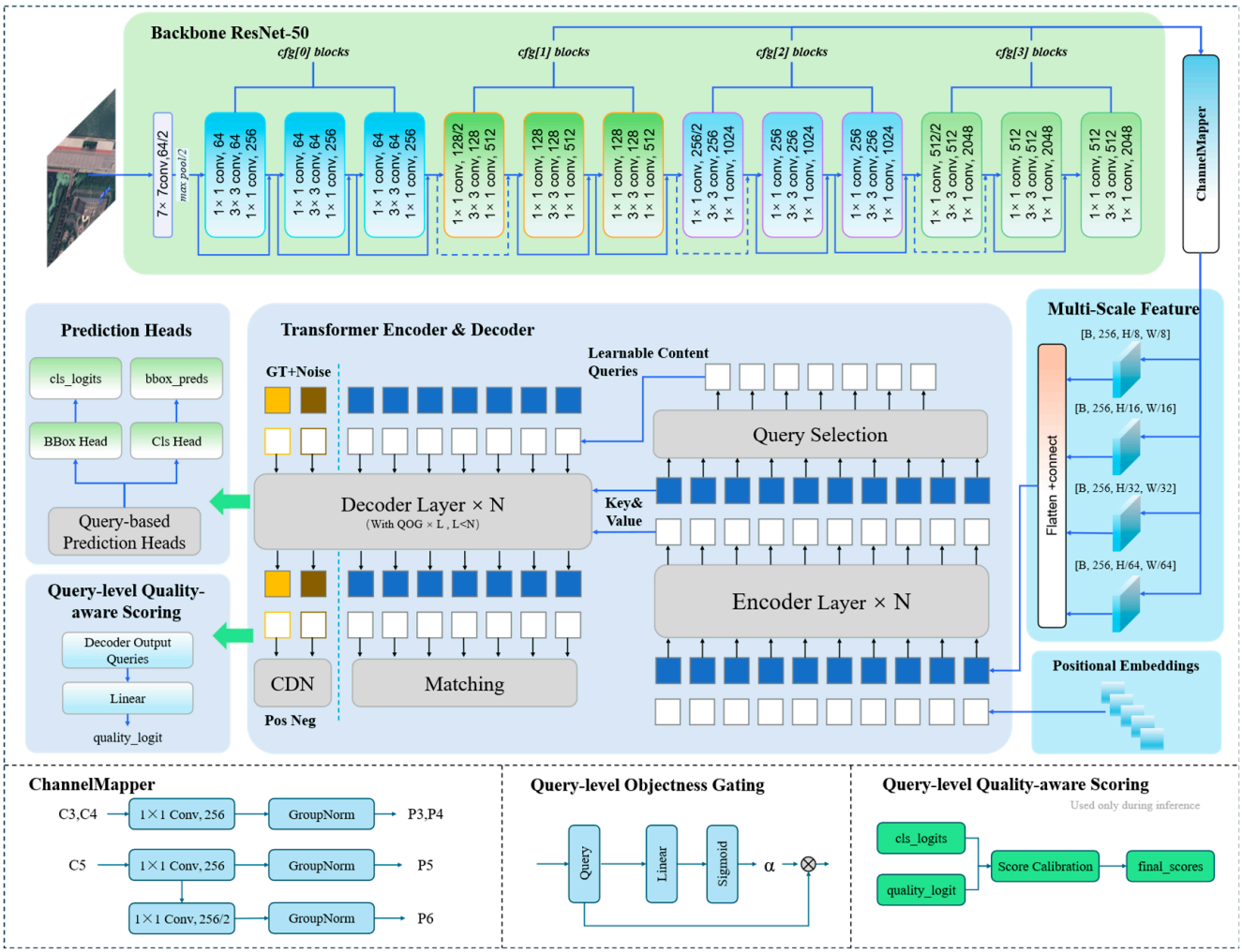


Figure 1. Schematic diagram of the improved model.

To alleviate query interference caused by complex backgrounds and cross-domain appearance variations, this paper introduces the QOG module into the Transformer decoder of DINO. Unlike feature-level domain alignment methods, QOG operates directly on decoder queries. For each decoder query, a lightweight linear layer followed by sigmoid activation predicts an objectness-aware gating coefficient. The coefficient is then multiplied by the corresponding query embedding, enabling the model to suppress queries that are weakly related to transmission towers or strongly affected by background textures. In this way, object-relevant queries are preserved and noisy queries are weakened before subsequent decoding and prediction.

2.2. DINO Baseline

DINO is a recent advancement in DETR-like detectors proposed by Zhang et al. [13] in 2023. This model incorporates three key techniques derived from Deformable DETR [14]. First, contrastive denoising training, which is closely related to the query denoising strategy introduced in DN-DETR [15], generates corrupted queries by adding noise to the ground-truth bounding boxes during training. This strategy enables the decoder to learn to recover the original box positions from noisy inputs, thereby significantly accelerating convergence and improving localisation robustness. Second, hybrid query selection uses the content features of the encoder output together with a learnable anchor-box prior for query initialisation, ensuring that the initial queries contain both semantic information and spatial priors. Finally, two-stage inference uses the decoder's intermediate outputs as candidate boxes for preliminary screening, which are then refined by a secondary detection head for further optimisation. In this study, we adopt the official implementation of DINO from the MMDetection library. The backbone network is ResNet-50, and the neck channel mapper projects multi-scale features to 256 dimensions. Both the encoder and decoder consist of six layers. The detection head includes a classification branch and a bounding-box regression branch. The original regression loss is defined as a weighted combination of L1 loss and GIoU loss [16], with weights of 5.0 and 2.0, respectively. The classification loss is defined as Focal Loss, and its corresponding weight is 2.0. Pretrained weights are initialised from the official DINO model trained on the COCO dataset, followed by full-parameter fine-tuning on the power tower cross-domain dataset.

2.3. Query-Level Objectness Gating

In the DINO decoder, each layer employs query vectors to aggregate target-related information from the encoder feature maps through a cross-attention mechanism. However, not all queries are strongly associated with real objects in the early decoding layers. Zhao et al. [17] highlighted, in their study of cross-modal visual localisation, that intermediate queries produced by the first two decoder layers of a DETR-like model have low confidence in their predicted frame locations and ambiguous semantic representations. These queries compete indiscriminately for attention with high-quality queries in subsequent layers, which significantly dilutes the model's attention weights on the actual target region.

Query modulation methods can be broadly categorised into two approaches. The first category emphasises screening during query initialisation. Representative works include Deformable DETR [14], which initialises queries by selecting the top K positions as reference points based on classification scores generated by the encoder, and DINO [13], which introduces a hybrid query selection method that integrates anchor-frame priors with encoder content features. The second category focuses on dynamic feature recalibration, in which feature responses are adaptively reweighted according to their estimated importance. However, these approaches still have limitations for the present task: query initialisation filtering is mainly effective at the decoder entrance and does not explicitly modulate intermediate queries as they evolve through subsequent decoder layers. In addition, feature-level recalibration does not directly address query-level background interference in DETR-like decoders. Therefore, applying Query-level Objectness Gating to dynamically screen intermediate decoder queries represents a targeted design for cross-domain transmission tower detection.

In this paper, we introduce a Query-level Objectness Gating (QOG) module into the DINO decoder and adapt it to cross-domain transmission tower detection. As illustrated in Figure 2, QOG is inserted between adjacent decoder layers and operates on decoder queries. Its core idea is to dynamically generate objectness-aware gating coefficients from

decoder queries and use them to rescale the corresponding query embeddings. In this way, low-objectness or background-biassed queries are suppressed, while object-relevant queries are preserved for subsequent decoding and prediction.

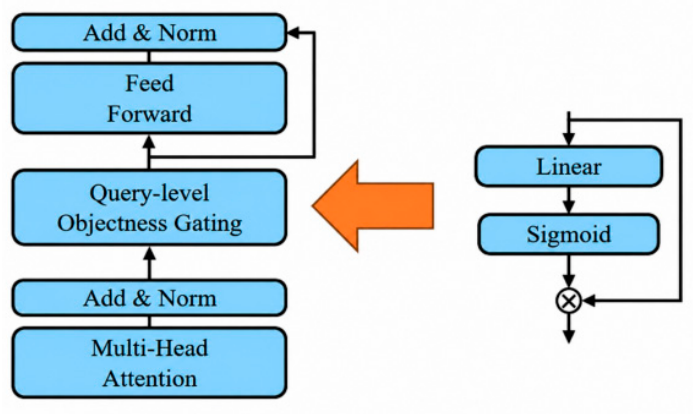


Figure 2. Schematic diagram of the Query-level Objectness Gating.

Given the decoder query embedding $q_\ell \in \mathbb{R}^{N\times C}$ at the ℓth decoder layer, where N denotes the number of queries and C denotes the channel dimension, the objectness-aware gating coefficient is computed as:

$$\alpha_\ell = \sigma(W_g q_\ell + b_g) \tag{1}$$

where W_g and b_g are the learnable parameters of the linear projection, and σ denotes the sigmoid activation function. The gating coefficient $\alpha_\ell \in \mathbb{R}^{N\times 1}$ represents the estimated objectness score of each query and is broadcast along the channel dimension during multiplication. The gated query embedding is then obtained as:

$$\widetilde{q}_\ell = \alpha_\ell \odot q_\ell \tag{2}$$

where $\odot$ denotes element-wise multiplication with broadcasting. The rescaled query $\widetilde{q}_\ell$ is then fed into the subsequent decoder computation. When the estimated objectness score of a query is low, the corresponding gating coefficient becomes small, reducing the influence of this background-dominated query. Conversely, queries with high objectness scores are largely preserved. Since the gating coefficient is predicted directly from the decoder query itself, QOG does not require additional objectness annotations or extra supervision. In this way, QOG enhances the decoder's focus on transmission tower structures and improves robustness under cross-domain background variations.

2.4. MPDIoU Regression Loss

Bounding-box regression (BBR) is fundamental to target detection, and the design of its loss function significantly influences the model's localisation accuracy. Since the introduction of the IoU loss [18], several enhanced variants have emerged in the academic literature. The GIoU loss [16] addresses gradient vanishing in non-overlapping regions by incorporating the minimum outer rectangle. The DIoU loss [19] further integrates centroid distance into the penalty term, while the CIoU loss [19] introduces aspect-ratio consistency constraints in addition to DIoU and has become the default configuration for mainstream detectors such as YOLOv7. However, these loss functions encounter substantial challenges in cross-domain power tower detection scenarios.

First, a power tower is a typical elongated target, and the Intersection over Union (IoU) metric is highly sensitive to variations in aspect ratio. When the centroids of the

predicted frame and the actual frame are perfectly aligned and the width and height are scaled proportionally, the IoU, Generalised IoU (GIoU), Distance IoU (DIoU) and Complete IoU (CIoU) metrics cease to be optimised, causing the model to converge to a local optimum. Second, existing IoU variants do not explicitly account for vertex alignment between the predicted and actual frames. The distance between vertices often reflects localisation quality more accurately than the overlap area, particularly when geometric distortions arise from changes in viewing angle and variations in tower morphology within cross-domain contexts. For instance, in cross-domain transmission tower detection, a predicted frame that is offset by 10 pixels vertically but maintains the correct width and height may yield a higher IoU than a predicted frame with perfectly aligned vertices but slightly altered width and height. However, the latter clearly represents a superior localisation outcome in engineering measurements.

To address this problem, Ma and Xu [12] proposed a new bounding-box similarity metric, MPDIoU (Minimum Point Distance Intersection over Union), in 2023. The core concept of this method is that the geometric similarity of bounding boxes can be fully represented by the normalised distances between two key vertices in the image coordinate system, namely the upper-left and lower-right corners. Unlike CIoU and EIoU, which indirectly model bounding-box similarity by combining multiple geometric factors, MPDIoU directly uses vertex coordinates as the optimisation objective. It also incorporates three key factors—overlap region, centroid distance and width–height deviation—in a concise mathematical form, with significantly lower computational complexity than CIoU.

The geometric relationship between the predicted box, the ground-truth box and the vertex-distance penalty is illustrated in Figure 3. MPDIoU is defined as follows: given a coordinate frame $B = (x_1, y_1, x_2, y_2)$ and a real coordinate frame $B^{gt} = \left(x_1^{gt}, y_1^{gt}, x_2^{gt}, y_2^{gt}\right)$, where (x_1, y_1) is the upper-left coordinate and (x_2, y_2) is the lower-right coordinate, the standard Intersection over Union ratio of the two boxes is first calculated:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \tag{3}$$

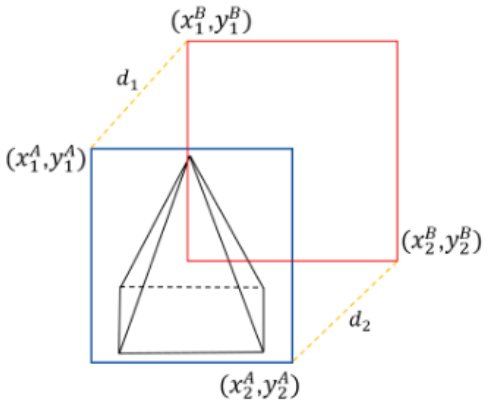


Figure 3. Schematic diagram of the MPDIoU loss function.

MPDIoU introduces the Euclidean distance between vertices as a penalty term based on IoU and normalises the penalty by the image size:

$$MPDIoU = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \tag{4}$$

$$d_1 = \sqrt{\left(x_1 - x_1^{gt}\right)^2 + \left(y_1 - y_1^{gt}\right)^2} \; d_2 = \sqrt{\left(x_2 - x_2^{gt}\right)^2 + \left(y_2 - y_2^{gt}\right)^2}$$

where d_1 and d_2 denote the Euclidean distances between the corresponding upper-left and lower-right vertices of the predicted and ground-truth boxes, respectively; w is the width and h is the height of the input image, and the corresponding loss function of MPDIoU is defined as:

$$\mathcal{L}_{MPDIoU} = 1 - MPDIoU \tag{5}$$

$$\mathcal{L}_{reg} = \lambda_{L1}\mathcal{L}_{L1} + \lambda_{MPDIoU}\mathcal{L}_{MPDIoU} \lambda_{MPDIoU} = 2.0, \lambda_{L1} = 5.0$$

In this paper, the original regression loss is replaced with MPDIoU in the DINO detection head, while the classification loss and other training configurations are retained to ensure a consistent loss-weight balance with the baseline.

This loss function has three important properties. First, completeness: MPDIoU equals 1 and the loss is zero only when the coordinates of the four vertices of the predicted frame coincide precisely with those of the actual frame; any deviation at any vertex incurs a penalty. Second, scale invariance ensures that the loss value remains unaffected by the absolute size of the input image. Third, computational efficiency: MPDIoU does not require calculation of the inverse-tangent aspect ratio, avoids complex weighting of geometric factors and uses approximately 60% of the floating-point operations required by CIoU.

2.5. Quality-Aware Scoring Module (QSM)

The Quality-aware Scoring Module (QSM) introduces a lightweight quality head to estimate the localisation reliability of candidate boxes. In the implementation, each decoder layer is equipped with an independent linear quality head that takes the corresponding decoder-query feature as input and outputs one quality logit for each query. During training, the quality head is supervised on positive denoising queries. The target is the IoU between the detached predicted box and the matched ground-truth box, which is clipped to the range [0, 1], and the optimisation objective is binary cross-entropy with logits. The loss weight of this quality branch is set to 1.0.

During inference, the quality prediction is used to calibrate the classification confidence before candidate ranking. For the i-th query in the l-th decoder layer, the original classification output and the predicted quality output are fused to obtain the calibrated confidence. The fusion is defined as follows:

$$\hat{s}_{\ell,i} = s_{\ell,i}^{cls} + \alpha \cdot q_{\ell,i} \tag{6}$$

where the hatted term denotes the calibrated confidence, the superscript '*cls*' denotes the original classification output, q denotes the quality logit predicted by the quality head, and α is the fusion coefficient. In the implementation, α corresponds to quality_fuse_pow and is set to 1.0. The operation is performed on logits before the final activation and top-k selection, which is consistent with the code-level implementation in which the quality logit is added to the classification logits. Therefore, Equation (6) should be interpreted as logit-level score calibration rather than direct addition of two post-sigmoid probabilities.

This calibration gives lower priority to boxes with high classification responses but unreliable localisation quality [20], while preserving boxes whose classification and localisation predictions are mutually consistent. For slender transmission towers, this design is useful because small box shifts can noticeably reduce localisation quality and may otherwise lead to high-confidence false detections.

2.6. Dataset and Experimental Settings

2.6.1. Dataset

Source-domain dataset: The TT dataset refers to the publicly available power transmission and distribution infrastructure imagery dataset released through the Duke Energy Initiative Data+ project [21]. This public dataset contains high-resolution remote sensing imagery collected across multiple countries and cities and provides annotations for electric power infrastructure objects, including transmission towers, distribution towers, transmission lines, distribution lines and transformer substations. The original imagery in the benchmark was resampled to a ground sampling distance of approximately 0.3 m, and the source-domain samples used in this study were cropped and standardised into 512 × 512 pixel image slices. In this study, the transmission tower samples from the TT dataset were used as the source-domain data. Table 1 summarises the cross-country distribution of transmission tower objects in the public TT dataset.

Table 1. Statistical characteristics of power towers in different countries.

Country	Involved Cities	Total Number of Objects
Brazil	Rio de Janeiro	55
China	Shanghai	307
New Zealand	Dunedin, Gisborne, Palmerston North, Rotorua, Tauranga	195
Sudan	Khartoum	84
USA	Tucson, Hartford, Colwich–Maize, Clyde, Wilmington	137

Target-domain dataset: The DNP dataset is a self-annotated target-domain dataset constructed from high-resolution Google Earth imagery covering the Jiaxing–Huzhou region in eastern China. The original scenes, with a spatial resolution of approximately 0.3 m, were cropped into 512 × 512 pixel image slices for model training and testing. The dataset was designed to represent local cross-region deployment scenes, including rural areas, urban fringes, woodland environments, greenhouses and water-adjacent backgrounds. The DNP dataset contains typical transmission tower types observed in eastern China, including cat-head towers, wine-glass towers, dry-shaped towers and compact or angle towers. It contains 5123 valid target-domain test images and 6104 annotated transmission tower instances. Compared with the public TT source-domain dataset, the DNP dataset exhibits different tower structures, scale distributions, background textures, land-cover contexts and imaging conditions. These differences increase the difficulty of cross-domain detection and make the DNP dataset suitable for evaluating cross-domain generalisation.

The self-annotated DNP dataset was labelled using LabelImg version 1.8.6. Following common object-detection annotation practice, each transmission tower was annotated with a tight horizontal bounding box enclosing the visible tower structure. The annotation rules were as follows: (1) each independently visible transmission tower was annotated as one object instance; (2) the bounding box was drawn to cover the visible tower body as completely as possible while excluding irrelevant background; (3) partially occluded towers were annotated when the main tower structure remained identifiable; (4) extremely ambiguous objects that could not be confidently recognised as transmission towers were excluded; and (5) all annotations were manually checked to correct missing boxes, duplicate boxes and boxes with poor boundary alignment.

Table 2 summarises the three cross-domain training configurations used in the controlled experiments. The source/target-domain definition and annotation-ratio design are described in Section 2.6.2.

Table 2. Cross-domain dataset construction.

Split	Subset Type	Image Count	Annotation Count
TT + 0% DNP	Training	735	1004
	Validation	101	167
	Test	5123	6104
TT + 5% DNP	Training	940	1268
	Validation	130	172
	Test	5123	6104
TT + 10% DNP	Training	1145	1510
	Validation	158	210
	Test	5123	6104

2.6.2. Source/Target Domain Split and Annotation Ratios

To assess cross-domain generalisation, this study constructed three dataset partitions by varying the proportion of annotated target-domain DNP samples introduced during training. The TT dataset, namely the public power transmission and distribution infrastructure imagery dataset released through the Duke Energy Initiative Data+ project, is defined as the source domain. The DNP dataset, namely the self-annotated Jiaxing–Huzhou target-domain dataset, is defined as the target domain. In the 0% target-domain annotation setting, the training data contain only source-domain TT samples, and the model is evaluated on the fixed DNP target-domain test set. In the 5% and 10% settings, limited annotated DNP samples are added to the training data, while the same DNP test set is retained for evaluation. This design enables controlled assessment of how limited target-domain annotations affect cross-domain detection performance. The detailed image and annotation counts for each split are shown in Table 2.

As shown in Table 2, the training and validation subsets vary across the 0%, 5% and 10% target-domain annotation settings. The 0% setting uses 735 training images and 101 validation images without annotated target-domain training samples. The 5% setting uses 940 training images and 130 validation images, while the 10% setting uses 1145 training images and 158 validation images. The DNP test set is fixed across all settings and contains 5123 images and 6104 annotated transmission towers.

To further characterise the geometric distribution of transmission tower instances, Table 3 reports the bounding-box statistics of the training and test sets. Here, the training set denotes the source-plus-target training samples used for model optimisation, whereas the test set denotes the fixed DNP target-domain test set. Compared with the training set, the fixed target-domain test set has larger average width, height and area, indicating a noticeable difference in tower scale and bounding-box geometry between the training data and the target-domain test data. This difference further reflects the cross-domain distribution shift considered in this study.

Table 3. Statistical metrics of bounding boxes in the training and test sets.

Dataset	Average Width (pixel)	Average Height (pixel)	Average Area (pixel^2)	Area Variance (pixel^4)
train	84.82	78.43	8353.08	145,294,900.33
test	98.05	96.31	10,272.04	106,096,110.02

2.7. Implementation Details

All models were executed in a Python 3.8.20 and PyTorch 2.1.0 environment using GPUs. The AdamW optimiser was employed with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-4}, together with gradient clipping. The learning rate schedule combined linear warm-up with a multi-step decay strategy. The learning rate was first linearly increased from 1×10^{-7} to 1×10^{-4} over the first 500 iterations and was then reduced at the 16th and 22nd epochs with a decay rate of 0.1. The backbone network used a smaller learning-rate multiplier ($0.1\times$) for hierarchical learning-rate tuning.

To address the classification and localisation challenges in target detection, a multi-task loss function was employed, comprising Focal Loss, L1 Loss and GIoU Loss. Specifically, Focal Loss was applied to the classification task with a focus parameter of $\gamma = 2$ and a balance factor of $\alpha = 0.25$, which increases the emphasis on hard-to-classify samples by reducing the weight of easily classified samples. L1 Loss and GIoU Loss were used for bounding-box regression and Intersection over Union optimisation, respectively, with weights of 5.0 and 2.0. All models were trained for 36 epochs.

2.8. Evaluation Metrics

We quantitatively evaluated model performance using standard object-detection metrics, including precision, recall, F1-score, average precision at an IoU threshold of 0.5 (AP@0.5) and the reported mAP@0.5. Because transmission tower detection is a single-class detection task in this study, the reported mAP@0.5 is numerically equivalent to AP@0.5 but is retained for consistency with common object-detection reporting conventions. The main comparison tables report mAP@0.5 for all evaluated detectors. In the statistical robustness analysis, mAP@0.5:0.95 is additionally reported for the DINO baseline and the complete proposed model to examine localisation stability across stricter IoU thresholds. The metrics are defined below.

Precision measures the proportion of samples predicted as positive that are actually positive:

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

where TP and FP denote true positive cases and false positive cases, respectively.

Recall measures the proportion of actual positive samples that are correctly predicted:

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

where FN denotes false negative cases.

F1 is the harmonic mean of precision and recall and summarises model performance:

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{9}$$

The reported mAP@0.5 is calculated as the area under the precision-recall curve at the IoU threshold of 0.5:

$$mAP = \int_0^1 p(r)\, dr \approx \sum_{k=1}^{n} p(k)\Delta r(k) \tag{10}$$

where $p(r)$ denotes the precision-recall curve, and $p(k)$ and $\Delta r(k)$ denote the interpolated precision sampling point and the corresponding recall increment, respectively. For the robustness analysis, mAP@0.5:0.95 denotes the mean average precision averaged over IoU thresholds from 0.50 to 0.95 with a step size of 0.05.

3. Results

3.1. Cross-Domain Performance Under Different Target-Domain Annotation Ratios

Tables 4–6 compare YOLOv8s, YOLOv9s, YOLOv11n, the original DINO baseline and the proposed method under the 0%, 5% and 10% target-domain annotation settings. In addition to detection accuracy, the tables report the number of parameters, FLOPs, FPS and GPU memory consumption to evaluate the computational cost of each detector. The inclusion of the original DINO baseline enables direct assessment of the improvements introduced into the DINO framework.

Table 4. Detection accuracy and parameters under the 0% target-domain annotation setting for different methods.

Model	Precision	Recall	F1-Score	mAP@0.5	Params (M)	FLOPs (G)	FPS	Memory (MB)
YOLOv11n	0.2902	0.1207	0.1767	0.0952	2.59	4.122	64.686	187.4
YOLOv8s	0.3161	0.1892	0.2367	0.1445	11.136	18.334	84.4	277.9
YOLOv9s	0.1850	0.1381	0.1681	0.0690	7.288	17.527	39.942	232.4
DINO	0.3227	0.1641	0.2176	0.1148	47.54	80.705	13.23	531.8
Ours	0.2956	0.1972	0.2366	0.1480	47.542	80.706	15.267	531.8

Table 5. Detection accuracy and parameters under the 5% target-domain annotation setting for different methods.

Model	Precision	Recall	F1-Score	mAP@0.5	Params (M)	FLOPs (G)	FPS	Memory (MB)
YOLOv11n	0.7836	0.6245	0.6951	0.7173	2.59	4.122	69.555	190
YOLOv8s	0.8180	0.7107	0.7606	0.7893	11.136	18.334	82.191	363.7
YOLOv9s	0.7529	0.6445	0.6945	0.7187	7.288	17.527	39.928	233
DINO	0.8430	0.7448	0.7909	0.7795	47.54	80.705	14.831	531.8
Ours	0.8498	0.7652	0.8053	0.8077	47.542	80.706	15.842	531.8

Table 6. Detection accuracy and parameters under the 10% target-domain annotation setting for different methods.

Model	Precision	Recall	F1-Score	mAP@0.5	Params (M)	FLOPs (G)	FPS	Memory (MB)
YOLOv11n	0.8257	0.7128	0.7651	0.8002	2.59	4.122	67.787	190
YOLOv8s	0.8434	0.7363	0.7862	0.8253	11.136	18.334	81.613	363.7
YOLOv9s	0.7709	0.6445	0.7021	0.7245	7.288	17.527	42.030	233
DINO	0.8748	0.8126	0.8425	0.8531	47.54	80.705	14.732	531.8
Ours	0.8947	0.8199	0.8556	0.8684	47.542	80.706	15.932	531.8

Under the 0% target-domain annotation setting, all methods are trained without annotated target-domain samples and evaluated on the fixed DNP target-domain test set. The proposed method achieves the highest recall (0.1972) and mAP@0.5 (0.1480), outperforming the original DINO baseline in recall, F1-score and mAP@0.5. Specifically, compared with DINO, the proposed method improves recall from 0.1641 to 0.1972, F1-score from 0.2176 to 0.2366 and mAP@0.5 from 0.1148 to 0.1480. This indicates that the proposed modules improve target-domain generalisation even when no target-domain annotations are available during training.

When 5% annotated target-domain samples are introduced, the proposed method further improves upon the original DINO baseline. Compared with DINO, precision increases from 0.8430 to 0.8498, recall from 0.7448 to 0.7652, F1-score from 0.7909 to 0.8053 and mAP@0.5 from 0.7795 to 0.8077. These results show that limited target-domain supervision

can substantially reduce the domain gap and that the proposed method benefits more effectively from the additional target-domain annotations.

Table 6 presents the 10% target-domain annotation setting. Under this setting, the proposed method achieves the best overall performance, with a precision of 0.8947, a recall of 0.8199, an F1-score of 0.8556 and an mAP@0.5 of 0.8684. Compared with the original DINO baseline, the proposed method improves precision by 0.0199, recall by 0.0073, F1-score by 0.0131 and mAP@0.5 by 0.0153. Compared with the YOLO-series detectors, the proposed method provides a more balanced improvement in recall and localisation-related metrics, which is important for transmission tower detection under geographic domain shift.

The complexity results show that the proposed model introduces only a negligible computational increase compared with the original DINO baseline. Across the three annotation settings, the number of parameters increases from 47.540 M to 47.542 M, while FLOPs increase only from 80.705 G to 80.706 G. GPU memory remains unchanged at 531.8 MB. The measured FPS of the proposed model is also comparable to that of the DINO baseline and is slightly higher in the current evaluation environment. These results support the claim that the proposed modules improve cross-domain detection accuracy without introducing a meaningful additional inference burden.

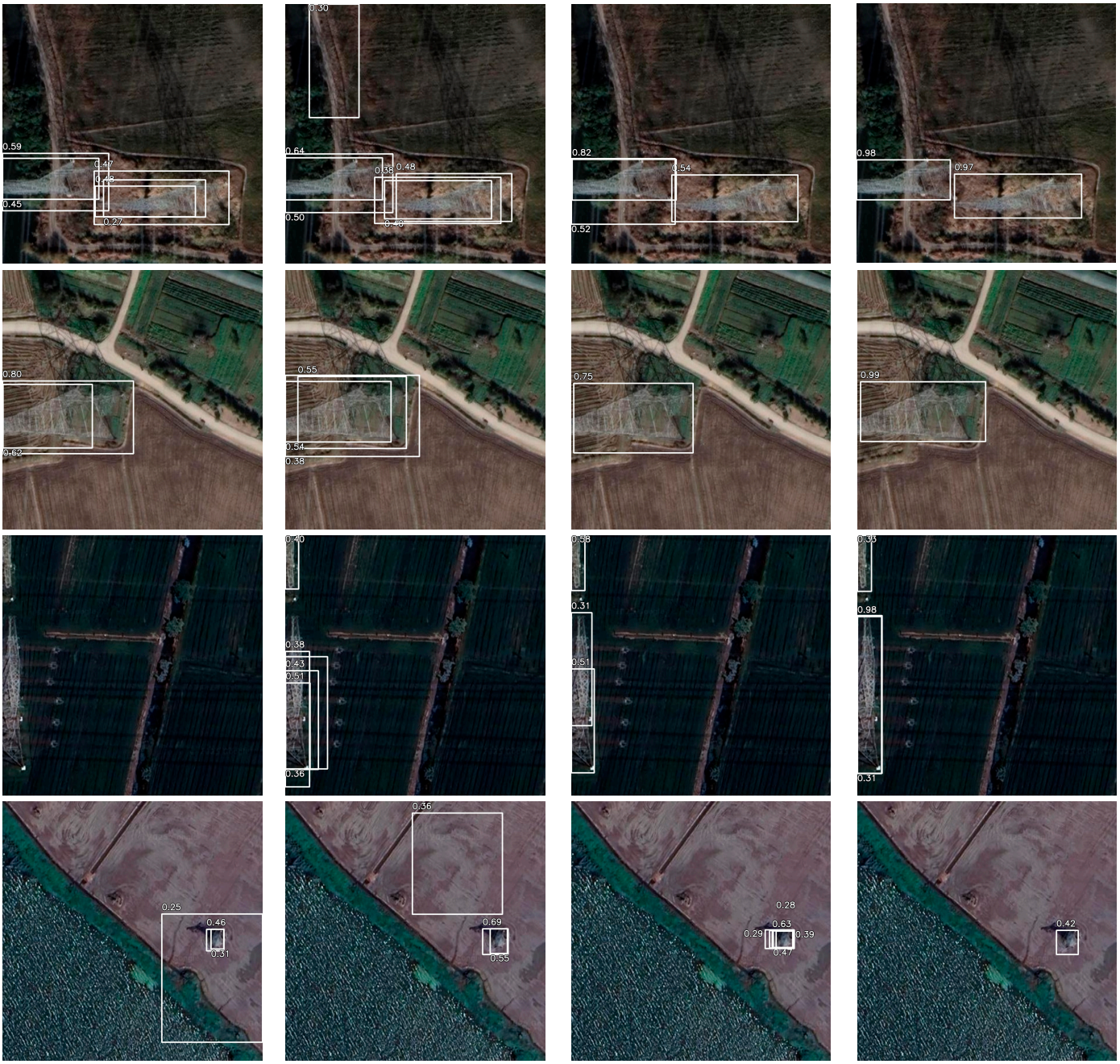
To further evaluate the statistical robustness of the improvement over the DINO baseline, three independent runs with different random seeds were conducted under the 10% target-domain annotation setting for the original DINO baseline and the proposed model. The mean and standard deviation on the fixed DNP target-domain test set are shown in Table 7.

Table 7. Statistical robustness comparison over three independent runs under the 10% target-domain annotation setting.

Model	Test Precision	Test Recall	Test mAP@0.5	Test mAP@0.5:0.95
DINO	0.8748 ± 0.0057	0.8123 ± 0.0068	0.8530 ± 0.0054	0.5007 ± 0.0055
Ours	0.8904 ± 0.0039	0.8168 ± 0.0053	0.8692 ± 0.0081	0.5334 ± 0.0088

The multi-run results show that the proposed model consistently outperforms the original DINO baseline. The mean test mAP@0.5 increases from 0.8530 ± 0.0054 to 0.8692 ± 0.0081, corresponding to an improvement of approximately 1.9%. Moreover, the range of test mAP@0.5 obtained by the proposed model (0.8616–0.8777) does not overlap with that of DINO (0.8476–0.8584), indicating that the improvement is larger than the observed run-to-run variation. The improvement is more pronounced for mAP@0.5:0.95, which increases from 0.5007 ± 0.0055 to 0.5334 ± 0.0088. This suggests that the proposed QOG, QSM and MPDIoU components not only improve single-threshold detection accuracy but also enhance localisation stability across stricter IoU thresholds.

Figure 4 presents the qualitative evaluation results. Columns (a)–(d) show the results obtained by YOLOv11n, YOLOv8s, YOLOv9s and Ours, respectively. Rows 1–5 correspond to five representative scenarios: small-object detection near farmland roads, dense-object detection in regular farmland, multi-object detection in complex farmland backgrounds, low-contrast detection near rivers and object detection in complex urban construction backgrounds. The comparison illustrates representative detection behaviour under cross-domain remote sensing scenes, including duplicate, incomplete and missed detection.



(**a**) YOLOv11n (**b**) YOLOv8s (**c**) YOLOv9s (**d**) Ours

Figure 4. Detection results derived from the proposed model and mainstream models.

The performance trend across the 0%, 5% and 10% settings shows that target-domain annotations progressively improve cross-domain detection accuracy. However, the proposed method consistently remains competitive with or superior to the original DINO baseline, indicating that QOG, QSM and MPDIoU provide complementary benefits beyond simple target-domain fine-tuning. The improvement is particularly reflected in recall and mAP@0.5, suggesting that the enhanced model is more effective at recovering missed towers and improving bounding-box quality under domain shift.

3.2. Ablation Study

To assess the separate and combined effects of the proposed modules, ablation experiments were conducted using DINO as the baseline, as shown in Table 8. The symbol $\surd$ indicates that the corresponding module is included in the model configuration. Compared

with the baseline, adding QOG alone increases precision from 0.8748 to 0.8837, recall from 0.8126 to 0.8144 and F1-score from 0.8425 to 0.8476, although mAP@0.5 slightly decreases from 0.8531 to 0.8473. This suggests that QOG improves query filtering and reduces false responses, but when used alone it may also suppress some uncertain positive queries, which can affect localisation-sensitive average precision.

Table 8. Ablation study of QOG, QSM and MPDIoU.

Baseline	QOG	QSM	MPDIoU	Precision	Recall	F1-Score	mAP@0.5
√				0.8748	0.8126	0.8425	0.8531
√	√			0.8837	0.8144	0.8476	0.8473
√			√	0.8820	0.8201	0.8499	0.8602
√		√		0.8756	0.8132	0.8432	0.8533
√	√	√		0.8891	0.8153	0.8506	0.8582
√	√	√	√	0.8947	0.8199	0.8556	0.8684

The MPDIoU-only setting improves the baseline across all metrics, increasing precision to 0.8820, recall to 0.8201, F1-score to 0.8499 and mAP@0.5 to 0.8602. This confirms that MPDIoU directly strengthens bounding-box regression for elongated transmission tower targets. The QSM-only setting produces a smaller but consistent improvement relative to the baseline, with precision, recall, F1-score and mAP@0.5 increasing to 0.8756, 0.8132, 0.8432 and 0.8533, respectively. This indicates that localisation-aware confidence calibration contributes to reducing confidence-quality mismatch, although its effect is more limited when used without QOG and MPDIoU.

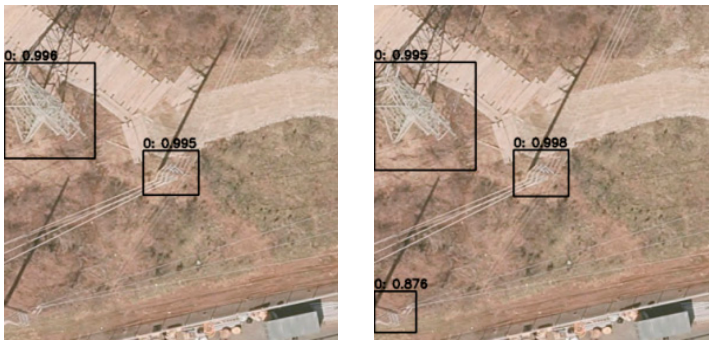
When QOG and QSM are combined, the model reaches a precision of 0.8891, a recall of 0.8153, an F1-score of 0.8506 and an mAP@0.5 of 0.8582. The complete model integrating QOG, QSM and MPDIoU achieves the best overall performance, with a precision of 0.8947, a recall of 0.8199, an F1-score of 0.8556 and an mAP@0.5 of 0.8684. These results demonstrate that the three modules are complementary: QOG improves query selection, QSM calibrates confidence using localisation quality and MPDIoU improves geometric alignment for slender targets.

In the qualitative ablation evaluation, Figure 5 compares different module combinations for transmission tower detection. The baseline DINO model tends to miss small transmission towers and sometimes produces inaccurate detections in darker or more complex backgrounds. With the gradual incorporation of QOG, QSM and MPDIoU, the model shows improved detection confidence and more accurate localisation. The final model is more stable in complex backgrounds and identifies previously missed small towers more effectively.



Base QOG

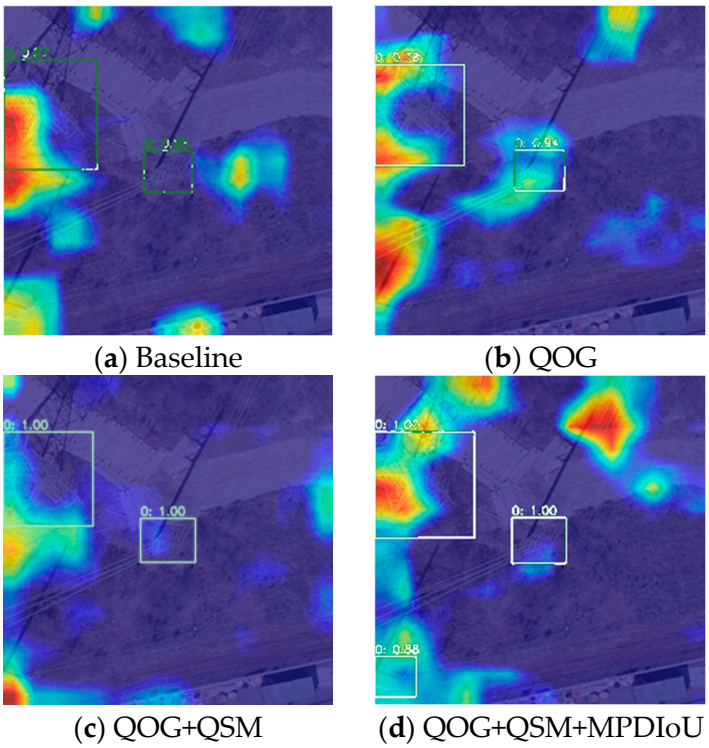
Figure 5. *Cont.*



QOG+QSM QOG+QSM+MPDIoU

Figure 5. Qualitative comparison results of ablation experiments.

Figure 6 illustrates the differences in feature focusing across the ablation models using attention heat maps. The baseline DINO model exhibits dispersed high-attention regions and is susceptible to interference from background textures. After introducing QOG, the attention response becomes more concentrated on the tower body. With the addition of QSM and MPDIoU, the high-attention regions cover the tower structure more completely and background interference is further reduced. This visual trend is consistent with the quantitative improvements reported in Table 8.



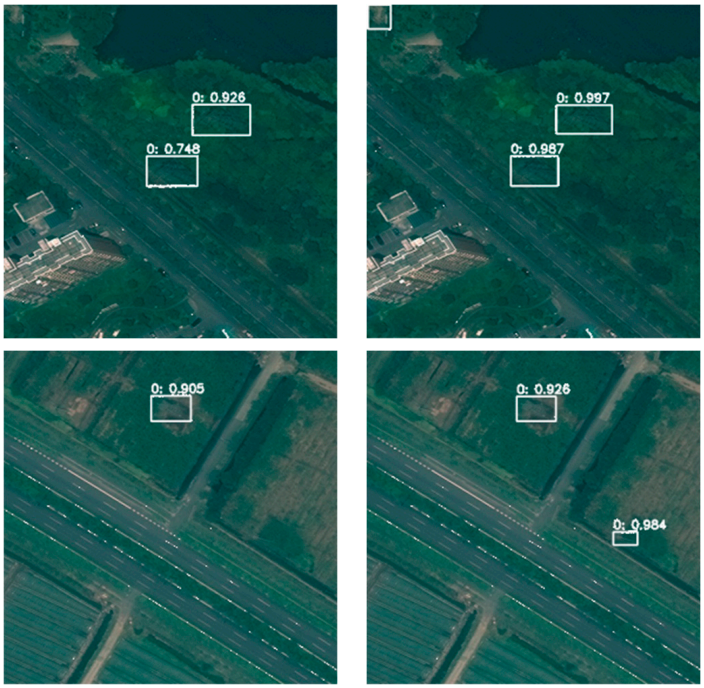
(**a**) Baseline (**b**) QOG

(**c**) QOG+QSM (**d**) QOG+QSM+MPDIoU

Figure 6. Comparative analysis of attention thermograms for ablation experiments.

In the cross-domain generalisation scenario, the full model benefits from the complementary effects of QOG, QSM and MPDIoU. Compared with the baseline DINO model, the enhanced model shifts attention from scattered background regions to the transmission tower structure. The baseline DINO model still exhibits attention responses that are vulnerable to terrain texture interference, whereas the final model achieves more complete coverage of tower structural features and more precise alignment between predicted boxes and target outlines.

The visual results also reveal remaining limitations. In images with dark or cluttered backgrounds, the proposed model may still misidentify ambiguous small structures as transmission towers. For example, in Figure 7, the small object in the lower-left region is difficult to distinguish visually and is detected as a transmission tower by the model, which may result in a small number of false positives.



(**a**) DINO (baseline) (**b**) Ours (proposed)

Figure 7. Comparison of detection results for a small target in a complex background.

Overall, the quantitative and qualitative results indicate that the proposed modules improve detection confidence, localisation accuracy and robustness on the cross-domain test set. The improvements are most evident when limited target-domain annotations are available, but the 0% setting also shows that the enhanced DINO framework provides better generalisation than the original DINO baseline under strict source-to-target transfer.

4. Discussion

The results indicate that geographic domain shift in transmission tower detection is mainly reflected in differences in tower morphology, object scale, background texture and imaging conditions. The results under the 0%, 5% and 10% target-domain annotation ratios show that increasing target-domain supervision improves performance, but the proposed model consistently provides stronger cross-domain detection capability than the original DINO baseline. This suggests that the proposed modules do not merely benefit from additional target-domain samples, but also improve the detector's ability to handle domain-induced query interference and localisation bias.

The ablation results further clarify the role of each module. QOG mainly improves decoder-query filtering and suppresses background-biassed responses, but its standalone use may slightly reduce mAP@0.5 because some uncertain positive queries can also be weakened. MPDIoU provides the most direct improvement in localisation quality by strengthening vertex-level geometric constraints for slender tower targets. QSM contributes by calibrating classification confidence with localisation quality, which is useful for reducing high-confidence detections with poor box alignment. Together, the three modules yield the strongest cross-domain performance, which is consistent with the individual ablation trends.

The comparison with YOLO-series detectors should be interpreted together with the differences in model architecture. YOLO models rely on convolutional feature extraction and predefined detection heads, whereas DINO uses transformer-based query interaction. By adding the original DINO baseline and reporting parameters, FLOPs, FPS and memory consumption, the revised experiments directly show that the proposed modules improve the underlying DINO detector while introducing almost no additional computational cost. This is consistent with the module design: QOG and the quality-aware scoring branch are lightweight query-level operations, and MPDIoU mainly changes the regression objective rather than adding a heavy inference subnetwork.

Although the enhanced model improves cross-domain detection performance, several limitations remain. Small transmission towers in dark or cluttered backgrounds may still be confused with visually similar structures, and the regional distribution of tower samples is imbalanced between datasets. The additional three-run experiment under the 10% target-domain annotation setting confirms that the improvement of the complete model over the DINO baseline is stable in terms of test mAP@0.5 and mAP@0.5:0.95. Nevertheless, this statistical robustness analysis is currently limited to the DINO baseline and the complete proposed model, while the YOLO-series comparisons and the 0% and 5% annotation settings are still reported as single-run results. Future work will extend the multi-seed evaluation to more baselines, annotation ratios and cross-region datasets.

5. Conclusions

This paper proposes an enhanced framework for the cross-domain detection of transmission towers to address the degradation in detection accuracy and localisation bias exhibited by target detectors under differences in tower morphology, background texture and object scale. Using DINO as the baseline model, the framework improves cross-domain detection through three lightweight enhancements: Query-level Objectness Gating, MPDIoU regression and query-level quality-aware scoring. Experiments under 0%, 5% and 10% target-domain annotation settings demonstrate that the proposed method consistently improves over the original DINO baseline and achieves the best overall performance under the 10% setting, with a precision of 0.8947, a recall of 0.8199, an F1-score of 0.8556 and an mAP@0.5 of 0.8684. Complexity analysis further shows that the accuracy gain is obtained with almost unchanged parameters, FLOPs and GPU memory relative to the DINO baseline, indicating that the proposed framework is suitable for robust cross-domain transmission tower detection in remote sensing images.

Author Contributions: Conceptualization, J.W., L.Z. and J.Y.; methodology, J.W., J.T. and N.L.; software, J.W.; validation, J.T. and L.S.; formal analysis, J.W.; investigation, J.T.; resources, N.L. and J.Y.; data curation, J.W. and J.T.; writing—original draft preparation, J.W.; writing—review and editing, L.Z., N.L. and L.S.; visualization, J.W.; supervision, L.Z. and J.Y.; project administration, L.Z; funding acquisition, L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Key Research and Development Program of China under Grant 2024YFC3810804, National Natural Science Foundation of China under Grant 62471337 and Grant 42201416, in part by National Natural Science Foundation of Wuhan under Grant 2026040301020027, and in part by Open Fund of Key Laboratory of National Geographical Census and Monitoring of Ministry of Natural Resources 2026NGCM2-05.

Data Availability Statement: The publicly available TT dataset can be accessed at Figshare (https://doi.org/10.6084/m9.figshare.6931088.v1) [21]. The DNP dataset generated in this study is not publicly available due to privacy restrictions and data management policies, but is available from the corresponding author upon reasonable request.

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

References

1. Brunelli, R. *Template Matching Techniques in Computer Vision: Theory and Practice*; Wiley: Hoboken, NJ, USA, 2009.
2. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2005; pp. 886–893.
3. Ma, W.; Xiao, J.; Zhu, G.; Wang, J.; Zhang, D.; Fang, X.; Miao, Q. Transmission tower and power line detection based on improved Solov2. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 1–11. [CrossRef]
4. Gwizdała, J.; Öner, D.; Roy, S.K.; Shah, M.A.; Eberhard, A.; Egorov, I.; Krüsi, P.; Yakushev, G.; Fua, P. Vision-based power line cables and pylons detection for low flying aircraft. *Mach. Vis. Appl.* **2025**, *36*, 46. [CrossRef]
5. Son, H.; Kusari, A. Quantifying context bias in domain adaptation for object detection. *arXiv* **2024**, arXiv:2409.14679.
6. Zhou, T.; Wang, Y. A survey on domain adaptive object detection. *Comput. J.* **2026**, *69*, 357–375. [CrossRef]
7. Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* **2016**, *17*, 1–35.
8. Guo, W.; He, B.; Wu, Y.; Li, X.; Wu, L. Cross-domain soft adaptive teacher for Syn2Real object detection. In *Pattern Recognition and Computer Vision: 6th Chinese Conference, PRCV 2023, Proceedings, Part VI*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 460–472.
9. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2223–2232.
10. Barišić Kulas, A.; Petric, F.; Bogdan, S. Aerial maritime vessel detection and identification. In *Proceedings of the 2025 International Conference on Unmanned Aircraft Systems (ICUAS)*; IEEE: New York, NY, USA, 2025; pp. 44–51. [CrossRef]
11. Danish, M.S.; Iqbal, J.; Ali, M.; Sarfraz, M.S.; Khan, S.; Khan, M.H. Perturbing dominant feature modes for single domain-generalized object detection. In *Proceedings of the 25th International Conference on Digital Image Computing: Techniques and Applications (DICTA)*; IEEE: New York, NY, USA, 2024; pp. 93–100. [CrossRef]
12. Ma, S.; Xu, Y. MPDIoU: A loss for efficient and accurate bounding box regression. *arXiv* **2023**, arXiv:2307.07662. [CrossRef]
13. Zhang, H.; Li, F.; Liu, S.; Zhang, L.; Su, H.; Zhu, J.; Ni, L.M.; Shum, H.Y. DINO: DETR with improved denoising anchor boxes for end-to-end object detection. In Proceedings of the International Conference on Learning Representations (ICLR), Kigali, Rwanda, 1–5 May 2023.
14. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable DETR: Deformable transformers for end-to-end object detection. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 3–7 May 2021.
15. Li, F.; Zhang, H.; Liu, S.; Guo, J.; Ni, L.M.; Zhang, L. DN-DETR: Accelerate DETR training by introducing query denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 21–24 June 2022; pp. 13619–13627.
16. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
17. Li, T.; Wang, C.; Tian, S.; Zhang, B.; Wu, F.; Tang, Y.; Zhang, H. TACMT: Text-aware cross-modal transformer for visual grounding on high-resolution SAR images. *ISPRS J. Photogramm. Remote Sens.* **2025**, *222*, 152–166. [CrossRef]
18. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T.S. UnitBox: An advanced object detection network. In Proceedings of the ACM International Conference on Multimedia (MM), Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
19. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. *Proc. AAAI Conf. Artif. Intell.* **2020**, *34*, 12993–13000. [CrossRef]
20. Lin, C.; Shi, Y.; Xie, C.; Li, M. Quantification and mitigation of border-level localization deviation for object detectors. *Expert Syst. Appl.* **2025**, *278*, 127435. [CrossRef]
21. Bradbury, K.; Han, Q.; Nair, V.; Pathirathna, T.; You, X. *Electric Transmission and Distribution Infrastructure Imagery Dataset*; Figshare: London, UK, 2018. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.