



# Article Optimizing Insulator Defect Detection with Improved DETR Models

Dong Li<sup>1</sup>, Panfei Yang<sup>1</sup> and Yuntao Zou<sup>2,\*</sup>

- <sup>1</sup> China Electronic Product Reliability and Environmental Testing Research Institute, Guangzhou 511300, China; lidong@ceprei.com (D.L.); yangpanfei@ceprei.com (P.Y.)
- <sup>2</sup> School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
- \* Correspondence: zouyuntao@hust.edu.cn

Abstract: With the increasing demand for electricity, the power grid is undergoing significant advancements. Insulators, which serve as protective devices for transmission lines in outdoor high-altitude power systems, are widely employed. However, the detection of defects in insulators captured under challenging conditions, such as rain, snow, fog, sunlight, and fast-moving drones during longdistance photography, remains a major challenge. To address this issue and improve the accuracy of defect detection, this paper presents a novel approach: the Multi-Scale Insulator Defect Detection Approach using Detection Transformer (DETR). In this study, we propose a multi-scale backbone network that effectively captures the features of small objects, enhancing the detection performance. Additionally, we introduce a self-attention upsampling (SAU) module to replace the conventional attention module, enhancing contextual information extraction and facilitating the detection of small objects. Furthermore, we introduce the insulator defect (IDIoU) loss, which mitigates the instability in the matching process caused by small defects. Extensive experiments were conducted on an insulator defect dataset to evaluate the performance of our proposed method. The results demonstrate that our approach achieves outstanding performance, particularly in detecting small defects. Compared to existing methods, our approach exhibits a remarkable 7.47% increase in the average precision, emphasizing its efficacy in insulator defect detection. The proposed method not only enhances the accuracy of defect detection, which is crucial for maintaining the reliability and safety of power transmission systems but also has broader implications for the maintenance and inspection of high-voltage power infrastructure.

Keywords: insulator defect detection; transformer; self-attention upsampling; small insulator defect

**MSC:** 68T10

### 1. Introduction

Insulators in high-voltage power transmission are components used to support and secure transmission lines. Their primary function is to support conductors on transmission lines while preventing the current from flowing through the supporting structures into the ground or other objects that should not be electrified. Insulators are typically made from materials such as ceramics, fiberglass-reinforced plastics, or rubber to ensure they can effectively prevent current leakage. Additionally, insulators can prevent the formation of electrical arcs caused by high voltages, which could damage transmission equipment and lead to power transmission interruptions. In high-voltage power transmission, insulators play a crucial role in ensuring the safe operation of power lines.

Damage to insulators can lead to a series of severe consequences. First, there is the interruption of power transmission; damaged insulators can cause a short circuit in power lines, leading to transmission interruptions. Second, there are safety risks; if an insulator is damaged, the current may flow down the pole to the ground, increasing the risk of



Citation: Li, D.; Yang, P.; Zou, Y. Optimizing Insulator Defect Detection with Improved DETR Models. *Mathematics* 2024, *12*, 1507. https:// doi.org/10.3390/math12101507

Academic Editor: Alessandro Niccolai

Received: 9 April 2024 Revised: 7 May 2024 Accepted: 8 May 2024 Published: 11 May 2024



**Copyright:** © 2024 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 (https:// creativecommons.org/licenses/by/ 4.0/). electrocution. Furthermore, damaged insulators can cause electrical arcs, posing a fire hazard. Third, there are economic losses; failures in transmission lines can disrupt the power supply to businesses and households, leading to corresponding economic losses. Finally, short circuits caused by damaged insulators can damage electrical equipment. Even if there is no immediate apparent equipment failure, it can lead to wear and tear on power system equipment over time, reducing equipment life and posing risks to the stability of the power system. Therefore, maintenance and inspection of insulators are important measures to ensure the safe and stable operation of high-voltage power transmission systems.

Insulators are key electrical components in overhead transmission systems, bearing the critical responsibility of preventing current leakage and maintaining line stability. Unfortunately, these devices may suffer from self-explosion, breakage, or flashover issues due to pollution, as they are constantly exposed to strong electric fields and harsh environmental conditions. Specifically, surface failures caused by contamination and their damage to small insulators have become a leading cause of power grid failures. It is estimated that such defects account for over half of all grid failures. Therefore, developing methods for the rapid and accurate detection of insulator conditions is essential for ensuring the stable operation of power grids.

With the development of drone and computer vision technology, optical remotesensing images, including drone imagery, have become increasingly important in both military and civilian applications [1]. The ability to locate and identify objects in these images is a fundamental challenge in image processing, with wide-ranging implications for disaster monitoring, urban management, and precision control [2–4]. Currently, traditional detection methods that rely on helicopter or manual field inspections, as well as the analysis of drone-captured images, are inefficient and costly in the face of China's complex power grid structure. This is particularly true for the identification of small defect targets in drone images, where the complex background increases the difficulty of recognition. Given this, the development of a technology suitable for detecting small target faults in complex backgrounds is urgently needed to enhance the efficiency and accuracy of detection efforts.

In recent years, DEtection TRansformer (DETR) [5] has made significant progress in the field of object detection, which has revolutionized object detection by treating it as a set prediction task and leveraging the powerful relational modeling capabilities of the transformer. The transition from traditional detection methods to advanced solutions such as DETR is not only expected to solve the problem of low efficiency but also to address the unique challenges posed by complex environments such as those in drone images. This study aims to achieve the accuracy and efficiency requirements of insulator defect detection tasks by enhancing the ability of the DETR model to detect small objects.

Although DETR variants offer numerous advantages, they also face several challenges when applied to the detection of insulator defects. Firstly, as deep networks progress, the scale of feature layers diminishes, resulting in the gradual loss of small defect features. This loss of information limits the model's ability to accurately detect minor defects. Secondly, the self-attention operations in DETR tend to overpower the limited information present in small defects, causing contamination from background features. This issue leads to decreased detection accuracy. Lastly, the Hungarian matching algorithm, which is commonly used in DETR, has inherent shortcomings. Minor shifts in the predicted bounding boxes can cause significant fluctuations in Intersection over Union (IoU) [6], making the optimization process difficult.

This study addresses these challenges and provides the following contributions:

- 1. To tackle the confusion between foreground and background features, we propose a context-based attention module. This module enables the model to effectively learn the relationship between defects and their backgrounds, improving detection accuracy.
- We proposed the insulator defect IoU (IDIoU) loss function, which specifically addresses the instability issues caused by small defects in the matching process. The IDIoU loss helps the model optimize the detection of minor defects and accelerates the training speed.

By incorporating these improvements into DETR variants, we aim to enhance the detection performance of insulator defects, particularly minor defects. The experimental results demonstrate the effectiveness of our proposed methods and highlight the potential for their application in real-world scenarios. The application of unmanned aerial vehicle remote sensing images has become the norm, and efficient and accurate image detection technology is needed in conjunction to form large-scale applications. This study has made progress in the accuracy and efficiency of small defect detection in complex backgrounds, laying a solid foundation for insulator defect detection in drone images. The application of this technology will greatly improve the efficiency of insulator defect detection and provide guarantees for the stable operation of the power grid.

#### 2. Related Work

### 2.1. CNN-Based Methods

Deep learning methods, particularly Convolutional Neural Networks (CNN), have shown great promise in object detection models for remote sensing images [7,8]. These models can be broadly classified into two categories: anchor-based approaches and anchorfree approaches.

Anchor-based approaches involve manually setting anchor sizes, which can lead to redundancy in computation and memory load [9]. While these methods have shown good performance, the reliance on manually defined anchors can be limiting in certain scenarios. On the other hand, anchor-free approaches directly provide object location and class information, eliminating the need for anchor adjustment [10]. By converting the traditional bounding box prediction problem into a corner prediction problem, anchor-free approaches avoid unnecessary hyperparameters and streamline the detection process.

The YOLO series models have been widely used in CNN-based object detection models in recent years. YOLOv5 [11] is renowned for its fast processing speeds, making it suitable for real-time detection tasks. However, its accuracy is not high in detecting very small or highly occluded objects. YOLO7 [12] and IYOLO7 [13] provide improvements in detection accuracy and speed, particularly in handling complex object detection scenarios with better handling of occlusions and small objects. However, the increased complexity leads to longer training times, and YOLO7 and IYOLO7 require more computational power for optimal performance.

### 2.2. Transformer-Based Object Detection Technology

Most modern detection systems rely on manually designed components such as anchors and proposals, along with post-processing steps like Non-Maximum Suppression (NMS). Carion et al. introduced DETR [5], an end-to-end system simplifying target detection into a direct set prediction problem. It utilizes the deep relational modeling capabilities of Transformers for object queries and pairs sample matching and box positioning with bipartite graph matching the Hungarian algorithm. However, DETR still has shortcomings in the convergence speed and detection performance of small objects. To address this, Zhu et al. adopted a deformable self-attention mechanism [14], which enhances the recognition of small objects by strategically sampling key areas of the image and fusing multi-scale features. Dai et al. further developed dynamic attention techniques [15], dynamically adjusting according to the importance of scale, spatial location, and feature dimensions to optimize performance and accelerate model convergence. However, their methods may cause feature confusion during self-attention computations by merging features of different scales into a single input for the encoder. This study proposes an improved multi-scale DETR design, implementing cross-scale local attention in different paths of the feature pyramid. This not only maintains feature consistency but also effectively reduces the confusion between small defects and the background by focusing on foreground features.

In the single-stage models of DETR, object queries are randomly initialized. DINO [16] enhances position queries by selecting top-K features from the Encoder output, enriching the prior information of object queries. RT-DETR [17] adopts an IoU-aware query

selection method, picking high-scoring feature samples to optimize the decoder's input. DEYOv3 [18] combines elements of DETR with the YOLO architecture to achieve real-time object detection capabilities. Conditional DETR [19] introduces a conditioning mechanism on the decoder's attention, allowing for more flexible and targeted learning, which helps the model to better understand complex scenes and detect objects more accurately. DN-DETR [20] improves upon the deformable DETR by integrating a denoising strategy that enhances the model's robustness and accuracy, particularly in cluttered environments.

These models build on the foundation of the DETR by introducing modifications that target specific weaknesses, such as slow convergence times and difficulty with small object detection. These advancements reflect ongoing efforts to optimize the efficiency and applicability of Transformer-based object detection systems.

### 2.3. Detection of Small Objects

Current technologies for detecting small objects primarily modify and optimize standard object detection frameworks to accommodate small-sized targets. These techniques can generally be classified into four mainstream directions based on their improvement strategies:

- 1. Data augmentation methods are crucial for training data-driven deep learning models, especially in the field of small object detection, where standard sampling techniques often fail to provide sufficient training samples [21]. One approach is to replicate and randomly transform small objects within the same image, enlarging the sample size of small objects and enhancing the model's detection capability. Another strategy involves new label assignments to include more small objects in training.
- 2. Multi-scale fusion methods allow for recognition by high-level features in deep networks, which are typically small in scale and do not encompass low-level features [22]. Considering that small object features are often present in low-level features, PANet [23] leverages the concept of inter-layer feature interaction from FPN, enhancing feature hierarchy via bi-directional paths, thus improving the precision of deep feature localization. Yang et al. [24] adopted Scale-Dependent Pooling (SDP) to choose appropriate feature layers for the pooling steps of small objects. MS-CNN [25] generates object proposals for specific scale ranges, providing optimal receptive fields for small objects.
- 3. Attention mechanism methods [26] mimic the human ability to rapidly scan scenes while ignoring irrelevant parts, with our perceptual system utilizing visual attention mechanisms crucially during recognition. Attention models assign different weights to different parts of the feature map, highlighting important areas while suppressing secondary information.
- 4. Contextual information methods utilize spatial correlations, providing additional clues and evidence. Contextual information plays a crucial role in human visual systems and scene-understanding tasks. Some methods enhance the detection of small objects by leveraging contextual clues. Chen et al. [27] used context representations encompassing the proposal region for recognition. Hu et al. [28] explored encoding areas beyond the object range and modeled local context information in a scale-invariant manner to detect small objects. Based on this, we combine contextual information with attention to better distinguish the relationships between foreground and background.

### 3. Method

### 3.1. Overview

The aim of this research is to utilize the DETR architecture for the identification of defects in insulators, with a special emphasis on enhancing the detection precision of minute flaws. As demonstrated in Figure 1, the IF-DETR (Insulators Flaws DETR) model is an end-to-end crater identification model based on Transformer technology.



**Figure 1.** Structure of the IF-DETR model. There are two innovations in this model; one is the fusion of low-level features from the backbone network with high-level features from the transformer via the SAU module to further capture contextual information. The other is to solve the instability problem caused by small defects in the matching process via the IDIoU loss function and to speed up the training.

This model employs a multi-scale feature fusion approach, integrating features from before and after the encoder at various scales. We use the SAU (Spatial Attention Upsampling) [29] operation for upsampling, allowing the enhanced high-level features from the encoder to extend to the low-level features through cross-scale fusion, further capturing contextual information. As shown in Figure 1, the low-level feature F2 from the backbone network is fused with the high-level feature S3 from Transformer to obtain S2 through the SAU module, and F2 is fused with S2 in the same way. In the Head section, multi-scale features processed by cross-scale fusion are fed into the transformer decoder, thus increasing the number of queries involved in training. Ultimately, to simplify the optimization of small objects, we have designed a specialized stable IoU (Intersection over Union) loss function [30] for small objects.

Multi-scale Architecture In convolutional neural networks, it has been proven by Cai [31] that multi-scale feature fusion can effectively improve detection accuracy. However, the increased length of input sequences still leads to significant computational demands. To overcome this obstacle, we rethought the encoder structure and designed a series of variants with different encoders, as shown in Figure 1. Feature fusion at two scales not only achieves good detection results but also avoids introducing excessive computation. Applying self-attention to high-level features with richer semantic concepts can capture the relationships between conceptual entities within images, aiding subsequent modules in the detection and identification of objects in images. Moreover, the fusion of features across multiple levels can effectively extract contextual information around defects to prevent confusion between the foreground and background.

The algorithm steps for IF-DETR are as follows (Algorithm 1):

Algorithm 1. IF-DETR
Input: Images from drones capturing high-altitude power lines
Output: Annotated images with detected insulator defects
Begin
// Step 1: Preprocess the Input Images
for each image in drone-captured images do
Normalize(image)
Resize (image, required_input_size)

Algorithm 1. Cont.
// Step 2: Feature Extraction with Multi-Scale Backbone
Initialize multi_scale_backbone_network
for each preprocessed_image in preprocessed_images do
$features \leftarrow multi\_scale\_backbone\_network (preprocessed\_image)$
$enhanced_features \leftarrow apply\_context\_based\_attention (features)$
enhanced_features $\leftarrow$ self_attention_upsampling (enhanced_features)
// Step 3: Defect Detection with Enhanced DETR
Initialize enhanced_DETR_model
for each feature_set in enhanced_features do
detection_results $\leftarrow$ enhanced_DETR_model (feature_set)
// Step 4: Loss Function Optimization
Initialize IDIoU_loss_function
$total_loss \leftarrow 0$
for each result in detection_results do
$loss \leftarrow IDIoU\_loss\_function (result, ground\_truth)$
$total\_loss \leftarrow total\_loss + loss$
Optimize enhanced_DETR_model using total_loss
// Step 5: Post-Processing and Matching
Initialize Hungarian_algorithm
matched_results $\leftarrow$ Hungarian_algorithm (detection_results, ground_truth)

#### End

### 3.2. Self-Attention Upsampling

During feature fusion, traditional upsampling methods primarily include interpolation algorithms and deconvolution. Interpolation has limitations in capturing rich semantic information, and deconvolution is limited due to the use of a uniform kernel. To address this, CARAFE (Content-Aware ReAssembly of FEatures) [32] introduced a novel convolutional upsampling method, which reassembles features in a content-aware manner. However, this approach only integrates high-level features into low-level features without addressing cross-scale feature interactions. We propose a self-attention upsampling layer, as shown in Figure 2. This layer utilizes cross-scale local cross-attention, where each element in the low-resolution feature map calculates cross-attention with the corresponding high-resolution feature region. This ensures semantic consistency and enhances feature expressiveness by fusing high and low-level features. Moreover, it effectively preserves the spatial information of small defects and their context in low-level features, preventing foreground-background confusion.

High-level features  $S_2 \in \mathbb{R}^{c \times h/2 \times w}$  are enhanced by the features  $F_1 \in \mathbb{R}^{c \times h/2 \times w/2}$ from the transformer encoder, and the high-resolution low-level features  $F_1 \in \mathbb{R}^{c \times h \times w}$ are extracted from the backbone network. A "2 × 2" sliding window with a stride of 2 is used. As the sliding window moves across  $F_1$  once, a  $Query_i \in \mathbb{R}^{c \times w \times w}$  is obtained via the self-attention mechanism. Meanwhile, the i-th feature pixel in S is considered as  $Key_i \in \mathbb{R}^{c \times 1 \times 1}$  and  $Value_i \in \mathbb{R}^{c \times 1 \times 1}$ . After linearly projecting  $Query_i$  and  $Key_i$ ,  $Q_i$  and  $K_i$ are obtained. The Key and Value with index i correspond to the Query with index i. The i-th attention weight unit  $A_i$  is defined as the dot product of  $Q_i$  and  $K_i$ . The final weights are obtained using a Softmax operation and are then used to perform a dot product with features  $S_2$  and fuse with low-level features F to obtain the final result S.



**Figure 2.** Self-attention upsampling layer. This figure describes the detailed process of fusing the low-level feature F1 from the backbone network with the high-level feature S2 from the Transformer in Figure 1 to illustrate the working principle of the SAU module.

The calculation of the weights is shown in Equation (1):

$$A_i = Q_i K_i^T, \ W = softmax\left(\frac{Ai}{\sqrt{d}}\right) \tag{1}$$

where d is the normalization factor.

The computation is defined by Equation (2):

$$V = \begin{bmatrix} V_0, V_1, \dots, V_{\frac{hw}{4}} \end{bmatrix}^T$$
(2)

In the matrix *V*, each row is populated with the projection of the *i*-th high-level feature pixel, denoted as Vi.

The resulting feature matrix *S* is obtained as shown in Equation (3):

$$S = V \bigodot W + F \tag{3}$$

#### 3.3. Insulator Defect IDIoU Loss Function

The loss function for two-stage DETR variants typically comprises three components: classification loss, box L1 loss, and IoU-based loss. The loss function of IF-DETR is defined as follows:

$$L(y, y^{gt}) = \lambda_{cls} L_{cls}(c, c^{gt}) + \lambda_{l1} \|b - b^{gt}\|_1 + \lambda_{\text{IDIoU}} L_{\text{IDIoU}}(b, b^{gt})'$$
(4)

Hyperparameters are utilized to modulate the scaling of the three constituent parts of the loss. Additionally, y = [1] and  $y^{gt} = \{c^{gt}, b^{gt}\}$ , respectively denote the prediction and ground truth, where *c* and *b* represent the class and bounding box.  $L_{cls}()$  and  $L_{IDIoU}()$  are the classification loss and IDIoU loss that we have designed to optimize the difficulty associated with defect detection. The classification loss employs Focal loss is defined as follows:

$$L_{cls}(c, c^{gt}) = \begin{cases} -(1-c)^{\gamma} logc, & \text{if } c^{gt} = 1\\ -c^{\gamma} log(1-c), & \text{if } c^{gt} = 0 \end{cases}$$
(5)

We have observed that optimizing small defect detection during the training process presents challenges, primarily due to the defect features being easily conflated with background features. Furthermore, enhancing stability in one-to-one matching is also challenging. The original DETR uses IoU loss to measure the overlap between predictions and ground truth, but IoU loss does not accurately reflect the distance between prediction boxes and actual boxes. Consequently, many DETR variants have opted for GIoU (Generalized Intersection over Union) loss [13] in place of IoU loss. However, when two boxes have a containment relationship, GIoU loss degrades into IoU loss, which does not convey the boxes' relative positions, disadvantaging datasets with densely packed small defects. For example, during the initial stages of training, a single predicted box may encompass multiple small defects. In such cases, GIoU loss contributes little to one-to-one matching, whereas classification loss and box L1 loss become more significant. Additionally, we found that one-to-one matching using only classification scores and L1 scores yields disparate outcomes, leading to instability in the matching process as the model does not optimize toward a unified objective.

To address these issues, we propose the IDIoU loss for small defects based on EIoU (Efficient-IoU) Loss [33]. EIoU Loss consists of three components: IoU loss, center distance loss, and aspect ratio loss. We found that even if the predicted box is within the actual box, the EIoU Loss can still effectively discriminate the predicted box's relative position and incorporate the supervision of the box L1 loss into the IoU loss. Our specially designed IDIoU Loss for small defects uses classification scores and optimizes all components in a unified direction. We introduce the Defect IoU Loss to accommodate this enhancement, as illustrated by Equation (6):

$$L_{\text{IDIoU}} = 1 + rIoU - IoU + \frac{\rho^2(o, o^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2}$$
(6)

where *o* and  $o^{gt}$  denote the central prediction box and ground truth box,  $w^c$  and  $h^c$  are the width and height of the smallest enclosing box of the two boxes.  $\rho^2()$  represents the Euclidean distance between two points. r represents the classification scores.

Figure 3 illustrates the multiple defects within the prediction box, as well as the different matching results of individual classification scores and individual L1 scores. The term "L1" means the L1 scores, and the term "CLS" means the classification cores.



Figure 3. The graphical explanation of the mathematical model of the loss function.

The prediction Box A in the figure contains multiple small defects. The GIoU between the GT (ground truth) of these defects and prediction Box A degrades to IoU. The prediction Box B has a low L1 score and high classification score, and the prediction Box C is the opposite. This problem can lead to different results when choosing who matches the GT with A or B at different stages.

# 4. Experiment

# 4.1. Dataset

In the realm of insulator fault detection, data are relatively scarce, and research typically relies on proprietary datasets. Presently, there are two mainstream publicly available datasets: The Chinese Power Line Insulator Dataset and the augmented version of the same dataset. We conducted comparative experiments on these two datasets against state-of-the-art methods and performed ablation studies on the backbone network structure, loss function selection, and upsampling mechanism choice. As the primary objective of this paper is to enhance detection accuracy, our experiments primarily utilize recall and precision as evaluation metrics.

IFDD Dataset

The IFDD dataset includes 1600 image samples, which are divided in a 7:2:1 ratio. It encompasses a sufficient number of defect samples, covering most of the defect scenarios found in insulators in daily life, with a wide distribution of data. Figure 4 shows some samples. The dataset contains data of various sizes, predominantly small and medium-sized targets, which aligns more closely with real-world scenarios.



Figure 4. Samples of the IFDD Dataset.

CPLID Dataset

The Chinese Power Line Insulator Dataset (CPLID) comprises images of normal and defective insulators captured by Unmanned Aerial Vehicles (UAVs) for the State Grid Corporation of China. Specifically, CPLID contains 600 images of normal insulators and 248 images of defective insulators; Figure 5 shows some samples. Given the limited number of defective insulator images, some were generated by segmenting defective insulators and superimposing them onto different backgrounds.



Figure 5. Samples of the CPLID Dataset.

AI-TOD Dataset

The AI-TOD (Aerial Images-Tiny Object Detection) is a large-scale horizontal box dataset released by Wang et al. [34] from Wuhan University in 2020 for small object detection

in optical remote sensing images. This dataset consists of 28,036 images, all of which are 800 pixels  $\times$  800 pixels in size. The average size of object instances is approximately 12.8 pixels. The proposal of these data fill the gap in the field of small object detection in optical remote sensing images, providing data support for the development of subsequent small object detection tasks in optical remote sensing images.

### 4.2. Training Settings

In our experiments, we utilized a Resnet model [35] pre-trained on ImageNet (IS-LVRC2012) as the backbone network for extracting multi-scale features. We adopted a linear warm-up strategy for learning rate optimization, settling on a final learning rate of 0.0002. In addition, the loss function was optimized over 200 epochs using the AdamW (Adam with Weight Decay Fix) optimizer with a weight decay of 0.0001. The anchors within the auxiliary head were aligned with those used in YOLOv3 [36]. All experiments were implemented in PyTorch 1.7.0 and carried out on a Tesla A100 GPU.

### 4.3. Evaluation Metrics

To assess the detection quality of our model, we used precision and recall as performance metrics on the dataset, similar to most detection methods. We categorized predicted boxes with an IoU greater than 0.5 compared to the ground truth as true positives and the rest as false positives. The specific formulas are as follows:

$$Precision = \frac{\|TP\|}{\|TP\| + \|FP\|}$$
(7)

$$Recall = \frac{\|TP\|}{\|TP\| + \|FN\|}$$
(8)

Prediction boxes that do not contain defects are classified as false negatives (FN). Prediction boxes that contain defects and have an IoU greater than 0.5 with the ground truth are classified as true positives (TP), while those with an IoU less than or equal to 0.5 are classified as false positives (FP). ||TP||, ||FP||, and ||FN|| represent the counts of true positives, false positives, and false negatives, respectively. Higher accuracy and recall rates indicate better detection performance.

### 4.4. Compared with Other Methods

On the IFDD dataset, we compared the IF-DETR with other advanced DETR variants. Table 1 displays the results of these comparisons. It is evident from the table that our model achieved the highest precision and recall with fewer parameters and made progress in detecting small defects (APs).

Compared to high-performance models, our method averaged a 7.47% improvement in detection accuracy, as shown in Figure 6. These models struggle to identify defects with weak features or in complex backgrounds because they unify multi-scale features into a single input for the encoder, leading to increased computational cost and potential feature confusion. In contrast, our model integrates a feature fusion step between the encoder and decoder, effectively ensuring that features of small craters in high-resolution feature maps are not overlooked during the learning process. Additionally, our training strategy ensures that the encoder learns higher-quality features.

Compared to the baseline model RT-DETR, IF-DETR showed more than a 2% increase in precision. This is due to the fact that RT-DETR's interpolation upsampling and convolutional pooling fail to ensure semantic coherence of features. Moreover, RT-DETR's sparse sampling method for object query selection limits the decoder's decoding capabilities. Our proposed SAU operator ensures the consistency of upsampled features, which are suitable for networks based on self-attention computation. DINO [16]

RT-DETR [28]

IF-DETR (Ours)

R50

R50

R50

12

72

72

		-					
Model	Backbone	#Epcohs	#Param (M)	FLOPS	Precision (%)	Recall (%)	Aps (%)
Faster-RCNN [37]	R50	24	-	-	80.5	65.2	29.5
ATSS [38]	R50	24	-	-	80.6	70.3	32.5
YOLOv5 [11]	-	275	-	-	81.3	67.7	28.5
YOLOv7 [12]	-	275	-	-	89.5	86.7	30.2
IYOLOv7 [13]	-	275	-	-	91.1	88.4	34.3
Deformable-DETR [19]	R50	50	40.2	170.1	81.2	81.5	26.2
Conditional-DETR [39]	R50	50	43.3	63.7	82.1	68.3	25.5
Group-DETR [40]	R50	50	44.1	63.2	78.3	78.2	31.6
Focus-DETR [41]	R50	36	49.5	113.6	88.1	83.1	33.4

Table 1. Comparison of different methods in IFDD Dataset.

47.6

41.4

39.3

Note: The missing part of the data is based on CNN methods, which are not comparable to those based on DETR methods.

89.8

90.6

92.3

83.4

87.5

89.6

178.4

93.4

90.1



Figure 6. Comparison of accuracy on the IFDD dataset [16,27,35-43].

In Figure 7, we make a separate comparison for small defects. Compared with the baseline model, our model's instability score drops faster in the early stages of training, and it is more stable in the later stages of training. It could present that our method effectively mitigates the instability of small defect matching and accelerates the optimization process.



Figure 7. Comparison of RT-DETR and IF-DETR instability scores on small defect.

Figure 8 showcases the detection results of four high-performance DETR variants and our proposed method. Figure 8 illustrates the three sample cases. In the sample case

33.7

34.1

35.2

of Figure 8a, only two defects were found by YOLO5, and three defects were detected by Deformable-DETR, DINO, and RT-DETR. Four defects were detected by our IF-DETR model, which has higher reliability. In the case of Figure 8b, the other models detect only one defect, and the IF-DETR model detects three defects. The results show that YOLO5, Deformable-DETR, DINO, and RT-DETR perform poorly in identifying small defects and have missed detections. Our algorithm's precise design effectively identifies small craters with less evident edge features, thus demonstrating exceptional detection performance on such objects.



ALL AND

RT-DETR

Ours (a)



RT-DETR

Ours (b)

Figure 8. Cont.



**Figure 8.** (a) Comparison of detection results containing many defects. (b) Comparison of detection results for small defects. (c) Comparison of detection results in complex backgrounds.

In Figure 8c, our model displays an advantage in handling images with complex backgrounds. The complex texture features in these images pose a challenge to detection as they resemble defects. Although complex backgrounds challenge all detectors, our method is least affected and accurately detects defects where other methods fail, benefitting from our model's proposed self-attention upsampling module, which effectively differentiates between foreground and background, resulting in superior feature learning capabilities and enhanced robustness of the model.

Overall, our algorithm has clear advantages in feature learning and distinguishing between foreground and background features. It improves the accuracy of defect detection, especially for small defects, using more effective multi-scale feature fusion techniques and a more stable matching process.

### 4.5. Ablation Study

We conducted a series of ablation experiments to assess the specific contributions of each component to the model's performance. By incrementally incorporating the three innovative points we proposed into the robust baseline model, the detection results improved. These experimental results are summarized in Table 2. Initially, we utilized a multi-scale backbone network structure within the classic DETR, which improved the precision of defect detection by 0.9%, underscoring the benefits of multi-scale feature fusion under the DETR framework. Subsequently, by integrating the SAU module into the multi-scale bottom-up path, the model's detection precision increased by 0.4%, reflecting the efficacy of the SAU module in emphasizing small defect features. The model's performance increased by 0.4%, demonstrating the appropriate upsampling enhancement of the model's ability to differentiate foreground from background. Finally, by replacing the original loss function with the IDIoU loss, the model's detection precision increased again by 0.4%, indicating the beneficial impact of IDIoU loss on network training. In summary, our experiments demonstrate that each component of the proposed algorithm effectively enhances the model's performance, and they work in coordination without any conflict.

Model	Multi-Scale	SAU	IDIoU	Precision (%)	Recall (%)
Baseline	-	-	-	90.6	87.5
	$\checkmark$	-	-	91.5 ↑0.9	88.3 ↑0.8
Baseline + Parts	$\checkmark$	$\checkmark$	-	91.9 †1.3	$88.9\uparrow1.4$
	$\checkmark$	$\checkmark$	$\checkmark$	92.3 †1.7	89.6 †2.1

Table 2. Influence of each component.

# 4.6. Generalization Study

To further validate the robustness and broad applicability of the proposed model in handling insulator defect detection, we conducted experiments on a subset of the CPLID dataset. According to the results in Table 3, our model continues to demonstrate exceptional detection capabilities. From the experimental results, the defects in the CPLIID data are relatively simple, and existing algorithms have very high detection accuracy. Our model has an accuracy of 99.2% on the CPLID dataset, which is only lower than the 99.4% of Hybrid YOLO [42]. Compared with the best algorithm, Hybrid-YOLO, our method is comparable in precision, but we have improved by 2.6% in recall rate. This indicates that our method can identify as many defects as possible, thanks to a series of innovative points we proposed, which are more conducive to actual industrial applications. In addition, compared with DETR-based methods, our method also has significant improvements, which shows that our method is strong and robust and can achieve better experimental results across multiple data distributions.

Table 3. Comparison of different methods in the CPLID dataset.

Model	Precision (%)	Recall (%)
YOLO-5 [11]	95.2	96.3
Hybrid-YOLO [42]	99.4	93.1
Deformable-DETR [19]	90.2	84.4
Focus-DETR [41]	91.5	85.2
DINO [16]	97.1	92.1
RT-DETR [28]	98.3	93.6
IF-DETR (Ours)	99.2	95.7

To illustrate the robustness and generalization of the proposed model, especially for tiny objects, we do further experiments on a portion of the AI-TOD dataset.

As shown in Table 4, our model maintains the best detection performance. Most aerial images of AI-TOD are taken over the sea. Therefore, they do not have as distinctive textural features as crater images. IF-DETR has powerful feature learning capabilities and can fully fuse features from different scales. In addition, the proposed method can still use color information better to distinguish foreground and background.

Table 4. Comparison of different methods in AI-TOD dataset.

Model	mAP (%)	AP <sub>50</sub> (%)	APs (%)
Deformable-DETR [19]	10.13	30.50	12.57
DAB-Deformable-DETR [43]	10.86	30.84	11.29
DN-Deformable-DETR [20]	12.75	32.58	13.39
H-Deformable-DETR [44]	13.93	36.62	14.01
Focus-DETR [41]	16.09	41.89	16.21
DINO [16]	12.96	34.65	13.18
RT-DETR [18]	17.97	41.73	18.58
IF-DETR (Ours)	18.13	42.08	19.70

## 5. Discussion

### 5.1. Practical and Social Implications

The improvements made in insulator defect detection via the refinement of the DETR model have achieved significant advancements, particularly in detecting minor defects. This progress holds substantial practical importance for the maintenance and management of power systems, including:

- Enhanced detection efficiency and accuracy: Optimized detection models employing deep learning technology can significantly improve the identification efficiency and accuracy of insulator defects within power systems, aiding in the timely discovery and management of potential faults and reducing the incidence of electrical grid failures.
- 2. Reduced maintenance costs and risks: Compared to traditional manual inspection methods, automated defect detection technology reduces manpower requirements, lowers inspection costs, and minimizes the safety hazards of personnel working in high-risk environments.
- 3. Optimized maintenance strategy: Accurate defect detection results provide a reliable basis for electrical grid maintenance and repair, enabling more rational resource allocation and more targeted and efficient maintenance planning.

Furthermore, the study's advancements in insulator defect detection have had positive impacts on environmental protection, economic development, and public safety and welfare. By improving the efficiency and accuracy of grid maintenance, this research contributes to reducing energy waste and lowering the risk of grid accidents, thereby protecting the environment from pollution and degradation. Additionally, the development and application of efficient defect detection technology drive innovation and industry upgrades, providing momentum for the growth of emerging industries such as smart grids and automated detection equipment manufacturing while also creating new employment opportunities. The improved technology also enhances the reliability of the power system, boosting public confidence in grid safety and thereby promoting public safety and social welfare. Overall, this research showcases how technological advancements can play a significant role in environmental protection, economic growth, and the enhancement of public safety.

### 5.2. Limitations and Future Research Directions

Despite progress, some limitations exist in the study. For instance, the currently used dataset may not fully cover all types of insulator defects, especially those internal cracks that are difficult to detect using external imaging. Moreover, although the model performs well on the selected dataset, its generalization capability across different electrical grid structures and environmental conditions remains to be verified. The efficiency of defect detection against complex backgrounds and the computational efficiency and processing speed of the model in real-time or near-real-time applications also require further attention in future research.

Future research should focus on several key areas to refine the defect detection model further:

Dataset Expansion and Diversification: Enhancing the dataset to include a wider variety of defect types, especially those that are less obvious or occur internally, will be vital. Incorporating a broader range of defect scenarios and conditions will help improve the model's generalization capabilities.

Cross-Environment Adaptability Studies: Conducting detailed studies on the model's performance across various electrical grid structures and environmental conditions will help determine its adaptability and robustness. This involves testing the model under different weather conditions and with various insulator types and configurations to validate its effectiveness universally.

Algorithm Optimization and Architectural Innovation: Optimizing existing algorithms and exploring new model architectures can significantly enhance computational efficiency.

This development is crucial for implementing the model in real-time applications, ensuring it operates both swiftly and accurately.

Advanced Background Processing Techniques: Developing techniques that can effectively segregate defects from complex backgrounds will enhance detection accuracy. This may involve the use of advanced image segmentation and enhancement technologies that can isolate and highlight defect features more distinctly.

Integration of Multi-source Data: Utilizing multimodal learning methods to integrate data from multiple sources could substantially improve the accuracy and reliability of defect detections.

By addressing these areas, future research can significantly advance the state of insulator defect detection, making it more reliable, efficient, and applicable across a broader range of conditions and scenarios.

#### 6. Conclusions

In this comprehensive study, an innovative adaptation of the Detection Transformer (DETR) model, termed IF-DETR, was meticulously developed to specifically target the intricate task of identifying defects within insulators. The primary aim of this model was to substantially enhance the accuracy with which small, often elusive, defects are detected, marking a significant leap forward in the field of defect detection.

The foundation of this advancement lies in the integration of a sophisticated multiscale backbone network seamlessly combined with the classical DETR framework. This strategic integration served to amplify the visibility of smaller defects, thereby facilitating their detection. Building upon this, the study introduced a novel upsampling module, dubbed the SAU (Spatial Attention Upsampling), ingeniously designed to leverage crossscale local cross-attention computation. This module was ingeniously incorporated into the bottom-up pathway of the multi-scale backbone network. Its primary function was to amalgamate features across a myriad of scales, adeptly enhancing the model's capability to discriminate between minute defects and their surrounding background. To further refine the model's accuracy, particularly in the context of one-to-one matching of small defects often likened to small craters—a novel loss function, the IDIoU loss, was conceptualized and implemented. This loss function plays a pivotal role in the model by calculating the matching cost and the box regression loss, thereby substantially bolstering the stability of matching small defects, which is crucial for reliable defect detection.

The proposed IF-DETR model improves the detection accuracy of small target defects in complex backgrounds, which is crucial for the reliability and safety of insulator equipment and has broader implications for the maintenance and inspection of high-voltage power infrastructure. Future endeavors will focus on expanding the current dataset. This expansion is not merely about increasing the quantity of data but is aimed at encompassing a broader spectrum of defect types, including internal cracks. Such an enriched dataset will not only enhance the model's detection capabilities but will also contribute to the broader applicability and reliability of the IF-DETR model in real-world scenarios.

**Author Contributions:** Data curation, D.L. and P.Y.; Software, D.L. and P.Y.; Validation, D.L. and P.Y.; Writing—review & editing, D.L.; Investigation, P.Y.; Conceptualization, Y.Z.; Formal analysis, Y.Z.; Methodology, Y.Z.; Writing—original draft, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

**Data Availability Statement:** The IFDD Dataset can be downloaded from https://doi.org/10.608 4/m9.figshare.21200986 (accessed on 7 May 2024). The CPLID Dataset can be downloaded from https://github.com/InsulatorData/InsulatorDataSet (accessed on 7 May 2024). The AI-TOD Dataset can be downloaded from https://github.com/jwwangchn/AI-TOD (accessed on 7 May 2024).

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

### References

- 1. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 2016, 117, 11–28. [CrossRef]
- 2. Khan, A.; Gupta, S.; Gupta, S.K. Multi-hazard disaster studies: Monitoring, detection, recovery, and management, based on emerging technologies and optimal techniques. *Int. J. Disaster Risk Reduct.* **2020**, *47*, 101642. [CrossRef]
- Shojanoori, R.; Shafri, H.Z.M.; Mansor, S.; Ismail, M.H. The use of WorldView-2 satellite data in urban tree species mapping by object-based image analysis technique. *Sains Malays.* 2016, 45, 1025–1034.
- 4. Tian, C.; Zhang, X.; Liang, X.; Li, B.; Sun, Y.; Zhang, S. Knowledge Distillation with Fast CNN for License Plate Detection. *IEEE Trans. Intell. Veh.* **2023**, 1–7. [CrossRef]
- Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 213–229.
- 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, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666.
- Zhang, X.; Cheng, S.; Wang, L.; Li, H. Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.* 2023, *61*, 1–15. [CrossRef]
- 8. Wang, L.; Li, H. HMCNet: Hybrid Efficient Remote Sensing Images Change Detection Network Based on Cross-Axis Attention MLP and CNN. *IEEE Trans. Geosci. Remote Sens.* 2022, *60*, 1–14. [CrossRef]
- 9. Zhang, X.; Wan, F.; Liu, C.; Ji, R.; Ye, Q. Freeanchor: Learning to match anchors for visual object detection. *Adv. Neural Inf. Process. Syst.* **2019**, *32*, 147–155.
- 10. Zhu, C.; He, Y.; Savvides, M. Feature selective anchor-free module for single-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 840–849.
- 11. Jocher, G.; Stoken, A.; Borovec, J.; Changyu, L.; Hogan, A.; Diaconu, L.; Poznanski, J.; Yu, L.; Rai, P.; Ferriday, R. ultralytics/yolov5: v3. 0. Zenodo 2020. [CrossRef]
- 12. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
- Zheng, J.; Wu, H.; Zhang, H.; Wang, Z.; Xu, W. Insulator-Defect Detection Algorithm Based on Improved YOLOv7. Sensors 2022, 22, 8801. [CrossRef]
- Zhu, X.; Cheng, D.; Zhang, Z.; Lin, S.; Dai, J. An empirical study of spatial attention mechanisms in deep networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6688–6697.
- Dai, X.; Chen, Y.; Yang, J.; Zhang, P.; Yuan, L.; Zhang, L. Dynamic detr: End-to-end object detection with dynamic attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2988–2997.
- 16. 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. *arXiv* 2022, arXiv:2203.03605.
- 17. Lv, W.; Xu, S.; Zhao, Y.; Wang, G.; Wei, J.; Cui, C.; Du, Y.; Dang, Q.; Liu, Y. Detrs beat yolos on real-time object detection. *arXiv* **2023**, arXiv:2304.08069.
- 18. Ouyang, H. DEYOv3: DETR with YOLO for Real-time Object Detection. arXiv 2023, arXiv:2309.11851.
- 19. Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; Dai, J. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv* **2020**, arXiv:2010.04159.
- Li, F.; Zhang, H.; Liu, S.; Guo, J.; Ni, L.M.; Zhang, L. DN-DETR: Accelerate DETR Training by Introducing Query DeNoising. IEEE Trans. Pattern Anal. Mach. Intell. 2024, 46, 2239–2251. [CrossRef] [PubMed]
- 21. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. arXiv 2019, arXiv:1902.07296.
- 22. Zeng, N.; Wu, P.; Wang, Z.; Li, H.; Liu, W.; Liu, X. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–14. [CrossRef]
- Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9197–9206.
- Yang, F.; Choi, W.; Lin, Y. Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2129–2137.
- Liu, Y.; Liu, J.; Ning, X.; Li, J. MS-CNN: Multiscale recognition of building rooftops from high spatial resolution remote sensing imagery. *Int. J. Remote Sens.* 2022, 43, 270–298. [CrossRef]
- 26. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* 2021, 452, 48–62. [CrossRef]
- 27. Cheng, G.; Yuan, X.; Yao, X.; Yan, K.; Zeng, Q.; Xie, X.; Han, J. Towards large-scale small object detection: Survey and benchmarks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, 1–20. [CrossRef]

- 28. Hu, P.; Ramanan, D. Finding tiny faces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 951–959.
- 29. Xie, Y.; Zheng, S.; Li, W. Feature-guided spatial attention upsampling for real-time stereo matching network. *IEEE MultiMedia* **2020**, *28*, 38–47. [CrossRef]
- 30. Yu, J.; Jiang, Y.; Wang, Z.; Cao, Z.; Huang, T. Unitbox: An advanced object detection network. In Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, 15–19 October 2016; pp. 516–520.
- Cai, Z.; Fan, Q.; Feris, R.S.; Vasconcelos, N. A unified multi-scale deep convolutional neural network for fast object detection. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 354–370.
- 32. Wang, J.; Chen, K.; Xu, R.; Liu, Z.; Loy, C.C.; Lin, D. Carafe: Content-aware reassembly of features. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3007–3016.
- 33. Zhang, Y.-F.; Ren, W.; Zhang, Z.; Jia, Z.; Wang, L.; Tan, T. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing* **2022**, *506*, 146–157. [CrossRef]
- Wang, J.; Yang, W.; Guo, H.; Zhang, R.; Xia, G.S. Tiny Object Detection in Aerial Images. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 3791–3798.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- 36. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. arXiv 2018, arXiv:1804.02767.
- 37. Sun, X.; Wu, P.; Hoi, S.C.H. Face detection using deep learning: An improved faster RCNN approach. *Neurocomputing* **2018**, 299, 42–50. [CrossRef]
- 38. Gomes, M.; Silva, J.; Gonçalves, D.; Zamboni, P.; Perez, J.; Batista, E.; Ramos, A.; Osco, L.; Matsubara, E.; Li, J. Mapping utility poles in aerial orthoimages using atss deep learning method. *Sensors* **2020**, *20*, 6070. [CrossRef] [PubMed]
- Meng, D.; Chen, X.; Fan, Z.; Zeng, G.; Li, H.; Yuan, Y.; Sun, L.; Wang, J. Conditional detr for fast training convergence. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3651–3660.
- 40. Chen, Q.; Chen, X.; Zeng, G.; Wang, J. Group detr: Fast training convergence with decoupled one-to-many label assignment. *arXiv* 2022, arXiv:2207.13085.
- Zheng, D.; Dong, W.; Hu, H.; Chen, X.; Wang, Y. Less is more: Focus attention for efficient detr. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 6674–6683.
- 42. Souza, B.J.; Stefenon, S.F.; Singh, G.; Freire, R.Z. Hybrid-YOLO for classification of insulators defects in transmission lines based on UAV. *Int. J. Electr. Power Energy Syst.* **2023**, 148, 108982. [CrossRef]
- 43. Liu, S.; Li, F.; Zhang, H.; Yang, X.; Qi, X.; Su, H.; Zhu, J.; Zhang, L. DAB-DETR: Dynamic Anchor Boxes are Better Queries for DETR. *arXiv* 2022, arXiv:2201.12329.
- 44. Jia, D.; Yuan, Y.; He, H.; Wu, X.; Yu, H.; Lin, W.; Sun, L.; Zhang, C.; Hu, H. Detrs with hybrid matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 19702–19712.

**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.