



## Article

# Detection of Glass Insulators Using Deep Neural Networks Based on Optical Imaging

Jinyu Wang<sup>1,2</sup>, Yingna Li<sup>1,3,\*</sup> and Wenxiang Chen<sup>1,3</sup>

<sup>1</sup> Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

<sup>2</sup> Electric Power Research Institute of Yunnan Power Grid Co., Ltd., Kunming 650217, China

<sup>3</sup> Computer Technology Application Key Lab of the Yunnan Province, Kunming 650500, China

\* Correspondence: liyingna@kust.edu.cn

**Abstract:** As the pre-part of tasks such as fault detection and line inspection, insulator detection is a crucial task. However, considering the complex environment of high-voltage transmission lines, the traditional insulator detection accuracy is unsatisfactory, and manual inspection is dangerous and inefficient. To improve this situation, this paper proposes an insulator detection model Siamese ID-YOLO based on a deep neural network. The model achieves the best balance between speed and accuracy compared with traditional detection methods. In order to achieve the purpose of image enhancement, this paper adopts the canny-based edge detection operator to highlight the edges of insulators to obtain more semantic information. In this paper, based on the Darknet53 network and Siamese network, the insulator original image and the edge image are jointly input into the model. Siamese IN-YOLO model achieves more fine-grained extraction of insulators through weight sharing between Siamese networks, thereby improving the detection accuracy of insulators. This paper uses statistical clustering analysis on the area and aspect ratio of the insulator data set, then pre-set and adjusts the hyperparameters of the model anchor box to make it more suitable for the insulator detection task. In addition, this paper makes an insulator dataset named InsuDaSet based on UAV(Unmanned Aerial Vehicle) shoot insulator images for model training. The experiments show that the insulator detection can reach 92.72% detection accuracy and 84FPS detection speed, which can fully meet the online insulator detection requirements.

**Keywords:** high-voltage transmission lines; insulator detection; deep neural networks; Siamese network



**Citation:** Wang, J.; Li, Y.; Chen, W. Detection of Glass Insulators Using Deep Neural Networks Based on Optical Imaging. *Remote Sens.* **2022**, *14*, 5153. <https://doi.org/10.3390/rs14205153>

Academic Editors: Liangxiu Han, Wenjiang Huang, Yanbo Huang, Jiali Shang and Stefano Pignatti

Received: 25 August 2022

Accepted: 12 October 2022

Published: 15 October 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Introduction

The insulator is a device used to connect wire and tower, which plays the role of traction wire and electrical insulation [1]. As one of the most critical power transmission methods, high-voltage transmission lines are mainly distributed in non-populated areas such as mountainous areas, forests, grasslands, and farmland. The complex and changeable environment makes the task of insulator detection extremely challenging [2]. Insulator detection methods are divided into power and non-power detection [3]. Power detection uses an electrical way to detect whether the insulator has a leakage phenomenon. Non-power detection is based on the insulator's color, shape, texture, and other image information, using image processing technology to detect whether the insulator has burst, missing sheds, or other faults [3]. Insulators are subjected to the action of voltage and mechanical stress in the atmospheric environment for a long time. They are highly susceptible to failures such as bursting, and cap loss, which will cause the performance of the insulator to decline, thus threatening the safe operation of the line [4]. Therefore, as the pre-part of tasks such as fault detection and line inspection, insulator detection is a crucial task.

At present, the inspection task of insulators mainly relies on artificial visual observation, which is inefficient, costly, and prone to missed detection. Due to their low cost, strong

adaptability, and portability [5], Unmanned Aerial Vehicles (UAVs) have been widely used in the military, industry, ecology, agriculture, geography, environmental science, and other fields. Because drones can collect information about surveillance or high-risk activities, they are widely used in armed conflicts and military operations [6]. Zhang et al. [7] used UAVs for power line identification and vegetation detection tasks. Lyu et al. [8] applied unmanned aerial vehicle remote sensing to grassland ecosystem detection to strengthen regional grassland management and respond to sustainable development strategies. Honkavaara et al. [9] used spectral images of wheat taken by UAV to estimate biomass to guide sowing and fertilization for plant conservation and improve resource utilization. Qin et al. [10] utilize images captured by UAVs at different times for automatic and frequent monitoring of urban areas. Reis et al. [11] used multi-spectral imaging data captured by UAV to monitor forest restoration in real-time to save workforce and material resources. The rapid development of UAV technology has brought many conveniences but also brought long-distance transmission [12], communication capabilities in challenging application scenarios [12], security and privacy issues [13], and other challenges. However, considering the inspection task of UAV high-voltage transmission lines, it is not affected by factors such as harsh environments. It can obtain images that are multiple in the number of manual shots in equal time. This paper focuses on detecting insulators using images of insulators captured by drones.

Power detection uses electrical methods to detect whether there is a leakage phenomenon on the surface of the insulator, including infrared detection [14], red-heat external detection [15], ultraviolet detection [16], Etc., to detect the state of the insulator continuously. However, on the one hand, this detection method is susceptible to electromagnetic interference, which affects the accuracy of judging the leakage phenomenon of the insulator. On the other hand, the state detection of insulators still requires experienced power engineers, which severely limits the efficiency of state detection of insulators. Therefore, research on an automatic insulator detection method with high detection accuracy and fast detection speed has become a hot topic recently.

Non-power detection mainly uses image processing technology to detect insulators on transmission lines. The traditional insulator detection method has fast calculation speed and low resource consumption, but manual feature extraction has low efficiency and poor generalization ability. In addition, it can only locate insulators and detect faults under specific detection conditions [17]. At the same time, it does not fully consider the impact of complex environments such as occlusion and illumination intensity on insulator detection in practical applications.

With the continuous improvement of computer computing power, deep learning algorithms have made remarkable achievements in the field of target detection, and it is also the main research direction in recent years [18,19]. The deep learning method uses its particular network model structure and powerful feature learning capability to learn the internal features of the image rather than a single feature designed manually. The deep learning method significantly improves the recognition accuracy of the target and can meet the actual needs of the insulator detection. The challenge of the deep neural network is to improve the detection speed of the model while the detection accuracy is as high as possible.

Using drones to carry imaging equipment is an efficient means of image acquisition. Compared with imaging devices such as hyperspectral and infrared sensors, optical sensors have the advantages of small size, fast imaging speed, portability, and ease of use [20]. Monitoring transmission lines' critical components based on UAVs carrying optical imaging sensors has received increasing attention [21]. Liu et al. [22] used a YOLOv3-dense network to detect insulators, improving the detection accuracy of different sizes. Ge et al. [23] used a GAN network for defect detection of insulators in complex backgrounds to improve the detection accuracy of insulators in complex environments. Lei et al. [24] used a Faster R-CNN network to complete the high-voltage transmission line insulator detection.

Insulators vary greatly in images due to the differences in light and shooting Angle. At the same time, the background of high-voltage transmission lines is more complex, which brings significant challenges to the research of insulator detection based on optical images. Existing work has achieved certain gains in detection accuracy, but the robustness of the model is generally poor, and the detection speed is unsatisfactory. Therefore, this paper proposes an insulator detection model based on a deep neural network, aiming to achieve a balance between detection accuracy and speed. The main contributions of the article are as follows.

- First, in this paper, the original image is processed by mean grayscale to obtain a grayscale image against when the insulator target and background differences are minor. The canny-based edge detection operator is used to highlight the edge information of the insulator to obtain more semantic information to achieve the purpose of image enhancement.
- Second, based on the Darknet53 backbone network, this paper introduces the Siamese network, which can raise the edge information of the insulator into the model training as incremental information to achieve more fine-grained feature extraction, thereby improving the detection accuracy of the insulator.
- Third, considering that the shape of the insulator in aerial photography of the UAV is a rectangle, the insulators will vary in size due to the different shooting angles and distances. Therefore, to make the insulator data set more suitable for the model in this paper, through clustering statistics on the area and aspect ratio of the insulator dataset, pre-analyze and adjust the hyperparameters of the anchor box of the model to make it fitter for the insulator detection task.

The structure of the rest of this paper is arranged as follows. Section 2 introduces the related work of insulator and insulator fault detection based on traditional images and neural networks. Section 3 introduces the relevant knowledge of the benchmark model IN-YOLO in this paper. Section 4 elaborates on the proposed method in detail. The experimental details and experimental comparison results are analyzed in Section 5. Section 6 summarizes the work done in this paper and looks forward to the next work.

## 2. Related Work

There are mainly two types of insulator inspections based on optical images: traditional image processing and deep learning-based methods to identify and locate insulator defects.

Conventional insulator detection is usually realized by image processing technology. Liao et al. [25] used a method combining local features of insulator images and spatial order to identify insulators and improved the detection accuracy of insulators through a coarse-to-fine strategy. Zhai et al. [26] established a color model for the different color characteristics of glass and ceramic insulators, combined the spatial features of the insulator for its identification and localization, and finally performed morphological positioning of the located insulators to complete insulator fault detection. Yu et al. [27] combined the unique shape features and texture features of insulators and proposed an active contour model to achieve the segmentation of insulators. To solve the problem of unclear insulator fault location, Zhai et al. [17] separated the target and background in the insulator image by fusing the color and gradient features of the insulator. Then, they used adaptive morphology to detect and locate the fault location of the insulator. Zhao et al. [28] proposed a method to identify and detect insulators with a combination of azimuth detection and binary shape prior knowledge (OAD-BSPK) for the problem of changeable insulator angles in aerial images. Generally, this method is faster, but it is affected by uncertain factors such as lighting conditions, shooting angles, and complex backgrounds; it is impossible to detect insulators accurately in complex backgrounds.

In recent years, deep learning models have gradually become a new direction in object detection research with the continuous development of artificial intelligence technology. Chen et al. [4] proposed an insulator detection method based on YOLOv4 and attention mechanism for different sizes of insulators and complex insulator backgrounds, aiming

to accurately locate the position of insulators under the premise of ensuring detection speed. Li et al. [29] introduced the FPN network into the Mask R-CNN model to accurately identify and position insulators of different sizes for the problem of insulators of different sizes and complex insulator backgrounds. Ma et al. [30] proposed a YOLOv4 insulator detection model based on the loss function of joint Gaussian distance intersection, which improved the problems of low detection accuracy and slow positioning speed of insulators. In addition, rectifying and repositioning the tilted insulators also significantly improve the insulator detection accuracy. Gao et al. [31] improved the detection performance of small insulator targets and insulator faults by combining the inclusion of a batch normalized convolutional fast attention model (BN-CBAM) and a feature fusion module. Since the fault area of the insulator occupies a small proportion of the insulator image and is challenging to detect, Zhang et al. [32] proposed introducing a densely connected feature pyramid network into the YOLOv3. YOLOv3-dense network not only improves the detection accuracy of small fault insulators but also improves the detection speed of the model. Chen et al. [33] proposed a second-order fully convolutional network (SOF2Net) insulator detection method for the problem of difficulty in insulator fault detection in complex background images. Without extracting features from images and selecting classifiers, the detection accuracy is improved, and the computational complexity of the model is reduced. Han et al. [34] proposed a cascade model to complete the insulator multi-fault detection problem, using SPPNet to locate insulators from complex backgrounds, setting the positioned insulator strings as RoI, then trained the YOLOv3 model to detect multiple faults of insulators in the region of interest. Liu et al. [35] segmented insulators by combining the Mask R-CNN network and the curve-fitting method. Then defect segmentation of the insulator region is carried out by reconstruction and classification autoencoder network (RCCAEN). Finally, the density-based noise applied spatial clustering algorithm (DBSCAN) is used to evaluate the insulator fault level.

It can be seen from the above research that the complex background of the insulator image and the quality of the image will affect the detection accuracy of the insulator. We summarize these methods as follows.

- The traditional insulator detection algorithm has fast calculation speed and low resource consumption and is suitable for real-time computing on edge devices. However, such methods rely on complex feature extraction and have a low level of automation, making them difficult to apply to many practical applications. Moreover, the input image quality is highly required. The influence of complex environments such as occlusion and illumination intensity in insulator detection is not fully considered in practical applications. Therefore, the feature extraction effect will be affected for scenes with complex backgrounds, and the detection accuracy will also decrease.
- The insulator detection algorithm based on the deep neural network can automatically extract target features from input images in various application scenarios without prior knowledge through iterative training, including the target's spatial and semantic information. Moreover, many insulator detection models based on deep learning inevitably make concessions to insulator detection accuracy to meet real-time requirements.

To sum up, we hope to provide a method to reduce the consumption of hardware resources and improve the calculation speed of the model under the premise that the detection accuracy is as high as possible to match the hardware environment of UAVs and other devices. Therefore, combined with existing insulator datasets and insulator visual features from UAV aerial photography, this paper adopts a one-stage insulator detection scheme.

### 3. Structure of ID-YOLO

The background of high-voltage transmission lines is complex and extracting insulators' underlying features and high-level semantic information is difficult. In other words, insulators' details and location information are challenging to construct. The deep neural network has powerful information construction capabilities. This paper's backbone

network of ID-YOLO (Insulator Detection based on YOLO) is constructed based on the Darknet53 [36] network. The feature fusion network is built based on the Feature Pyramid Network (FPN) [37] and Path Aggregation Network (PAN) [38] structures. Finally, the detector of ID-YOLO is established based on YOLOv4 [39]. ID-YOLO can obtain fine-grained insulator features, which is conducive to improving the detection effect of the model. Meanwhile, the lightweight network structure can match the edge hardware where computational resources are scarce.

The target detection network usually requires a backbone to extract feature information. Classic backbones include ResNet, VGG16, Darknet53, Etc. Darknet53 utilizes the residual structure while abandoning the use of pooling layers, reducing the impact of gradient explosion or gradient disappearance. Compared with ResNet101, Darknet53 guarantees classification precision while reducing the number of network layers and improving the calculation speed.

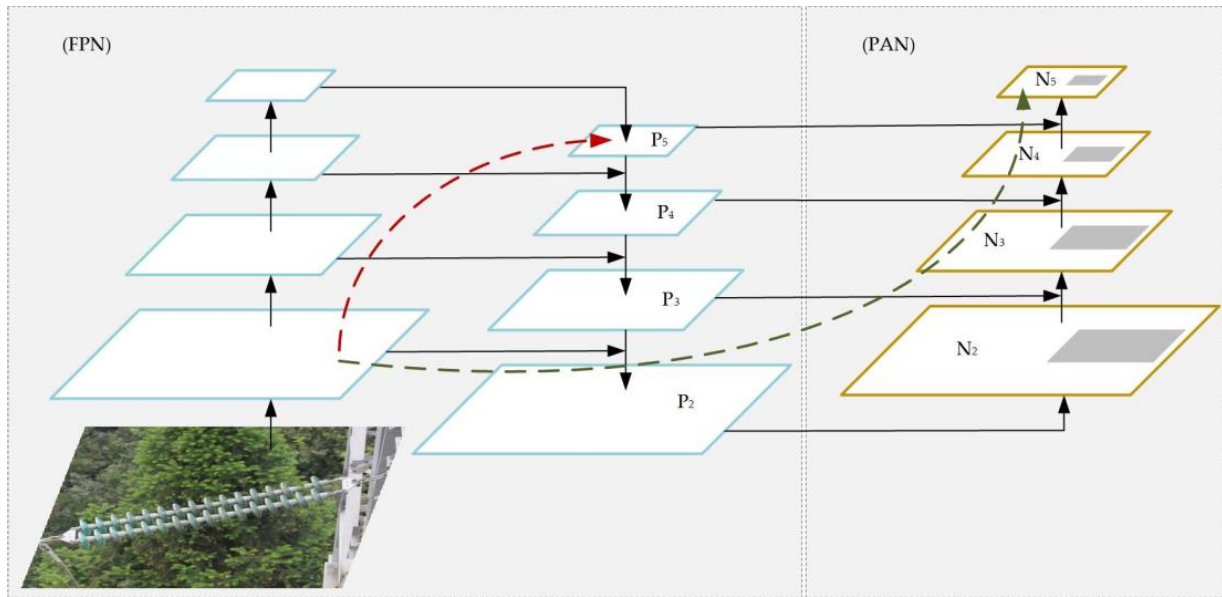
Therefore, this paper constructs a feature extraction network based on Darknet53, which is the backbone of ID-YOLO. This paper differs from the case where many types of targets exist in the universal target detection dataset. There are the only objects to be detected in this paper are the class of insulators. Therefore, as shown in Table 1, the backbone constructed in this paper reduces the use of residual structures. At the same time, the lightweight design can improve the overall operation speed of ID-YOLO.

**Table 1.** Applied kernels of backbone in ID-YOLO.

Number	Layer Information	Input	Output
	CONV-(N32,K3x3,S1,P1),IN,Leaky ReLU	(608,608,3)	(608,608,32)
	CONV-(N64,K3x3,S2,P1),IN,Leaky ReLU	(608,608,32)	(304,304,64)
Res1 ×	CONV-(N32,K1x1,S1,P0),IN,Leaky ReLU	(304,304,64)	(304,304,32)
	CONV-(N64,K3x3,S1,P1),IN,Leaky ReLU	(304,304,32)	(304,304,64)
	CONV-(N128,K3x3,S2,P1),IN,Leaky ReLU	(304,304,64)	(152,152,128)
Res2 ×	CONV-(N64,K1x1,S1,P0),IN,Leaky ReLU	(152,152,128)	(152,152,64)
	CONV-(N128,K3x3,S1,P1),IN,Leaky ReLU	(152,152,64)	(152,152,128)
	CONV-(N256,K3x3,S2,P1),IN,Leaky ReLU	(152,152,128)	(76,76,256)
Res4 ×	CONV-(N128,K1x1,S1,P0),IN,Leaky ReLU	(76,76,256)	(76,76,128)
	CONV-(N256,K3x3,S1,P1),IN,Leaky ReLU	(76,76,128)	(76,76,256)
	CONV-(N512,K3x3,S2,P1),IN,Leaky ReLU	(76,76,256)	(38,38,512)
Res4 ×	CONV-(N256,K1x1,S1,P0),IN,Leaky ReLU	(38,38,512)	(38,38,256)
	CONV-(N512,K3x3,S1,P1),IN,Leaky ReLU	(38,38,256)	(38,38,512)
	CONV-(N1024,K3x3,S2,P1),IN,Leaky ReLU	(38,38,512)	(19,19,1024)
Res2 ×	CONV-(N512,K1x1,S1,P0),IN,Leaky ReLU	(19,19,1024)	(19,19,512)
	CONV-(N1024,K3x3,S1,P1),IN,Leaky ReLU	(19,19,512)	(19,19,1024)

The current object detectors usually add a feature fusion layer named neck after the backbone to obtain feature maps at different semantic levels. The neck can solve the problem of the too significant size difference between the detected objects. At the same time, the feature maps with different resolutions are beneficial for improving the detection effect.

The FPN is a network consisting of bottom-up paths, top-down paths, and lateral connections, which enhances the communication of semantic and detailed information to improve the final detection accuracy. The PAN adds a bottom-up feature fusion layer based on FPN and uses adaptive feature pooling to aggregate features between different layers. Therefore, this paper constructed the neck part of ID-YOLO based on FPN and PAN, and its structure is shown in Figure 1.



**Figure 1.** The neck part of ID-YOLO based on FPN and PAN.

The head for predicting the object class and box position is the last part of the detector. The existing head part is divided into one-stage and two-stage. The most classic method in the two-stage class model is the R-CNN series, which first uses RPN to predict the box's position and category roughly, then gives more accurate results through the head. The two-stage method has better detection accuracy but high time consumption. The one-stage model is based on the feature map giving the prediction result directly. Although the detection effect of the one-stage model is slightly worse than that of the two-stage method, the detection speed is significantly improved. The more classic one-stage models include YOLOv4, SSD, RetinaNet, etc., among which YOLOv4 is a method for obtaining excellent performance in speed and detection accuracy. Therefore, this article builds the head part of ID-YOLO based on YOLOv4.

The ID-YOLO model divides the input image into  $N \times N$  regions and gives the prediction result based on the feature map output from the neck part. Therefore, the loss function of ID-YOLO consists of three components, the localization loss function  $L_{box}$ , the classification loss function  $L_{cls}$ , and the target confidence loss function  $L_{obj}$ .

$$L_{det} = L_{box} + L_{obj} + L_{cls} \quad (1)$$

The position regression loss is used to calculate the position information error between the ground-truth and the prediction box, which is commonly using IoU (intersection over Union) to measure:

$$L_{box} = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{i,j}^{obj} \left( 1 - \left( IoU - \frac{Distance\_2^2}{Distance\_C^2} - \frac{v^2}{(1 - IoU) + v} \right) \right) \quad (2)$$

$$IoU = \frac{|Box^p \cap Box^{gt}|}{|Box^p \cup Box^{gt}|} \quad (3)$$

$$v = \frac{4}{\pi^2} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w^p}{h^p} \right)^2 \quad (4)$$

where  $Box^p$  and  $Box^{gt}$  represents the predicted bounding box, and the ground-truth bounding box, respectively.  $IoU$  represents the intersection ratio of the predicted box and the ground truth box,  $Distance\_2$  represents the Euclidean distance between the center points of the predicted box and the ground-truth box,  $Distance\_C$  represents the diagonal distance

of the smallest closure region that can contain both the predicted box and the ground-truth box.  $w^p, h^p, w^{st}, h^{st}$  represent the width and height of the predicted box and the ground truth box, respectively.  $v$  is used to measure the similarity of aspect ratio and  $\frac{v}{(1-IoU)+v}$  is the weight coefficient.  $\lambda_{coord}$  indicates the weight coefficient,  $S^2$  indicates that the input image is divided into the number of grids,  $B$  indicates that each grid predicts  $B$  bounding boxes, and  $1_{i,j}^{obj}$  demonstrates that when there is an object detected in the  $j$ -th bounding box of the  $i$ -th grid, which is equal to 1, and it is equivalent to 0 if it does not exist.

The confidence loss  $L_{obj}$  is expressed by the cross-entropy loss function as:

$$L_{obj} = \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{i,j}^{noobj} (c_i - \hat{c}_i)^2 + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{i,j}^{obj} (c_i - \hat{c}_i)^2 \quad (5)$$

where  $s$  represents the auxiliary variable.  $\lambda_{obj}$  and  $\lambda_{noobj}$  are the weight coefficients used to balance positive samples and negative samples.  $c_i$  and  $\hat{c}_i$  denote the actual and predicted categories in the  $j$ -th candidate box of the  $i$ -th grid, respectively.

The classification loss  $L_{cls}$  is expressed by the cross-entropy loss function as:

$$L_{cls} = \lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{i,j}^{obj} \sum_{c \in classes} p_i(c) \log(\hat{p}_i(c)) \quad (6)$$

where  $\lambda_{class}$  indicates the class loss weight.  $p_i(c)$  and  $\hat{p}_i(c)$  represent the actual probability and predicted probability that the target belongs to category  $c$  when the  $i$ -th grid detects a certain target, respectively.

#### 4. Insulator Detection Using Siamese ID-YOLO

ID-YOLO has achieved good results in the insulator detection task, but there is still room for improvement in accuracy. As shown in Figure 2, this paper constructed a Siamese xxx network model that is more suitable for insulator detection tasks. Among them, this paper extracts the edge information in the image. Based on the Siamese network, input edge information as gain information into ID-YOLO. In addition, based on cluster analysis, the head part of ID-YOLO regarding the anchor is modified to make it more suitable for the insulator detection task. The training, and testing process of the Siamese ID-YOLO model is shown in Algorithm 1, and Algorithm 2, respectively.

---

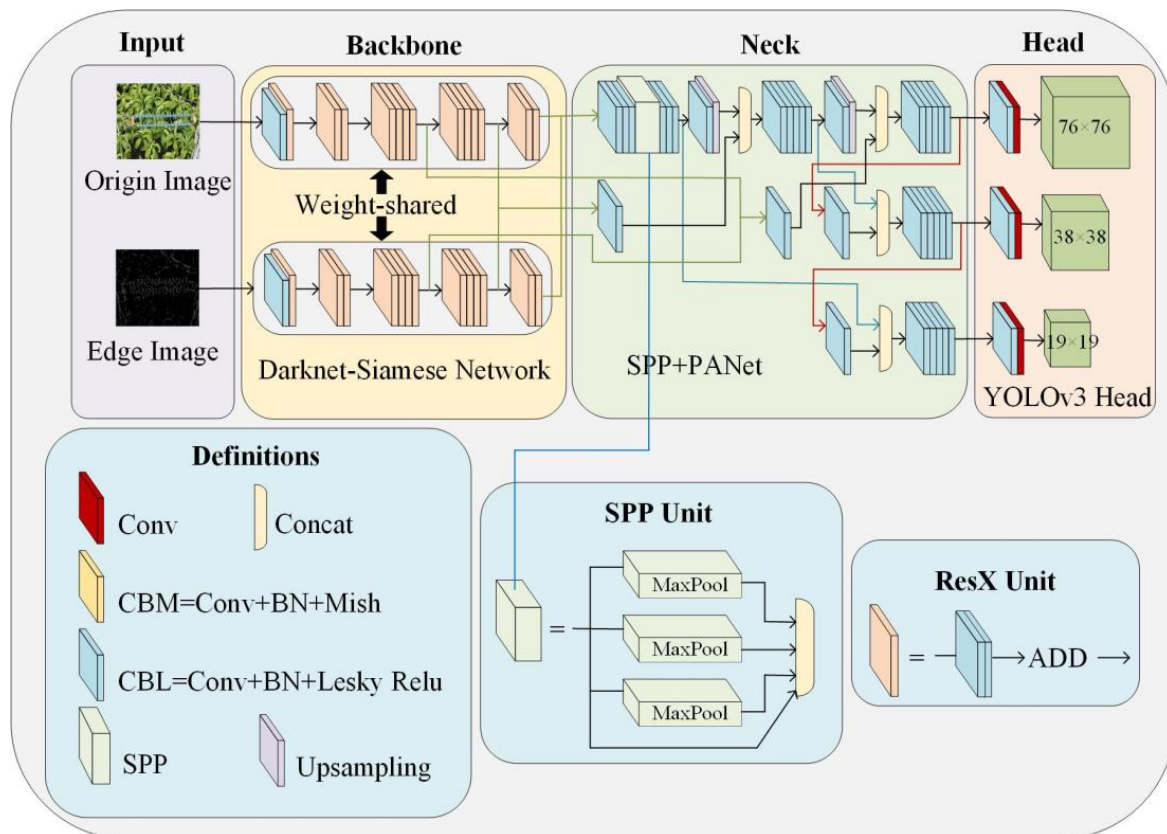
**Algorithm 1.** The training process of Siamese ID-YOLO.

---

Input: The edge image dataset  $Y = \{Y_1, \dots, Y_N\}$  and the labelling dataset  $Z = \{Z_1, \dots, Z_N\}$  of insulators. Among them, each image contains insulators.

Output: The trained model Siamese ID-YOLO.

- 1: Initialize Siamese ID-YOLO with random weights;
  - 2: repeat
  - 3: for  $i$  in 1~epochs do
  - 4: for  $j$  in 1~N do
  - 5: Extract the features of  $Y_i$  and  $Z_i$  using backbone network based on Darknet53;
  - 6: Presets adjust the size and scale of anchor boxes, and a feature map of  $Y_i$  and  $Z_i$  is obtained through the Siamese network and neck network;
  - 7: Output insulator detection results using the Siamese ID-YOLO;
  - 8: Calculate the penalty value via Equation (2), (5), and (6).
  - 9: Update parameters of Siamese ID-YOLO by minimizing Equation (1);
  - 10: end for
  - 11: end for
  - 12: until Siamese ID-YOLO finishes convergence
  - 13: return
-



**Figure 2.** The insulator detection model Siamese ID-YOLO includes the backbone, neck, head, and Siamese network. The backbone extracts feature information. The neck can solve the problem of the too significant size difference between the detected objects. The head used to predict the target category and box position is the last part of a detector. Based on the Siamese network, input edge information as gain information into Siamese ID-YOLO.

---

**Algorithm 2.** The testing process of Siamese ID-YOLO.

---

Input: The original insulator dataset  $T = \{T_1, \dots, T_M\}$  of insulators;

Output: Insulator detection results using Siamese ID-YOLO.

1: Siamese ID-YOLO training obtained weights;

2: repeat

3: for  $i$  in  $1 \sim M$  do

4: Feature map extraction for  $T_i$  using backbone network based on Darknet53 network, Siamese work, and neck network.

5: Output insulator detection results using the Siamese ID-YOLO;

6: end for

7: return

---

#### 4.1. Edge Extraction

The image quality has a considerable impact on the detection accuracy of the model. High-voltage transmission lines are primarily located in complex environments such as mountains, rivers, grasslands, and farmland, which leads to the difference between insulators and the background being small and difficult to distinguish. To obtain higher semantic information, this paper adopts an edge detection algorithm to highlight the contour information of insulators for image enhancement.

The edge detection algorithm needs to take a grayscale image as input. The typical image grayscale methods include the maximum value, average value, weighted average, and Gamma correction methods. The grayscale image by the maximum value method is



too bright. It loses more detailed information, so it is often used to process images with darker original images. The image after grayscale by the weighted average method is too dark, and it is not easy to distinguish the foreground and background, so it is often used to process the highlighted areas of the image. The effect of the Gamma correction grayscale is better. However, the calculation method is complicated, and the operation time is long, so it is not suitable for processing a large number of pictures. The effect of mean grayscale is similar to that of Gamma correction, which is simple to compute and has a better outcome. Therefore, this paper adopts mean grayscale for insulator image processing.

Based on the Canny algorithm, this paper uses the gradient of three different binary images [40] to obtain the edge image of the insulator. The insulator images obtained by UAV aerial photography are prone to over-exposure points due to the influence of illumination. Gaussian filtering [41] matches the target pixels through the convolution template to reduce the impact of over-exposure points.

Here, to conduct Gaussian filtering on the image, a Gaussian filter of  $(2k + 1) \times (2k + 1)$  is generally used to convolve the image. The Gaussian function is shown in Equation (7)

$$H_{i,j} = \frac{1}{2\beta\sigma^2} \exp\left(-\frac{(i - (k + 1))^2 + (j - (k + 1))^2}{2\sigma^2}\right); 1 \leq i, j \leq (2k + 1) \quad (7)$$

Among them,  $k$  ( $k = 1, 2, 3$ ),  $(2k + 1)$ , and  $(i, j)$  represent an integer, the convolution kernel size, and the coordinates of one of the points, respectively.  $(i - (k + 1))^2 + (j - (k + 1))^2$  represents the effective distance between the neighborhood pixels and the center pixel. Usually, a smaller value of this value means that the center pixel of the image will get a higher weight distribution and still retains a higher amount of noise. If the value is too large, the phenomenon of excessive denoising will occur and then will lead to image distortion. This paper compares the image denoising effect of  $3 \times 3$ ,  $5 \times 5$ , and  $9 \times 9$  convolution filters and decides to choose the  $5 \times 5$  convolution filter for image denoising.

After Gaussian denoising, take the gradient magnitude as the segmentation index and use the gradient magnitude Otsu to extract the strong edges of the target object from the original image, that is, the general outline of the target object, and obtain the high threshold  $th$  for the interclass variance. Between  $(0, th)$ , we use the gradient magnitude Otsu to separate the target from the noise and the background. Use the gradient magnitude Otsu to separate the target from the noise and the background as much as possible so that the complete contour of the target can be displayed; most of the noise and background are suppressed and obtain the middle threshold  $tm$  for the interclass variance. Between  $(0, tm)$ , use the gradient magnitude to separate weak edges from noise, extract the complete contour of the target and as little noise and background as possible from the original image, and obtain low threshold  $tl$  for the interclass variance. Finally, the gradient information of three different binary images is extracted according to three different thresholds (TH) values, among which the binary images are  $g1(x, y)$ ,  $g2(x, y)$ , and  $g3(x, y)$ . The principle of converting grayscale images to binary images is as in Equation (8).

$$\begin{cases} 1, & \text{if } f(x, y) \gg TH \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

The purpose of converting a grayscale image to a binary image is to increase the contrast between the pixels in the image. The purpose of using three different thresholds is to include all gray level changes in the image and increase the contrast between weak edge regions, thus ensuring that all true edges in the insulator image are identified. The gradient reflects the intensity of pixel changes in the local area. The more severe the gradient changes, the higher the probability that it is an edge. A complete gradient must be calculated in the horizontal and vertical directions, and the calculation Equation is (9) and (10).

$$\frac{\delta g_i}{\delta x_i} \approx \frac{g_i(x_i + 1, y_i) - g_i(x_i - 1, y_i)}{2} \quad (9)$$

$$\frac{\delta g_i}{\delta y_i} \approx \frac{g_i(x_i, y_i + 1) - g_i(x_i, y_i - 1)}{2} \quad (10)$$

After obtaining the gradients in the x and y directions of the three binary images according to the gradients in the horizontal and vertical directions, the direction  $\theta_i$  and the increment  $\|\nabla g_i\|$  of the gradients are calculated as follows:

$$\theta_i = \tan^{-1} \left( \frac{\delta g_i}{\delta y_i} / \frac{\delta g_i}{\delta x_i} \right) \quad (11)$$

$$\|\nabla g_i\| = \sqrt{\left(\frac{\partial g_i}{\partial x_i}\right)^2 + \left(\frac{\partial g_i}{\partial y_i}\right)^2} \quad (12)$$

All the grayscale changes have been included in the three binary images, and the gradient images will produce edge images with thicker edges. Therefore, the Canny algorithm uses non-maximum suppression (NMS) [41] to perform the suppression operation on the region with a slight gradient change. That is the refinement operation on the edges. The non-maximum value suppression algorithm retains the part with the largest grayscale change in the eight-value neighborhood of the pixel; that is, it seeks the local maximum of pixels point and removes the pixels corresponding to the non-maximum point. In this way, most non-edge points can be eliminated and achieve the purpose of edge refinement.

The non-maximum suppression algorithm can only achieve the purpose of thinning the edge, but it cannot guarantee that the remaining pixels must be edge pixels. Therefore, we use a dual-threshold method to maximize the suppression of false edges to generate a more accurate edge image. The dual-threshold algorithm generates a more precise edge image by choosing two thresholds, and the pixels above the strong edge threshold are the actual edge pixels. Pixels below the weak edge threshold represent background information and are directly suppressed. The pixels between the strong edge threshold and the weak edge threshold are called weak edge pixels. If a strong-edge pixel exists within the eight-neighborhood of a weak-edge pixel, the pixel is also reserved as an edge pixel. Finally, this paper applies all the previous operations to three binary images, generates three different edge images, and then performs OR operations on these three different edge images to generate the final edge image. The expression is shown below.

$$g = g_1 | g_2 | g_3 \quad (13)$$

#### 4.2. Siamese ID-YOLO

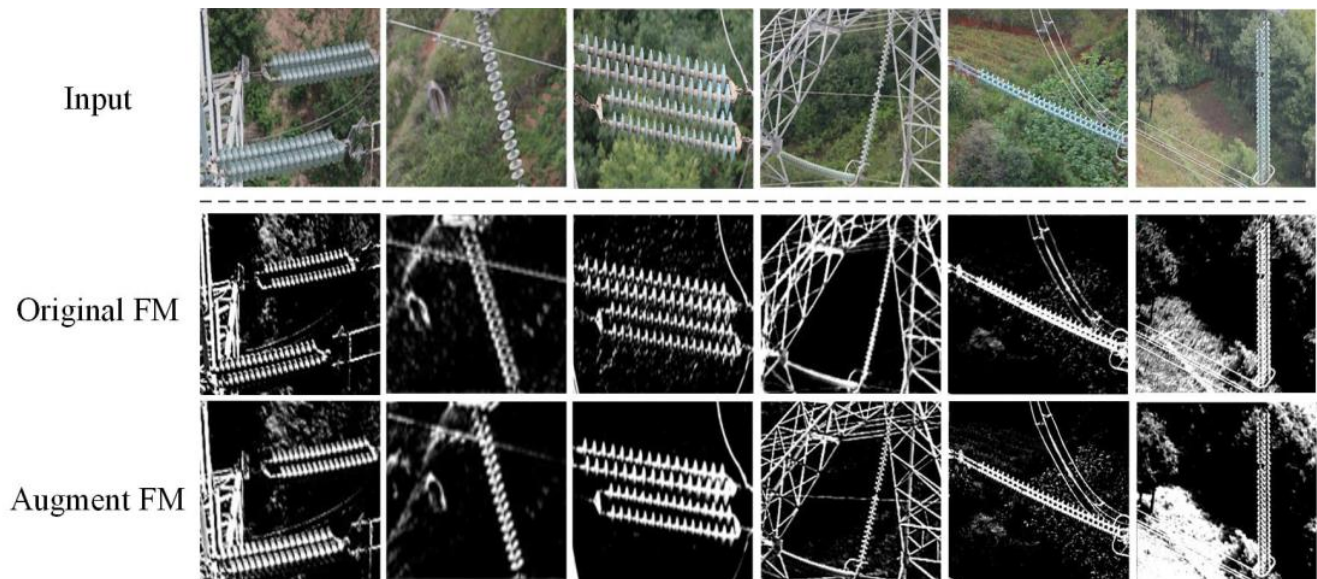
The insulator images gain richer semantic information after edge extraction. This paper expects to introduce the edge information of insulators into the model training to improve the performance of ID-YOLO in the insulator detection task. Therefore, the Siamese ID-YOLO network model is proposed based on the Siamese network.

The Siamese network was proposed by Bertinetto et al. in 2016 [42]. The Siamese network mainly solves the problem of object tracking and similarity judgment. All possible locations can be tested in detail to find the target's position in the new image. Select the candidate with the maximum similarity value to the previous target appearance to predict the target's position. The Siamese network uses the same transformation for both inputs, then combines the original image and the edge image, and is expressed by a similarity measure; similarity measure is accomplished through a cross-relevant layer:

$$g(z, x) = \varphi_\theta(z) * \varphi_\theta(x) + b \quad (14)$$

where  $\varphi_\theta(z)$  and  $\varphi_\theta(x)$  are the feature maps of the original image and edge image, respectively,  $b$  represents different bias values at each position.  $g(z, x)$  is not a single similarity score value but a score map, each value on the  $g(z, x)$  map corresponds to the similarity score of the original image and the edge image shifted sub-window, and the highest score is the sub-window where the target is located.

Based on the structure of the Siamese network, this paper performs feature extraction and feature fusion on the insulator edge image and the original image to obtain sufficient feature information and semantic information, thereby improving the feature extraction effect of the backbone part of ID-YOLO. As shown in Figure 3, Input, Original FM, and Augment FM represent the original insulator image, the feature map of the ID-YOLO output, and the feature map of the Siamese ID-YOLO output, respectively. Compared with the feature map obtained from the original ID-YOLO, the features of the insulators in the feature map output by Siamese ID-YOLO are more pronounced, which is beneficial to improving the effectiveness of insulator detection.



**Figure 3.** Input, Original FM, and Augment FM represent the original insulator image, the feature map of the ID-YOLO output, and the feature map of the Siamese ID-YOLO output, respectively. The features of the insulators in the feature map output by Siamese ID-YOLO are more pronounced, which is beneficial to improving the effectiveness of insulator detection.

#### 4.3. Hyperparameter Adjustment

Most detectors are currently pre-trained on ImageNet and COCO datasets to obtain prior knowledge of generic targets. Nonetheless, insulators do not fall into the scope of generic targets because it is a device that only appears in power facilities. Therefore, considering the shape characteristics of insulators, this paper adjusts some default hyperparameters before network training to make them more suitable for the insulator detection task.

The insulator has a lower proportion in the image and has a longer shape. As shown in Table 2, statistical analysis of the insulator calibration box's area found that the insulator's area size in the image is smaller than the initial set area of the anchor in YOLOv4. The maximum area of insulators does not exceed 600, and most are between 40 and 400. At the same time, we consider that the only type of object detected is the insulator and the effect of the number of hyperparameters on model training speed. This paper modifies the initial size hyperparameter of the anchor to [50, 80, 140, 200, 300] to match the area size of the insulator in the image.

**Table 2.** Comparison of the results of different anchor size settings.

Anchor Size	0–80	80–160	160–240	240–560
Proportion	16.66%	48.78%	24.94%	9.62%

According to the aspect ratio statistics, the shape of most insulators is long rectangular. As shown in Table 3, the aspect ratios of the insulators are mainly concentrated around 0.3 and 3. Compared with the initial setting of the aspect ratio [0.5, 1,2] of the anchor in ID-YOLO, the insulators' shapes are longer and more narrow. Therefore, this paper sets the aspect ratio hyperparameter of the anchor in ID-YOLO to [0.3, 1,3] to match the ratio of length and width of insulators in the image.

**Table 3.** Comparison of the results of different anchor ratio settings.

Anchor Ratio	0–0.3	0.3–1.0	1.0–5.0	5.0–12.5
Proportion	18.45%	32.16%	39.90%	9.49%

## 5. Experiments and Analysis

### 5.1. Experiment Description

#### 5.1.1. Dataset

Due to the limited number of insulator samples in the current publicly available insulator data sets CPLID [43], it cannot fully meet the application of the model in this paper in practical scenarios. Therefore, this article uses a 500 KV overhead transmission line in a particular place in Yunnan Province as the source of insulator data obtained. The insulator type is prioritized in line with toughened glass insulators, generally located at high altitudes and widely distributed in fields and forests. We filmed via UAV on a line insulator to improve sampling efficiency and then verify the feasibility and effectiveness of the model in this paper. Additionally, we use LabelMe to label insulator images. What is worth mentioning here is that in the process of labeling insulators, we inevitably mark some background information, which makes insulator identification extremely vulnerable to the interference of background information during the training process. To better separate the insulators from the background information, we use magenta to mark insulators to confirm the effectiveness of our proposed model. We name this dataset InsuDaSet, which contains three thousand image datasets containing insulators. The training set (TR-DB) and the test set (T-DB) include 2500 images and 500 images, respectively.

#### 5.1.2. Experiment Configuration

This paper's hardware configuration used for the experiments is AMD R5-4400G CPU, 32GB RAM, and NVIDIA RTX3060 GPU. The experimental environment is Ubuntu 18.04 64-bit Linux system, the software environment is Pytorch 1.4, and the test programs for the experiments were all written in Python-based language. Set the initial learning rate to 0.001 and trained all experimental results with 20 iterations of the best batch model parameters.

### 5.2. The Baselines

In the following experiments, we select several state-of-the-art deep learning networks on the insulator dataset to compare speed and precision.

**YOLOv4:** The YOLOv4 model is continuously improved based on YOLOv1, YOLOv2, and YOLOv3, and mainly uses the spatial pyramid pooling (SPP) and PANet to enhance the features, so that to obtain better detection accuracy and faster real-time operation.

**Faster R-CNN [44]:** As a two-stage target detection algorithm, the most significant contribution of this algorithm is that the RPN network is proposed first, which significantly saves the generation time of the proposal box. Secondly, RPN can be added to the end-to-end training process of the network by sharing the network weights with the object detection sub-network. Its detection accuracy is the best, but it also severely slows down the detection speed.

**SSD [45]:** The SSD algorithm adopts anchor box processing. With each pixel point of the feature map as the center point, generate a series of default boxes of different sizes. Finally, feature layers of different scales are used to detect objects of various sizes. The SSD algorithm learns from the idea of the YOLOv4 network to transform the target detection

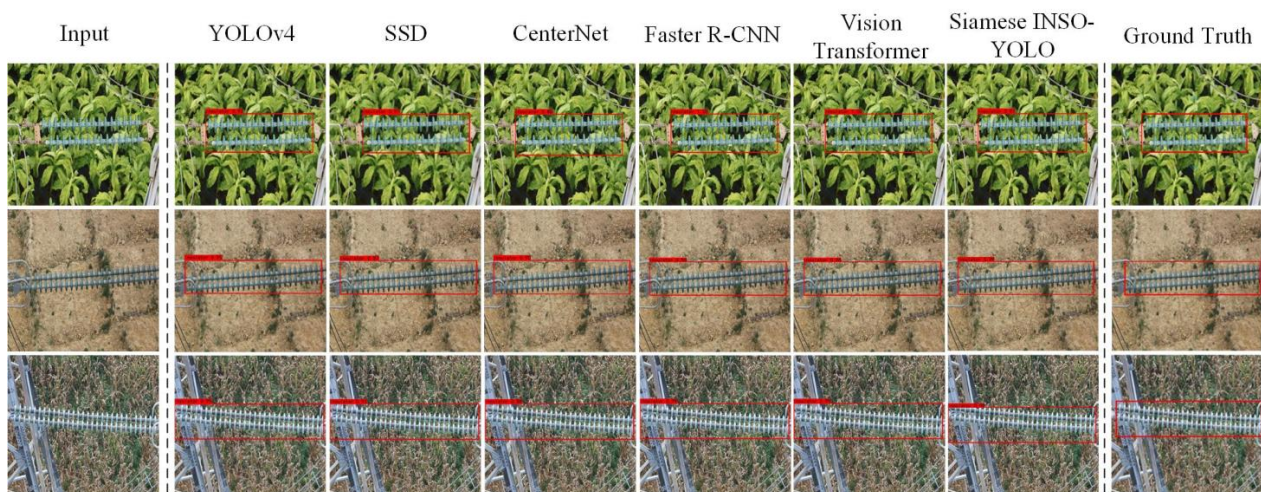
task from a classification problem into a regression problem without the extraction process of the region candidate box, which significantly shortens the detection time. At the same time, the SSD algorithm draws on the anchor point mechanism of the Faster R-CNN network. Namely, a certain number of aspect ratio determined default boxes are generated on each pixel of the feature map and make predictions on feature maps at different sizes. The purpose is to improve the detection accuracy of objects.

CenterNet [46]: CenterNet believes that if a critical point represents each object, the object detection problem is a key point detection problem. Moreover, since each target has only one point in the testing process, there will be no overlapping of multiple prediction boxes, and there is no need to use the NMS algorithm to deduplicate. Therefore, this method has the advantages of fast testing speed and small space occupation.

Vision Transformer: Compared with Faster R-CNN, the most significant feature of Vision Transformer is that it converts the target detection problem into an unordered collection prediction problem. Vision Transformer first extracts features from the image through a CNN backbone network. It then uses a transformer's encoder. The decoder structure directly obtains unordered predictions collection with a specified length of  $N$ . Each element in the collection contains the predicted parent's classification category and the prediction box coordinates.

### 5.3. Qualitative Evaluation

The image resolution of the input model is  $608 \times 608$ . Figure 4 shows our qualitative evaluation results. The results show that the test results of the model in this paper are more accurate in category and location. When using the InsuDaSet dataset to evaluate and test the current mainstream models, the detection effect of SSD is not as precise as YOLOv4. The test results of the model proposed in this paper are similar to Faster R-CNN, Vision Transformer in categories and locations. Still, the network parameters are less computationally intensive and consume fewer resources. The test results of CenterNet are general, and we believe this is due to the complex background of insulators and the small proportion of insulator targets in the image.



**Figure 4.** Test results of each model on the InsuDaSet dataset. The experimental results show that the test accuracy of Siamese ID-YOLO is similar to that of Faster R-CNN, Vision Transformer, and it is better than the single-stage model YOLOv4, SSD, and CenterNet based on the heatmap.

### 5.4. Quantitative Evaluation

Here, we evaluate the generation position of the detection box according to the average accuracy (AP) in the COCO dataset [25]. For object detection boxes obtained during testing and actual manually labeled ground truth IoU as correct criteria. Correct test results, incorrect results, and undetected ground truth are marked as true positives (TP), false

positives (FP), and false negatives (FN), respectively. IoU, accuracy, and recall are calculated as follows.

$$IoU = \frac{DefaultBox \cap GroundTruthBox}{DefaultBox \cup GroundTruthBox} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

The recall and accuracy of detection results under different IoU thresholds are calculated according to the confidence ranking of the predicted target frames, which can obtain the most commonly used COCO evaluation criteria, such as AP, AP<sub>IoU = 0.50</sub>, and AP<sub>IoU = 0.75</sub>. Because only one detection class is an insulator, AP is mAP, and IoU above the average value allows detectors to obtain better localization capability.

The results of different networks and improved methods are shown in Table 4. It can be seen from the results that the accuracy of the test results of Siamese ID-YOLO is not much different from that of Faster R-CNN, Vision Transformer, but it is significantly higher than that of single-stage SSD, YOLOv4. Nevertheless, CenterNet does not perform well on the insulator detection task, which we believe is caused by the complex background of insulator images and the low proportion of insulator images in the image. Compared with our base model ID-YOLO, the detection box of our model is more accurate, which indicates that edge extraction and the Siamese network play an important role. Experimental results show that Siamese ID-YOLO can effectively localize insulators in images.

**Table 4.** Comparison accuracy and speed table of different models.

Model	InsudaSet			CPLID		
	AP	AP <sub>50</sub>	AP <sub>75</sub>	AP	AP <sub>50</sub>	AP <sub>75</sub>
YOLOv4	60.41	88.25	80.69	60.76	88.49	79.98
SSD	58.36	86.87	78.26	59.64	87.63	78.83
Faster R-CNN	68.53	90.26	87.52	69.79	89.72	86.15
CenterNet	55.16	84.34	77.58	56.45	83.81	75.66
Vision Transformer	68.72	90.85	87.39	68.25	89.57	86.03
ID-YOLO	62.42	86.57	78.94	61.97	85.12	76.37
Siamese ID-YOLO	70.16	92.72	88.92	69.38	90.89	88.05

Also, in terms of test speed, FPS, Faster R-CNN takes the longest time, followed by Vision Transformer. Siamese ID-YOLO is slightly higher than single-stage YOLOv4 and SSD, and CenterNet has the fastest detection speed, followed by ID-YOLO. In general, the model in this paper improves the calculation speed of the model under the premise that the detection accuracy is as high as possible. The test speed of about 84 FPS is sufficient to support the real-time request, which can match the hardware environment of UAVs and other equipment. It can be applied in the natural environment of insulator detection.

### 5.5. Sensitivity Analysis

In this section, the sensitivity analysis of the things we improved is carried out to prove that the improvement of each part of our model is necessary, including backbone network selection, image preprocessing, Siamese network introduction, anchor box hyperparameter presetting, and the number of training iterations.

#### 5.5.1. The choice of the Backbone

On the premise of retaining other improvements of Siamese ID-YOLO, we conduct a sensitivity analysis on its backbone network. As shown in Table 5, the backbone network VGG16 has many parameters and is computationally complex. The backbone network based on the Darknet53 network used in this paper draws on the idea of a ResNet shortcut connection, and it performs better than ResNet101. However, we consider that we need

to detect only one type of insulator, and most of the insulators occupy a small proportion of the image. We believe that reducing the number of network layers and improving the feature extraction ability of the backbone network will help to improve these problems. The backbone network based on Darknet53 has the effect of streamlining the number of network layers; in addition, by introducing the twin network, it also achieves the purpose of improving the feature extraction capability of the backbone network. We finally decided to use the Darknet53 network and the Siamese network cooperatively use's backbone network as our feature extraction network.

**Table 5.** Comparison accuracy and speed table of different backbones.

Backbone	InsudaSet			FPS
	AP	AP <sub>50</sub>	AP <sub>75</sub>	
VGG16	60.41	84.14	79.17	56
ResNet101	64.73	90.82	87.69	70
Darknet53	68.82	91.27	87.63	73
ID-YOLO	62.42	86.57	78.94	96
Siamese ID-YOLO	70.16	92.72	88.92	84

### 5.5.2. Edge Extraction

On the premise of retaining other improvements of Siamese ID-YOLO, we conduct a sensitivity analysis on its edge extraction. Table 6 shows that the Sobel operator is a particularly unsatisfactory effect for images with complex noise. The Roberts operator positioning is relatively accurate, but because it does not include smoothing, it is susceptible to noise, which leads to its edge extraction effect is not very good. Since the Canny operator can detect the thin edge part of the image, we use the Canny-based edge detection operator to extract the edge of the insulator image.

**Table 6.** Comparison accuracy and speed table of different edge extraction methods.

Edge Extraction	InsudaSet			FPS
	AP	AP <sub>50</sub>	AP <sub>75</sub>	
Sobel	66.76	88.23	81.84	83.5
Roberts	67.94	89.05	82.63	83.2
Canny	68.28	90.35	85.64	83.8
Improved Canny	70.16	92.72	88.92	84

### 5.5.3. Hyperparameters Selection

In this part, to make the network more suitable for the insulator detection task, we consider changing the hyperparameters of the network. As seen in Table 7, according to the statistical results of ground-truth, we use the five values with the larger size in the insulator image [50, 80, 140, 200, 300] as the setting of the anchor size and find the effect of this setting significantly. After determining the anchor size settings, the initial value of the aspect ratio was next considered to be changed. We considered using the settings [0.3, 1.0, 3.0, 4.0] and the effect did improve. Finally, we determine the initial value of anchor size [50, 80, 140, 200, 300] and aspect ratio [0.3, 1.0, 3.0, 4.0], which can improve detection accuracy without adding much test time. These settings prove that it is necessary to initialize the anchor box size and aspect ratio.

### 5.5.4. Number of Epochs

The hyperparameter selection of the number of training iterations (epoch) is also significant. If the number of iterations is too low, the model cannot fit the entire sample distribution space. Overfitting is likely to occur if the number of iterations is too high. As shown in Table 8, as the number of experimental iterations increases, the model's performance on the test set will improve. However, after the number of iterations exceeds

200, the performance on the test set will decrease slightly; This is because the model generated an overfitting phenomenon on the training set. In summary, the number of iterations selected in this paper is 200.

**Table 7.** Comparison table of results for setting different anchor box sizes and aspect ratios.

Size	Scale	InsudaSet			FPS
		AP	AP <sub>50</sub>	AP <sub>75</sub>	
[32, 64, 128, 256, 512]	[0.5, 1.0, 2.0]	65.23	90.45	85.96	83.5
[50, 80, 140, 200, 300]	[0.5, 1.0, 2.0]	67.35	91.01	87.36	83.2
[32, 64, 128, 256, 512]	[0.3, 1.0, 3.0, 4.0]	68.69	92.25	87.66	83.8
[50, 80, 140, 200, 300]	[0.3, 1.0, 3.0,4.0]	70.16	92.72	88.92	84

**Table 8.** Comparison table of results for setting different epoch numbers.

Epoch Numbers	InsudaSet			FPS
	AP	AP <sub>50</sub>	AP <sub>75</sub>	
50	42.27	71.64	58.82	88
100	55.14	80.58	70.38	86
150	63.49	85.24	79.06	84
200	70.16	92.72	88.92	84
250	69.73	90.49	85.89	83

### 5.6. Ablation Analysis

To analyze the degree of influence of different components on the model, we performed an ablation analysis on the model proposed in this paper on the InsuDataset dataset. As shown in Table 9, the detection accuracy of model B is more evident than that of model A, and the improvement is the most, which indicates that the backbone based on Darknet53 plays a significant role with the assistance of the Siamese network. It allows the model to fuse more semantic information of edge images and original images to improve detection accuracy. Furthermore, the B model has lower computational complexity than other methods. Therefore, our insulator detection task must introduce a Siamese network based on the Darknet53 network. After model C performs an image enhancement operation on the image, the image quality is improved, and it has a certain positive effect on the improvement of the model detection accuracy. Module D made cluster analysis on the size and aspect ratio of the actual insulator data set and then preset adjustment of the size and proportion of the anchor box. Hyperparameter adjustment made the insulator data set more suitable for this paper's model and further improved insulator detection accuracy.

**Table 9.** The results of the ablation analysis.

Model	Architecture	AP	AP <sub>50</sub>	AP <sub>75</sub>
A	ID-YOLO	62.42	86.57	78.94
B	A + backbone	66.37	89.39	84.87
C	B + image augment	68.92	91.42	86.63
D	C+ hyperparameter adjustment	70.16	92.72	88.92

### 5.7. Computational Complexity

This section provides a comparative analysis of the models' computational overhead and training time. As shown in Table 10, there is little difference in the detection results of Siamese ID-YOLO compared to Faster R-CNN, Vision Transformer. However, the computational cost and training time of Siamese ID-YOLO are significantly reduced. The minor change in space complexity and time complexity of Siamese ID-YOLO compared to YOLOv4 and our benchmark model ID-YOLO is since we only modified the backbone network of the model. ID-YOLO designed a feature extraction network based on Darknet53



based on YOLOv4, and Siamese ID-YOLO introduced the Siamese network based on ID-YOLO, but the detection effect is much improved. In addition, CenterNet has the least computational overhead. SSD has a slightly higher computational overhead than Siamese ID-YOLO, but its detection results are unsatisfactory.

**Table 10.** Comparison table of computation cost and training time of different models.

Model	Param	Training Time (h)
YOLOv4	29 M	6.45
SSD	35 M	7.58
Faster R-CNN	181 M	48.83
CenterNet	15 M	4.08
Vision Transformers	156 M	40.57
ID-YOLO	21 M	5.23
Siamese ID-YOLO	32 M	6.96

In the sensitivity analysis experiment, we made sensitivity analysis on the feature extraction network, edge detection algorithm, and anchor box hyperparameter preset. Experiments show that the introduction of Siamese networks in the Darknet53-based network, the use of the Canny-based edge detection algorithm, and the preset adjustment of anchor box hyperparameters all have varying degrees of improvement in the detection effect of insulators. In the ablation analysis experiment, this paper conducts ablation analysis on different components, such as introducing the Siamese network, enhancing the data, and adjusting the hyperparameters of the anchor box. Experiments show that these components all improve the accuracy and speed of insulator detection to varying degrees. Among them, the introduction of the Siamese network improves accuracy the most. To sum up, the model Siamese ID-YOLO in this paper has an excellent insulator detection performance, achieving a detection accuracy of 92.72% and a detection speed of 84FPS. Siamese ID-YOLO model can match the hardware environment of UAVs and other equipment to detect insulators online.

## 6. Conclusions

This paper proposes an insulator detection model Siamese ID-YOLO, which has stronger robustness than the traditional detection methods. In addition, this paper makes an insulator dataset named InsuDaSet based on UAV aerial images. In this paper, the original image is processed by mean grayscale to obtain a grayscale image against when the insulator target and background differences are slight. The edge detection algorithm based on the Canny operator is used to get the insulator's edge image to enrich the image's semantic information. To integrate insulator edge information as gain information into model training, this paper introduces the Siamese network based on the Darknet53 network to fuse the semantic information and position information of edge image and original image to improve the detection accuracy of the insulator. Finally, to make the insulator data set more suitable for the model in this paper, by clustering statistics on the aspect ratio and size of the insulator data set, the hyperparameters of the model anchor box are pre-analyzed and adjusted. The proposed model can accurately output insulator recognition images compared to mainstream models. Furthermore, we only detect glass insulators on high voltage transmission lines, not fault detection. However, we believe that our model is equally applicable to the detection of ceramic insulators, compliant insulators, and other electrical components. Therefore, we will consider the classification detection of insulator faults in the next stage and continue exploring this model's application to other electrical components.

**Author Contributions:** Conceptualization, J.W. and Y.L.; methodology, J.W., Y.L. and W.C.; validation, J.W.; formal analysis, J.W. and Y.L.; investigation, J.W., Y.L. and W.C.; resources, J.W., Y.L. and W.C.; writing—original draft preparation, J.W.; writing—review and editing, J.W., Y.L. and W.C.; visualization, J.W.; supervision, Y.L.; project administration, Y.L.; All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the National Natural Science Foundation of China, grant number 61962031; the National Natural Science Foundation of China, grant number 51667011; and the Applied Basic Research Project of Yunnan province, grant number 2018FB095.

**Data Availability Statement:** The data in this paper are undisclosed due to the confidentiality requirements of the data supplier.

**Acknowledgments:** We thank the Yunnan Electric Power Research Institute for collecting the transmission line UAV inspection data, which provided a solid foundation for the verification of the model proposed in this paper. At the same time, we thank the reviewers and editors for their constructive comments to improve the quality of this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Zheng, H.; Liu, Y.; Sun, Y.; Li, J.; Shi, Z.; Zhang, C.; Lai, C.S.; Lai, L.L. Arbitrary-Oriented Detection of Insulators in Thermal Imagery via Rotation Region Network. *IEEE Trans. Industr. Inform.* **2022**, *18*, 5242–5252. [\[CrossRef\]](#)
- Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Qu, W. Detection of Transmission Line Insulator Defects Based on an Improved Lightweight YOLOv4 Model. *Appl. Sci.* **2022**, *12*, 1207. [\[CrossRef\]](#)
- Wang, Z.; Liu, X.; Peng, H.; Zheng, L.; Gao, J.; Bao, Y. Railway Insulator Detection Based on Adaptive Cascaded Convolutional Neural Network. *IEEE Access* **2021**, *9*, 115676–115686. [\[CrossRef\]](#)
- Chen, W.; Li, Y.; Zhao, Z. Missing-Sheds Granularity Estimation of Glass Insulators Using Deep Neural Networks Based on Optical Imaging. *Sensors* **2022**, *22*, 1737. [\[CrossRef\]](#)
- Li, X.; Li, Z.; Wang, H.; Li, W. Unmanned Aerial Vehicle for Transmission Line Inspection: Status, Standardization, and Perspectives. *Front. Energy Res.* **2021**, *9*, 21–34. [\[CrossRef\]](#)
- Giones, F.; Brem, A. From toys to tools: The co-evolution of technological and entrepreneurial developments in the drone industry. *Bus. Horiz.* **2017**, *60*, 875–884. [\[CrossRef\]](#)
- Zhang, Y.; Yuan, X.; Li, W.; Chen, S. Automatic Power Line Inspection Using UAV Images. *Remote Sens.* **2017**, *9*, 824. [\[CrossRef\]](#)
- Lyu, X.; Li, X.; Dang, D.; Dou, H.; Wang, K.; Lou, A. Unmanned Aerial Vehicle (UAV) Remote Sensing in Grassland Ecosystem Monitoring: A Systematic Review. *Remote Sens.* **2022**, *14*, 1096. [\[CrossRef\]](#)
- Honkavaara, E.; Saari, H.; Kaivosoja, J.; Pölonen, I.; Hakala, T.; Litkey, P.; Mäkynen, J.; Pesonen, L. Processing and Assessment of Spectrometric, Stereoscopic Imagery Collected Using a Lightweight UAV Spectral Camera for Precision Agriculture. *Remote Sens.* **2013**, *5*, 5006–5039. [\[CrossRef\]](#)
- Qin, R. An Object-Based Hierarchical Method for Change Detection Using Unmanned Aerial Vehicle Images. *Remote Sens.* **2014**, *6*, 7911–7932. [\[CrossRef\]](#)
- Reis, B.P.; Martins, S.V.; Fernandes-Filho, E.I.; Sarcinelli, T.S.; Gleriani, J.M.; Leite, H.G.; Halassy, M. Forest restoration monitoring through digital processing of high resolution images. *Ecol. Eng.* **2019**, *127*, 178–186. [\[CrossRef\]](#)
- Xiao, Z.; Zhu, L.; Liu, Y.; Yi, P.; Zhang, R.; Xia, X.-G.; Schober, R. A Survey on Millimeter-Wave Beamforming Enabled UAV Communications and Networking. *IEEE Commun. Surv. Tutor.* **2021**, *24*, 557–610. [\[CrossRef\]](#)
- Siddiqi, M.A.; Iwendi, C.; Jaroslava, K.; Anumbe, N. Analysis on security-related concerns of unmanned aerial vehicle: Attacks, limitations, and recommendations. *Math. Biosci. Eng.* **2022**, *19*, 2641–2670. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhao, Z.; Xu, G.; Qi, Y. Representation of binary feature pooling for detection of insulator strings in infrared images. *IEEE Trans. Dielectr. Electr. Insul.* **2016**, *23*, 2858–2866. [\[CrossRef\]](#)
- Gong, X.; Yao, Q.; Wang, M.; Lin, Y. A Deep Learning Approach for Oriented Electrical Equipment Detection in Thermal Images. *IEEE Access* **2018**, *6*, 41590–41597. [\[CrossRef\]](#)
- Wang, S.; Jiang, T.; Li, W.; Niu, L.; Lv, F. Infrared and UV Imaging Characteristics of Insulator Strings Containing a Zero-Value Insulator Based on Experiment and Multi-Physics Simulation. *IEEJ Trans. Electr. Electron. Eng.* **2021**, *16*, 374–382. [\[CrossRef\]](#)
- Zhai, Y.; Wang, D.; Zhang, M.; Wang, J.; Guo, F. Fault detection of insulator based on saliency and adaptive morphology. *Multimed. Tools Appl.* **2016**, *76*, 12051–12064. [\[CrossRef\]](#)
- Yin, J.; Lu, Y.; Gong, Z.; Jiang, Y.; Yao, J. Edge Detection of High-Voltage Porcelain Insulators in Infrared Image Using Dual Parity Morphological Gradients. *IEEE Access* **2019**, *7*, 32728–32734. [\[CrossRef\]](#)
- Xia, H.; Yang, B.; Li, Y.; Wang, B. An Improved CenterNet Model for Insulator Defect Detection Using Aerial Imagery. *Sensors* **2022**, *22*, 2850. [\[CrossRef\]](#)
- Baik, H.; Valenzuela, J. Unmanned Aircraft System Path Planning for Visually Inspecting Electric Transmission Towers. *J. Intell. Robot. Syst.* **2018**, *95*, 1097–1111. [\[CrossRef\]](#)

21. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-Based Oblique Photogrammetry for Outdoor Data Acquisition and Offsite Visual Inspection of Transmission Line. *Remote Sens.* **2017**, *9*, 278. [[CrossRef](#)]
22. Liu, C.; Wu, Y.; Liu, J.; Sun, Z. Improved YOLOv3 Network for Insulator Detection in Aerial Images with Diverse Background Interference. *Electronics* **2021**, *10*, 771. [[CrossRef](#)]
23. Ge, B.; Hou, C.; Liu, Y.; Wang, Z.; Wu, R. Anomaly detection of power line insulator from aerial imagery with attribute self-supervised learning. *Int. J. Remote Sens.* **2021**, *42*, 8819–8839. [[CrossRef](#)]
24. Lei, X.; Sui, Z. Intelligent fault detection of high voltage line based on the Faster R-CNN. *Measurement* **2019**, *138*, 379–385. [[CrossRef](#)]
25. Liao, S.; An, J. A Robust Insulator Detection Algorithm Based on Local Features and Spatial Orders for Aerial Images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 963–967. [[CrossRef](#)]
26. Zhai, Y.; Chen, R.; Yang, Q.; Li, X.; Zhao, Z. Insulator Fault Detection Based on Spatial Morphological Features of Aerial Images. *IEEE Access* **2018**, *6*, 35316–35326. [[CrossRef](#)]
27. Yu, Y.; Cao, H.; Wang, Z.; Li, Y.; Li, K.; Xie, S. Texture-and-Shape Based Active Contour Model for Insulator Segmentation. *IEEE Access* **2019**, *7*, 78706–78714. [[CrossRef](#)]
28. Zhao, Z.; Liu, N.; Wang, L. Localization of multiple insulators by orientation angle detection and binary shape prior knowledge. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 3421–3428. [[CrossRef](#)]
29. Li, T.; Hao, T. Damage Detection of Insulators in Catenary Based on Deep Learning and Zernike Moment Algorithms. *Appl. Sci.* **2022**, *12*, 5004. [[CrossRef](#)]
30. Ma, B.; Fu, Y.; Wang, C.; Li, J.; Wang, Y. A high-performance insulators location scheme based on YOLOv4 deep learning network with GDIoU loss function. *IET Image Process.* **2022**, *16*, 1124–1134. [[CrossRef](#)]
31. Gao, Z.; Yang, G.; Li, E.; Liang, Z. Novel Feature Fusion Module-Based Detector for Small Insulator Defect Detection. *IEEE Sens. J.* **2021**, *21*, 16807–16814. [[CrossRef](#)]
32. Zhang, X.; Zhang, Y.; Liu, J.; Zhang, C.; Xue, X.; Zhang, H.; Zhang, W. InsuDet: A Fault Detection Method for Insulators of Overhead Transmission Lines Using Convolutional Neural Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [[CrossRef](#)]
33. Chen, J.; Xu, X.; Dang, H. Fault Detection of Insulators Using Second-order Fully Convolutional Network Model. *Math. Probl. Eng.* **2019**, *2019*, 1–10. [[CrossRef](#)]
34. Han, J.; Yang, Z.; Xu, H.; Hu, G.; Zhang, C.; Li, H.; Lai, S.; Zeng, H. Search Like an Eagle: A Cascaded Model for Insulator Missing Faults Detection in Aerial Images. *Energies* **2020**, *13*, 713. [[CrossRef](#)]
35. Liu, W.; Liu, Z.; Wang, H.; Han, Z. An Automated Defect Detection Approach for Catenary Rod-Insulator Textured Surfaces Using Unsupervised Learning. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 8411–8423. [[CrossRef](#)]
36. Redmon, J.; Farhadi, A. YOLOv4v3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
37. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 28 December 2018; pp. 936–944. [[CrossRef](#)]
38. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; Institute of Electrical and Electronics Engineers (IEEE): New York, NY, USA, 2018; pp. 8759–8768. [[CrossRef](#)]
39. Bochkovskiy, A.; Wang, C.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934. Available online: <https://arxiv.org/abs/2004.10934v1> (accessed on 10 January 2022).
40. Al-Amaren, A.; Ahmad, M.O.; Swamy, M.N.S. Edge Map Extraction of an Image Based on the Gradient of its Binary Versions. In Proceedings of the 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), Lansing, MI, USA, 9–11 August 2021.
41. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1986**, *8*, 679–698. [[CrossRef](#)]
42. Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H.S. Fully-Convolutional Siamese Networks for Object Tracking. In *Computer Vision—ECCV 2016 Workshops. ECCV 2016; Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2016.
43. Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of Power Line Insulator Defects Using Aerial Images Analyzed With Convolutional Neural Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 1486–1498. [[CrossRef](#)]
44. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
45. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9905, pp. 21–37. [[CrossRef](#)]
46. Zhou, X.; Wang, D.; Krähenbühl, P. Objects as Points. *arXiv* **2019**, arXiv:1904.07850. Available online: <https://arxiv.org/abs/1904.07850v2> (accessed on 10 January 2022).