Accepted manuscript (author version)

To appear in: **Majlesi Journal of Electrical Engineering (MJEE)**

Online ISSN: 2345-377X **Print ISSN:** 2345-3796

Received: 15 Mar 2025

Revised: 14 June 2025

Accepted: 07 July 2025

DOI: 10.57647/j.mjee.2025.17412

Original Research

An Improved Deep Mask Region-based Convolutional Neural Network for the Detection of Power Transmission Insulator Defects

Haider Abdulzahra Saad **Alsaide**[1], Mohammadreza **Soltanaghaei**[2,*], Wael Hussein Zayer **Al-Lami**[3], Razieh **Asgarnezhad**[4]

[1]Department of Computer Engineering, Institute of Artificial Intelligence and Social and Advanced Technologies, Isf.C., Islamic Azad University, Isfahan, Iran.
headerabd917@gmail.com

[2]Department of Computer Engineering, Institute of Artificial Intelligence and Social and Advanced Technologies, Isf.C., Islamic Azad University, Isfahan, Iran.
soltan@iau.ac.ir (Correspond Author) 0000-0002-2930-9066

[3]Electronic Department, Amara Technical Institute, Southern Technical University, Missan, Iraq.
wael.zayer@stu.edu.iq

[4]Department of Computer Engineering, Institute of Artificial Intelligence and Social and Advanced Technologies, Isf.C., Islamic Azad University, Isfahan, Iran.
r.asgarnezhad@khuisf.ac.ir

Abstract

In the power industry, accurate and real-time defect detection in power transmission lines is crucial to reducing accidents and maintenance costs. Although various image processing and AI-based methods have been proposed using unmanned arial vehicles (UAV) images, their efficiency significantly drops under complex environmental conditions, such as high color or texture diversity in backgrounds. To address this challenge, this paper presents an improved version of the mask region-based convolutional neural network (R-CNN) network by redefining the architecture of the head and classification sections to enhance detection accuracy in complex scenarios. The proposed model introduces a novel rhombus-structured fully

connected layer in the classification branch for better feature encoding and decoding, alongside a drop-out layer in the bounding box detection branch to prevent overfitting. Additionally, ResNet-101 is employed as the backbone, and transfer learning is used to optimize training and reduce computational complexity. Experimental evaluations demonstrate that the improved Mask R-CNN achieves a mean average precision (mAP) of 92.07% and an accuracy of 92.50% in detecting power transmission line defects, outperforming existing methods. Furthermore, the proposed approach helps lower the cost and time required for fault detection and maintenance, making it a practical solution for real-world power transmission system inspections.

Keywords: Power transmission line, Defect detection, Deep learning, Feature extraction, Mask region-based convolutional neural network, Transfer learning

Introduction

In many developing nations, ensuring the uninterrupted operation of transmission lines is crucial for sustaining economic growth, fostering trade relationships, and supporting the daily lives of residents. A key aspect of transmission line maintenance is the accurate and real-time detection of defects [1]. Insulators, which are vital components of power transmission lines, play a critical role in supporting and isolating electrical conductors, preventing the current from passing through them. These insulators must endure vertical and horizontal loads, as well as conductor tension, while maintaining sufficient resistance to both operating voltage and overvoltage. Given that insulators are often deployed in varied natural and geographical settings for extended periods, they are prone to various failures or defects, such as erosion from foreign objects, shattering, string breakage, cracking, and self-explosion [2]. Any of these defects can result in the complete shutdown of a transmission line, jeopardizing the stability and continuous operation of the entire power network. Consequently, the precise and swift identification of insulator defects is essential to prevent power outages, minimize network losses, and reduce operational costs. Examples of such defects are depicted in Fig. 1.

Fig.1 Some examples of insulator defects in power transmission line

Most methods of Transmission line inspection involve manual, human-supervised helicopter, robotic, and unmanned aerial vehicle (UAV) inspection [3]. In traditional methods such as manual inspection, skilled electrical workers use visual observation and telescope-assisted observation along the entire length of the transmission line to decide whether the transmission line equipment is defect-free or defective, including anti-vibration hammers, insulators, and so on. These methods, which fully rely on human ability and knowledge, require considerable energy and physical resilience in addition to professional knowledge. This category of inspection methods requires higher technical and security requirements. These are more costly and risky, particularly for the inspection of transmission lines in mountainous regions and over large rivers. Therefore, the manual method is not effortless under certain geographical conditions and weather situations, and the efficiency of the inspection is not feasible. In non-contact inspection methods, imaging at different frequencies such as infrared, ultraviolet waves, ultrasonic, laser, infrared thermal imaging, and visible frequency is used. However, data analysis, fault diagnosis, and final status decision-making still depend on participating power engineers, which limits the effectiveness of power equipment quality diagnosis. Hence, the study has found more rapid and precise methods for automated fault detection in power transmission equipment, and insulators have become a research topic. Analysis of insulation images acquired by aerial imaging is a critical basis for making decisions about the performance status of transmission lines. Helicopter-based imaging usually has a high financial cost and many safety risks. Therefore, UAVs have become one of the most effective tools in the inspection of transmission lines due to their lower cost and greater flexibility [4]. In the last decade, methods for fault detection in power transmission line insulation based on machine learning and computer vision techniques have been presented. Most of the methods presented in this field can be divided into two groups: methods based on classical image processing techniques and methods based on deep learning (DL). Methods that use image processing techniques and classical machine learning methods such as supervised classifiers usually have less computational complexity, but their detection accuracy is much lower than DL-based methods. For example, Ni et al. in [5] developed an insulator Defect Detection (DD) approach based on handcrafted image features and watershed algorithms. In [5], the region of interest (ROI) is detected based on a variety of texture and formation characteristics. Also, according to the different properties of various defect types, an improved watershed algorithm is applied to achieve the segmentation of the insulator image and DD. In recent years, with the increase in computational power, the speed of methods based on deep neural networks has become acceptable. Xia et al. [6] designed an improved version of Deep CenterNet for insulator DD in transmission lines. Their proposed enhanced Deep CenterNet uses MobileNetV1 for feature extraction and depthwise separable convolution as the center unit, including depthwise convolution

and point-wise convolution. This center unit decreases the number of model parameters and runtime [6].

Aerial images taken by UAVs of transmission lines usually have complex and natural backgrounds. Therefore, extracting general features from these images is not a very logical approach and reduces the final accuracy of the system. Some methods use windowing techniques for local detection of defects, but this group of methods is highly dependent on the dimensions of the window. On the other hand, the contents of many windows are completely free of transmission lines and insulators, so the proposed deep network needs to learn natural environments in addition to defect-free insulators to be able to detect defects. These reasons make the accuracy of systems based on windowing techniques relatively low. In this article, a two-phase method is presented to detect insulation defects in power transmission lines. In the first phase, transmission lines are separated from the background. In the second phase, the insulator image is analyzed and the defective parts in the image are segmented. Mask R-CNN is a Convolutional Neural Network (CNN) that is known as an instance-based image segmentation network [7]. It conducts the segmentation at the pixel level on detected objects. The Mask R-CNN algorithm can handle multiple classes and overlapping objects. Mask R-CNN consists of a backbone network and a region-proposal network for feature extraction, learning desired instances, and ROI alignment. Next, a sequence of convolution, pooling, and associated layers is used for segmentation. While Faster R-CNN, as the base of Mask R-CNN, has two outputs per instance—the class label and bounding box offset—Mask R-CNN adds a third branch that outputs the object mask [8]. This supplemental output is distinguishable from the class and box outputs and enables the extraction of an additional refined spatial structure of an object.

In the current article, an enhanced and innovative Mask R-CNN model is proposed, based on a novel internal architecture for DD. Unlike the classical Mask R-CNN, the proposed improved version employs a rhombus-shaped structure to design the series of fully connected layers involved in the label classification stage of the network. In addition, to reduce over-fitting, a dropout layer is incorporated into the fully connected layers used in the mask prediction branch of the Mask R-CNN head. To reduce offline computational complexity, transfer learning is employed by fine-tuning a pre-trained ResNet-101, which serves as the backbone of the proposed architecture.

The effectiveness of the proposed method is evaluated on a well-known benchmark dataset in this field, the Insulator Defect Identification Dataset (IDID), using standard metrics including mean average precision (mAP), accuracy, precision, recall, and F1-score. The results show that the proposed method achieves significantly higher accuracy compared to existing approaches.

1.1. Main contributions

The principal objective of the current paper is to improve the accuracy of insulator DD in power transmission lines. In this regard, a new and enhanced version of the Mask R-CNN network has been presented. Therefore, the significant contributions of the current article can be mentioned as a concern:

- Using a rhombus-shaped structure in the arrangement of the fully connected layers in the classification branch of the Mask R-CNN network, instead of the classic sequential structure, can enhance the performance of defect detection.
- To reduce the effect of overfitting in the Mask R-CNN, which occurs during network training, a dropout layer in the segmentation branch can be used in the middle of the fully connected layers.
- Choosing a suitable network as the backbone in the Mask R-CNN can increase the final accuracy of DD. In this article, the transfer learning technique is used to train the model. Accordingly, ResNet-101 is fine-tuned as the backbone.

In order to prove the main contributions, experiments are designed and the results are reported in the Section 4.

1.2. Paper organization

The principal goal of the current article is to suggest an efficient method for automatic insulator DD in power transmission lines. Therefore, the remainder of this article is organized as follows: Related works in this field are examined to highlight their advantages and limitations in the following section. In Section 3, the proposed method is explained in detail. In this respect, first, the Mask R-CNN is described, and then the proposed deep backbone network and input channels are incorporated to design the improved deep network. In Section 4, datasets and performance evaluation metrics are introduced first. Next, the performance of the proposed method is evaluated in terms of precision and accuracy. At the end of this section, our improved deep R-CNN is compared with other effective existing methods in the field of insulator DD. Finally, in Section 5, the conclusion along with suggestions for future work is presented.

2. Related works

As mentioned above, the detection of defects in insulators of power transmission lines is a relatively new and active research topic that has attracted the attention of investigators in the last decade. The early methods that were proposed to solve this problem were based on image processing techniques, but in recent years, due to the increase in hardware computational power, most of the methods have been presented based on DL techniques. In this section, some of the efficient methods presented in this field are examined.

As one of the initial researchers, Jabid and Ahsan [9] proposed an unconventional approach for insulator DD based on statistical texture elements. In [9], a texture descriptor called Rotation-Invariant Regional Directional Pattern (RI-LDP) is proposed to describe insulators. RI-LDP is designed based on the Local Directional Pattern (LDP), in which local texture patterns are encoded into an 8-bit binary code. Insulator partitioning is performed based on an image segmentation technique. Next, the histogram of RI-LDP is used to extract features, and finally, Support Vector Machine (SVM) serves as the classifier [9]. As another machine learning-based method, Iruansi et al. [10] proposed a novel approach for power-line insulator DD based on the combination of supervised classifiers and texture features. In [10], preprocessing is first performed to reduce noise using a top-hat morphological filter. Next, a multi-step algorithm is applied for insulator

segmentation. A Local Binary Patterns (LBP) descriptor is employed to extract texture features from segmented regions of interest (ROI). Finally, by training on defect and defect-free samples, SVM is utilized to classify power-line insulator conditions [10].

As mentioned above, recent studies have utilized the potential of DL techniques to segment and classify insulator defects. Siddiqui et al. [11], unlike previous handcrafted-feature-based methods, presented an automatic framework based on convolutional neural networks. Their proposed method detects 17 different common insulator defect types. In this method, after insulator segmentation, cap cropping is performed based on geometric features [11]. Han et al. [12] proposed a multi-step cascaded method for fault detection in aerial images. First of all, a dataset of common faults called IMF-Detection is collected. Next, a spatial pyramid pooling network is used to locate the insulator string. Finally, a deep YOLOv3-Tiny network is trained to detect faults [12]. Panigrahy et al. [13] developed a convolutional neural network to classify insulator conditions in high-voltage power lines. They designed a classical CNN consisting of sequential blocks of convolution, pooling, and activation function layers. In their experiments, different optimizers such as Adagrad, Adadelta, Adam, Nadam, and RMSprop were used to train the designed CNN. Finally, the Adamax optimizer achieved the highest accuracy on the CPLID dataset.

Chen et al. [14] used a combination of multi-granularity conditional generative adversarial networks (MC-GANs) to detect faults in transmission lines. Their proposed generator framework includes three components: a coarse-grained module, a Monte Carlo search algorithm, and a fine-grained module [14]. In some aerial images, there is more than one defect in a transmission line. In this regard, Liu et al. [15] proposed an insulator DD approach based on DL. In [15], a novel deep network called Cross Stage Partial Dense YOLO (CSPD-YOLO) is presented, based on YOLOv3. YOLOv3 is a multi-scale deep neural network that consists of a feature extraction backbone and multiple detection heads to detect objects at different scales. YOLOv5 was applied by Liquan et al. [16] for automatic DD in power transmission lines. The main goal in [16] is to increase the speed of fault detection compared to previous methods. In this respect, the original ResNet in the CSP-X module of YOLOv5 is replaced by an improved ResNet, which consists of three branches [16].

Dong et al., [17] presented a hybrid method to detect transmission line components and insulator defects. In this respect, an improved version of cascade R-CNN is designed by incorporating Swin-V2 and a balanced feature pyramid [17]. The Swin-V2 is a kind of deep Transformer in which hierarchical feature maps by merging image patches are constructed. As a recent development, PDDD-Net is presented by Sui and Wang for DD in power lines [18]. PDDD-Net is based on two main mechanisms as parallel attention mechanism (PAM) and dual-channel spatial pyramid pooling-fast block (DC_SPPF). SPPF is used in this method to extract richer features and improve DD performance [18].

Zhang et al., [34] introduced an innovative Ghost module in the backbone of the basic YOLOv5 network to decrease the internal parameters and improve the performance of DD in UAV images. In addition, the backbone of YOLOv5 is optimized by involving convolutional block attention

modules (CBAM) to concentrate on essential information for insulators to conceal uncritical information and DD. Chen et al., [36] presented an improved version of YOLOv7 network. In this method, a novel Edge Detailed Shape Data Augmentation (EDSDA) technique is performed to improve the model's sensitivity to insulator's edge forms. Also, a Cross-Channel and Spatial Multi-Scale Attention (CCSMA) module is added to YOLOv7 to increase the attention of the network to high-level insulator defect elements. As main innovation, a Re-BiC module is designed to connect multi-scale contextual features. Also, the MPDIoU function is used in [36] to compute the model's localization loss. Wang et al., [38] proposed a trainable bag-of-freebies acquainted answer for insulator DD. Adaptable and efficient activity devices with their offered architecture are combined by Wang et al., in this method. Deep YOLOv7 network reaches all learned object sensors in both accuracy and speed in this reference [38].

Moradi et al. [39] presented a hybrid approach for fault diagnosis in power transformers, integrating robust algorithms with tree-based classifiers. Their method employs an artificial neural network (ANN) optimized to enhance diagnostic accuracy for identifying faults in transformer insulation systems. The proposed framework demonstrated improved precision in detecting and classifying multiple fault types in real-world scenarios, indicating the potential for hybrid approaches to boost reliability in power system monitoring. Ghahraei et al. [40] developed an intelligent harvesting robot for button mushrooms, combining an expert system with image processing techniques in a shelf cultivation environment. Their system utilized vision-based segmentation and classification to detect and pick mushrooms, achieving a low error rate of 9.45% in detection. This work illustrates the applicability of image segmentation and object recognition for automated systems, sharing common challenges with defect detection tasks in power transmission networks. Rezaee et al. [41] developed an interactive segmentation method for medical images using an active contour model with improved energy tuning. While applied to medical imaging, their approach highlighted the value of robust segmentation for precise defect localization. Kalluri et al. [42] reviewed various deep learning techniques and filtering approaches. They highlighted the significant role of image segmentation and classification in accurately characterizing plant traits and health status. Their comprehensive review of deep learning models and such as CNNs and hybrid models, for defect detection applications.

Table 1 summarizes recent studies on insulator defect detection (DD) in power transmission lines, highlighting key aspects such as proposed methods, core architectures, datasets, strengths, and limitations. Early works primarily relied on handcrafted features combined with traditional classifiers like SVM, offering simplicity and low computational cost but limited robustness in complex backgrounds. More recent approaches have employed CNN-based models for automatic feature extraction, significantly improving detection accuracy but requiring substantial training data. Advanced methods leveraging YOLO variants, GANs, and Transformer-based models have demonstrated superior performance in multi-scale feature extraction and real-time defect detection. However, these methods often involve higher computational complexity and demand

powerful hardware resources, particularly in cluttered or natural scenes. This progression illustrates a clear trend toward deep learning and hybrid architectures that balance detection accuracy and runtime performance, yet highlights the ongoing challenge of handling complex environmental backgrounds and ensuring generalizability across diverse datasets.

Table 1. Comparison of Previous Studies on Insulator Defect Detection (DD).

Author (Year) [Ref.]	Proposed Method	Main Architecture / Model	Strengths	Limitations
Jabid et al. (2018) [9]	Rotation invariant local directional pattern (RI-LDP) with SVM	Handcrafted features + SVM	Rotation-invariant texture features	Lower accuracy compared to deep learning methods
Siddiqui et al. (2018) [11]	CNN-based defect detection	CNN	Fully automated defect classification	Requires large training datasets
Iruansi et al. (2019) [10]	LBP combined with SVM	Handcrafted features + SVM	Simple and low-cost approach	Sensitive to lighting variations and complex backgrounds
Han et al. (2020) [12]	Cascaded approach with Spatial Pyramid Pooling and YOLOv3 Tiny	YOLOv3 Tiny + SPP	Balanced speed and accuracy	Limited precision in complex backgrounds
Panigrahy et al. (2021) [13]	Customized CNN architecture	Classical CNN	Improved accuracy with different optimizers	Lack of dataset diversity
Chen et al. (2021) [14]	MC-GANs with coarse and fine modules	GAN-based	Fine defect simulation with GANs	Heavy computational requirements
Liu et al. (2021) [15]	CSPD-YOLO based on YOLOv3	YOLOv3	Multi-scale defect detection	High computational resources needed
Liquan et al. (2022) [16]	YOLOv5s with enhanced ResNet in CSP-X module	YOLOv5s	Faster defect detection speed	Limited real-world validation
Rezaee et al. (2022) [41]	Interactive medical image segmentation using active contour model	Active contour-based	Accurate interactive segmentation	Limited to medical imaging contexts
Fadhil et al. (2022) [43]	Breast cancer diagnosis using convolutional autoencoders + feature selection	Convolutional autoencoders	High-precision classification in medical images	Requires large, diverse medical datasets
Dong et al. (2023) [17]	Improved cascade R-CNN with Swin Transformer	Swin-v2 + Cascade R-CNN	Strong multi-scale feature extraction	High computational complexity
Sui et al. (2023) [18]	PDDD-Net with PAM and DC-SPPF	CNN-based	Better feature extraction with parallel attention	High memory requirements
Zhang et al. (2023) [34]	Ghost module in YOLOv5	YOLOv5 + Ghost module	Parameter reduction	Limited improvement in cluttered scenes
Wang et al. (2023) [38]	YOLOv7 with optimized bag-of-freebies	YOLOv7	High speed and strong accuracy	No segmentation capability
Ghahraei et al. (2023) [40]	Mushroom harvesting robot with expert system and image processing	Vision-based + expert system	Low detection error in structured environment	Limited to agricultural applications

Chen et al. (2024) [36]	Improved YOLOv7 with EDSDA and CCSMA modules	YOLOv7 + attention modules	Enhanced edge sensitivity	High complexity and heavy resources needed
Kalluri et al. (2024) [42]	Review of deep learning in plant phenotyping	Hybrid CNN-transformer review	Comprehensive review and future directions	No direct experiments; general insights
Moradi et al. (2025) [39]	Hybrid approach for transformer fault diagnosis	ANN + robust/tree-based	Enhanced diagnostic precision in transformers	Focused on transformer applications

The review of the articles that have been presented in this field so far shows that binary classification as defect or defect-free insulator is not enough, and segmenting the defective part can increase the speed of repairs. Also, the review of the related papers shows that the use of classical methods such as texture and color analysis of objects is not very efficient and it is recommended to employ deep neural networks. Also, some of the studies of previous works indicate that it is more difficult to detect defects in complex situations, such as when there is more than one insulator in the captured image or when the background is very diverse, and it displays the network dependence on the training.

More recently, advanced object detection models such as YOLOv8, DETR, and Swin Transformer-based methods have emerged, showing remarkable improvements in both accuracy and real-time detection capabilities. YOLOv8, as an extension of YOLOv7, offers better performance through improved architecture and training strategies. Transformer-based models like DETR and Swin Transformer have demonstrated strong capabilities in capturing global context and multi-scale feature interactions, making them promising candidates for defect detection in complex scenarios. Hybrid CNN-transformer models combine the local feature extraction of CNNs with the global reasoning ability of transformers, potentially enhancing robustness in cluttered environments.

3. The proposed insulator defect detection approach

A Convolutional Neural Network (CNN) is a type of deep artificial neural networks (ANNs) widely utilized in data analysis and image recognition. CNNs serve as essential tools in diverse computer vision studies, such as image segmentation [19]. The conventional structure of CNNs contains three immediate layers: pooling, convolutional, and fully connected layers. By sequentially combining these layers, CNNs can learn to detect and recognize objects in an image. However, in scenarios involving multiple objects within an input image, the traditional CNN framework may not yield optimal results.

To address this limitation, Mask R-CNN, an advanced DL architecture, was introduced building on the R-CNN framework. The term R-CNN refers to region-based CNN, a specialized neural network designed for identifying Regions of Interest (ROI) [20]. The R-CNN model operates by placing bounding boxes around object regions and then applying convolutional networks independently to

each ROI, thereby classifying different segments of an image. Fast R-CNN, an enhanced version of R-CNN, features a two-stage structure consisting of a Region Proposal Network (RPN) and a standard CNN. The RPN component functions as a neural network that identifies multiple objects within an image [21]. This process involves extracting features from RPN employing RoIPool for each candidate box, followed by classification and bounding-box regression. RoIPool extracts a compact feature map from each detected ROI. Unlike the original R-CNN, Fast R-CNN eliminates the need for an extensive number of region proposals before passing them to the next CNN stage, significantly increasing processing speed.

Image segmentation involves dividing an input image into multiple continuous pixel groups, known as segments. This technique is commonly employed to detect objects and define their boundaries (e.g., lines and curves). Consequently, detecting surface defects can be framed as an image segmentation task, where the objective is to classify image regions into defective and non-defective areas. Mask R-CNN, a variant of CNNs, is particularly effective for image segmentation tasks [22]. The Mask R-CNN architecture was initially developed as an enhancement of Faster R-CNN. Unlike Faster R-CNN, in which two outputs were produced: class labels and bounding-box coordinates, Mask R-CNN introduces a third output known as the object mask. By incorporating this additional output, Mask R-CNN facilitates pixel-level instance segmentation alongside object detection. This mechanism ensures the generation of highly precise segmentation masks for each detected object, enabling fine-grained boundary delineation at the pixel level for more accurate segmentation.

The principal purpose of the current article is to detect defects in power transmission lines. When analyzing aerial images, two primary challenges arise: insulator segmentation and defect detection (DD). To address these issues, Mask R-CNN is chosen as the method due to its capability to provide three essential outputs: bounding box, mask, and label. This article presents an enhanced version of the Mask R-CNN model, aimed at improving the accuracy of DD in power transmission lines. The improvements are implemented in two phases. The first phase involves modifying the backbone of the Mask R-CNN to enhance the precision of insulator segmentation. In the second phase, adjustments are made to the network head to further increase the overall accuracy of DD. The architecture of the improved Mask R-CNN model is displayed in Fig. 2.

In Mask R-CNN, the backbone serves as the initial component responsible for extracting feature maps. In many studies and applications, ResNet has been widely adopted as the backbone for Mask R-CNN [23-25]. ResNet is favored over traditional CNNs due to its residual learning mechanism, which enables deep layers to learn directly from shallower layers, thereby accelerating network convergence. While most related studies have utilized ResNet-50 as the backbone [26-27], this paper employs ResNet-101. The rationale behind this choice is that in certain aerial images, multiple insulators may be present, and defect sizes can vary significantly. By using ResNet-101, the model is expected to better handle these variations and improve detection accuracy.

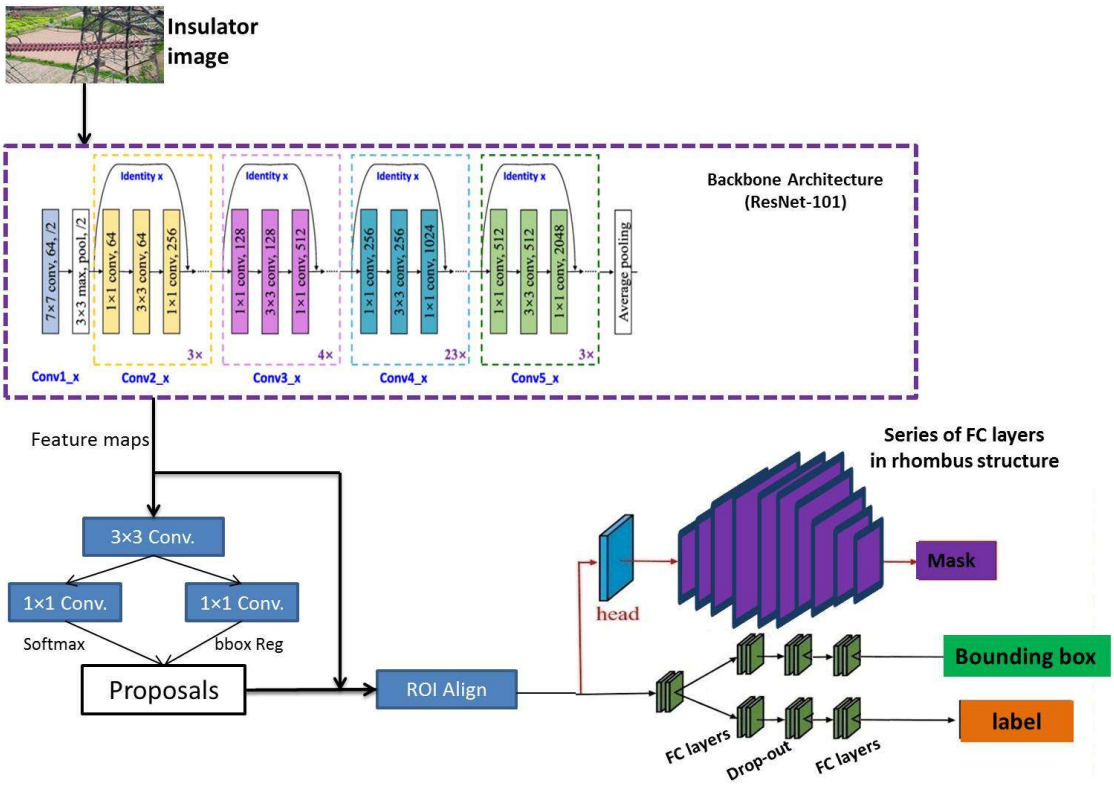

Fig.2 The architecture of the suggested improved Mask R-CNN

ResNet-101 consists of 101 layers, making it significantly deeper than ResNet-50 [28]. This increased depth enhances its ability to handle complex tasks such as insulator DD, leading to improved accuracy. However, while deeper versions of ResNet exist, their use considerably increases computational complexity and significantly reduces real-time processing capabilities. For instance, ResNet-152 comprises 58.5 million parameters, making it substantially slower compared to ResNet-101, which has only 42.8 million parameters [29]. The architecture of ResNet-101, including layer types and internal feature map dimensions, is depicted in Fig. 3. To extract feature maps, this study employs a pre-trained ResNet-101 model trained on the COCO dataset, with the final fully connected layer omitted.

Layer name	Output size	101-layer
Conv1	112×112	7×7, 64, stride 2
Conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \end{bmatrix}\times3$
Conv3_x	28× 28	$\begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \end{bmatrix}\times4$
Conv4_x	14 × 14	$\begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \end{bmatrix}\times23$
Conv5_x	7 × 7	$\begin{bmatrix} 1\times1, & 512 \\ 3\times3, & 512 \\ 1\times1, & 2048 \end{bmatrix}\times3$
	1 × 1	**Average pool, 1000-d fc, softmax**

Fig.3 The internal architecture of ResNet-101 with details

One of the key branches in the head of the Mask R-CNN is responsible for assigning labels to the input images. In DD tasks, two main classes are considered: normal insulators and defective insulators. This classification process is carried out using a series of fully connected layers designed to extract the final distinguishing features. The classification itself is performed by a softmax layer. In the standard Mask R-CNN architecture, multiple fully connected layers of equal dimensions are used. While this structure is trainable, it may not be optimal for detecting defects in power transmission lines.

Defects in power transmission lines often differ significantly from their surrounding background. Using fully connected layers of uniform dimensions can cause highly discriminative features, crucial for label classification, to be influenced by background features, reducing classification efficiency in the final layer. To address this issue, the improved version of Mask R-CNN introduced in this study utilizes a rhombus-shaped structure for the fully connected layers in the label detection branch. In this design, the output size of convolutional layers first increases alternately across some layers and then decreases periodically toward the end of the structure. This adaptive scaling enhances feature extraction and improves classification performance.

The result attribute map size of a convolution layer can be calculated using the next equation.

$$Output\ size = \frac{I - K + 2P}{S} + 1 \quad (1)$$

Where, I means the input feature map size, K shows the kernel size in square format and P is padding size. Also, S shows the stride size in convolution layer. As can be seen in the Eq. 1, padding size can be used to increase the output size in comparison with input size. If S=1 is considered, the following condition must be met to produce a larger feature map in the output.

If O – I > 0 then 2P – K > 0 (2P > K) (2)

The rhombus-shaped structure for a series of blocks included convolution, activation layer and pooling layers is shown in the Fig.4.

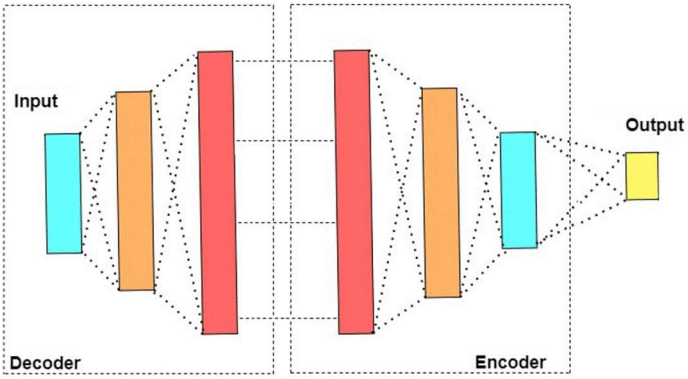

Fig.4 The deep rhombus-shaped structure for a series of blocks

Fig.4 illustrates the proposed deep rhombus-shaped classification head, designed to enhance feature abstraction in the final stage of the network. This architecture leverages variable-sized fully connected (FC) layers, a concept explored in recent adaptive deep network designs, such as pyramid residual attention modules and spatial–relation segmentation structures [43, 44]. These variable-dimension FC structures allow for gradual feature expansion and compression, promoting richer encoding while suppressing background interference. The alternating increase and decrease in layer size also encourage better generalization, as shown in lightweight segmentation frameworks that utilize adaptive pooling and residual connections [45]. This design provides a structured balance between high-level abstraction and the preservation of discriminative details, making it particularly effective for defect detection in complex backgrounds. To further validate the effectiveness of the rhombus structure, an ablation study was performed (Table 5), quantifying its contribution to detection accuracy.

An improved version of Mask R-CNN is presented in the current article which should be trained. Overfitting is a common problem that may occur in the learning process. Dropout is a regularization technique that is used in deep networks to prevent overfitting. The Dropout layer is a mask that removes the contribution of some hidden neurons toward the next layer. It randomly deactivates neurons during the training process and forces the deep network to learn more robust features. It improves generalization, training, and model performance on unseen data [30]. This layer sets selected input units to zero with a frequency of a given rate during training time, which aids in controlling overfitting. Neurons not set to zero are scaled up by 1 / (1 - rate) such that the sum of all inputs is unaffected. An example of the dropout layer process is demonstrated in Fig.5.

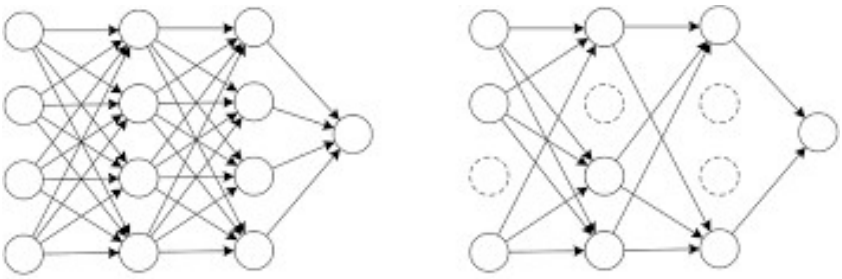

Fig.5 An example of Dropout layer process

Dropout can be used after any non-output layer in deep networks like Mask R-CNN. The general rule of thumb suggests adding this layer after the last pooling layer, but it depends to the size of the dataset and the acceptable complexity of the proposed model. So, it can be added to other layers as well. In our proposed improved Mask R-CNN a dropout layer is used after half of the fully connected layers in the mask detection branch of Mask R-CNN. The loss function of our proposed improved Mask R-CNN is calculated as follows:

$$L_{Total} = L_c + L_R + L_m \quad (3)$$

Where, L_c represents the classification loss parameter. Classification loss measures the discrepancy between the ground truth class labels for each area proposal and predicted class probabilities. Note that computed loss is of concern:

$$Lc = -(P_{true} \times \log(P_{pred}) + (1 - P_{true}) \times \log(1 - P_{pred})) \quad (4)$$

Where, P_{true} is the label of the ground truth which can be 0 or 1 and indicates the real true label of the sample. Also, P_{pred} is the anticipated likelihood of the positive class for the sample. L_R represents the regression loss, which estimates the contrast between the anticipated bounding box and the foundation fact bounding box in terms of coordinates for each region proposal. It is computed using the following equation.

$$L_R = \sum_i R(C_i - C_i^*) \quad (5)$$

Where, index i symbolizes the desired sample, and C_i is the anticipated framing coordinate vector for the i^{th} example. Also, C_i^* shows the coordinate vector corresponding to the frame of the real zone for the i^{th} sample, and R is the loss function, which computes the element-wise loss between the anticipated and target coordinates. Mask loss denoted by L_M, which computes the dissimilarity between the anticipated mask and the ground truth mask for each area suggestion.

$$L_M = \sum_{i=1}^{N} [x_i^* \times \log\big(P(x_i)\big) + (1 - x_i^*) \times \log(1 - P(x_i))] \quad (6)$$

Where **N** shows the whole number of pixels in the mask and parameter i iterates over the pixels. The term x_i^* is the classification label where the pixel is found, and $P(X_i)$ is the likelihood of the anticipated classification (X_i).

4. Experimental Results

The suggested method for detecting defects in transmission line insulators is implemented in a software environment that includes Python 3.11, TensorFlow 2.15, Keras 2.15, and Windows 10. The deep network is deployed on a physical system equipped with a 16-core CPU, 16 GB of RAM, and an RTX 4060 GPU.

For training the proposed model, the hyper parameters are set, including 1000 epochs, a batch size of 8, and a confidence threshold of 0.3. The learning rate is initially set to 0.001 for the first 900 epochs and is then reduced to 0.0001 for the final 100 epochs. To rev convergence, the pre-trained weights of the Mask R-CNN model trained on the COCO dataset are utilized [31].

4.1. Datasets

This investigation strives to develop an efficient method for DD in power transmission lines. To assess the usefulness of the suggested approach, a benchmark dataset in this domain is used, allowing for direct comparison with existing methods. The Insulator Defect Image Dataset (IDID) contains 7,568 images of power transmission lines captured in natural environments. These images are categorized into three groups: normal insulators, insulators with defects caused by flashover, and insulators with defects due to shell breakage. Specifically, the dataset contains 2,636 images of fully intact insulators, 1,140 images of insulators with damaged coatings, and 2,004 images of power transmission line insulators affected by flashover [32]. Examples of normal and defective insulators from the IDID dataset are illustrated in Fig.6.

(a) Normal insulator samples (b) Defective insulator sample

Fig.6 Some samples of CPLID dataset

4.2. Performance evaluation metrics

In the experiments, the performance of the suggested method is evaluated using common benchmark metrics, including recall (R), mean average precision (mAP), average precision (AP), and precision (P). The equations for calculating recall and precision are given below:

$$\text{Precision} = \text{TP}/\ \text{TP} + \text{FP} \tag{7}$$

$$\text{Recall} = \text{TP}/\ \text{TP} + \text{FN} \tag{8}$$

Where true positive (TP) symbolizes the number of correctly classified examples, TP + FP denotes the total number of predicted positive instances, and TP + FN symbolizes the total number of actual positive instances.

In multi-class classification problems such as DD, the definitions of mean average precision (mAP) and average precision (AP) are described as the next equation:

$$\text{AP} = \int_0^1 P(R)dr \tag{9}$$

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{10}$$

In many cases, the F1 score as an independent metric, which symbolizes the harmonic average of recall and precision. The definition of the F1 score is as follows:

$$F_1\ Score = 2 \times \frac{P * R}{P + R} \tag{11}$$

4.3. Performance analysis of the proposed method

The suggested deep neural network in this study generates three outputs for per input image: class label, bounding box, and object mask. Consequently, the effectiveness of the method has been evaluated in two key areas: insulator segmentation in power transmission lines and DD.

4.3.1. Results of insulator segmentation

As discussed earlier, the first phase of the improved Mask R-CNN involves insulator segmentation, which is handled by the bounding box segmentation branch in the proposed network. Some previous studies classify defects without explicitly identifying the insulator in the image, often analyzing the entire image for labeling. In contrast, one of the fundamental benefits of the presented method is its precise insulator segmentation, which enhances the background knowledge available to repair personnel, reduces operational costs, and minimizes the UAV's dependency on high zoom capabilities. To evaluate the usefulness of the proposed approach, experiments were conducted on the IDID. Since the dataset includes ground-truth annotations for all images, the output of the bounding box segmentation branch was quantitatively compared

against the ground truth at the pixel level. The performance results for insulator segmentation in power transmission line images with natural backgrounds are presented in the table below. Additionally, Fig.**7** illustrates examples of insulator segmentation results for both normal and defective insulators.

Table 2. The performance evaluation of the suggested method for insulator location on IDID

Metrics / Method	Precision	Recall	Accuracy
Improved Lightweight YOLOv4 [33]	93.29	91.77	91.89
Pre-trained Faster R-CNN	94.38	92.27	92.76
Proposed method (Improved Mask R-CNN)	98.89	98.60	98.81

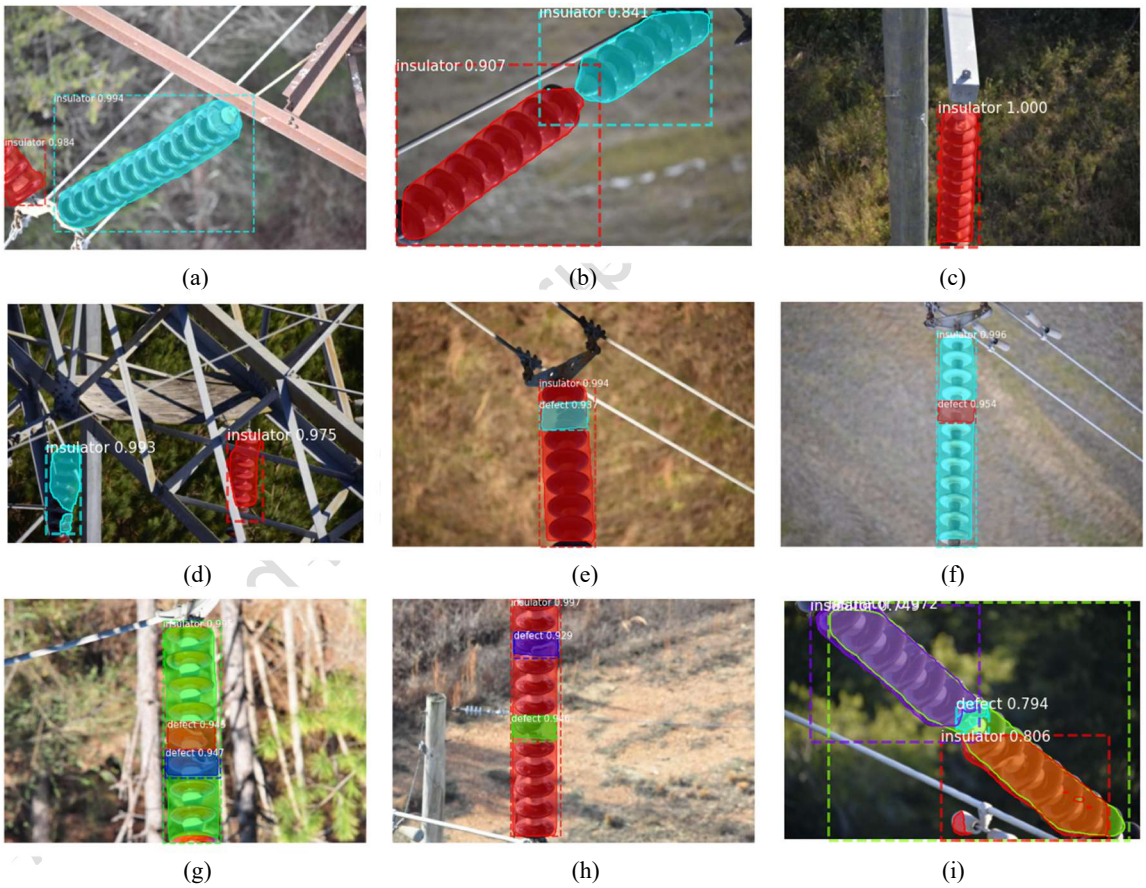

Fig.7 Some examples of the proposed output results for insulator segmentation

(a) Two normal cross insulator in power transmission line (b) Two normal continuous insulators (c) One normal insulator (d) Two normal parallel insulators (e) One defective insulator (f) One defective insulator (g) One insulator defected in two areas (h) One insulator defected in two areas (i) One defective insulator.

4.4. Comparison performance with existing methods in scope

To verify the performance of the suggested method, some efficient approaches in this scope were implemented. In this respect, three types of deep object detection methods were used such as Faster R-CNN, YOLOv3 and YOLOv4. The Insulator Defect Image Dataset (IDID) was likewise employed for training and testing the object detection models. The extracted outcomes are presented in the Table 3.

Table 3. The performance comparison of the suggested method with existing object detection methods

Method	AP (%)		mAP (%)	F_1 Score (%)	Accuracy
	Insulator	Defect			
Faster R-CNN	96.87	67.85	82.36	79.85	83.33
YOLOv3	92.18	82.14	87.16	86.32	87.50
YOLOv4	92.18	87.50	89.84	89.61	90.01
MobileNet_CenterNet [6]	97.20	81.70	89.45	NR	NR
YOLOv7t [38]	NR	83.8	84.3	86.6	NR
Proposed method (Improved Mask R-CNN)	**98.43**	**85.71**	**92.07**	**93.32**	**92.50**

In order to investigate the first contribution, the suggested model performance was evaluated based on several pre-trained deep neural networks as backbones. The evaluation outcomes are displayed in Table 4.

Table 4. The evaluation of the performance of the proposed method based on different deep networks as backbone

Backbone	AP (%)		mAP (%)	F_1 Score (%)	Accuracy
	Insulator	Defect			
ResNet-101	98.43	85.71	92.07	93.32	92.50
ResNet-50	97.36	84.67	90.92	92.29	91.43
MobileNetV2	98.30	85.55	91.98	93.17	92.28
VGG-16	95.89	82.68	89.47	90.36	89.66

An ablation study was conducted to quantify the contributions of the proposed rhombus-shaped classification head and the dropout layer individually. As shown in Table 5, the addition of the rhombus head improves the mAP and F1 scores by approximately 2%, while the dropout layer provides a slight improvements in generalization. The combined use of both modifications results in the highest performance, validating the effectiveness of these enhancements in improving defect detection accuracy.

Table 5. Ablation Study of the Proposed Model's Key Enhancements (Rhombus Head and Dropout Layer)

Model Configuration	mAP (%)	Accuracy (%)	F1 Score (%)
Baseline Mask R-CNN	82.5	89.2	88
Baseline + Rhombus Head	84.7	90.3	89.5
Baseline + Dropout	83.9	89.8	88.9
Improved Mask R-CNN (both)	85.7	92.5	92

The analysis of the experiment results displays that the method presented in this paper generally provides higher precision, mean average precision, F_1 score and accuracy than efficient methods in the field of power transmission line DD. The efficiency of the YOLOv4 network for detecting defects in images that actually have defects is about 1.79 percent higher than our presented improved Mask R-CNN. However, due to the fact that more power lines are defect-free, the mAP and accuracy of YOLOv4 network is about 2.23 percent and 2.49 percent lower than our proposed method. Also, the YOLOv4 network only performs the classification process and cannot segment the defect area. But our proposed improved Mask R-CNN has the capability of segmenting defects, locating insulator and even producing defect mask, which decrease the repairing process time and costs.

5. Discussion

As the outcomes demonstrate, the method presented in the current article has provided higher accuracy than the common deep networks in this field and related works. Using the rhombus structure instead of the classical structure in the classification branch of the Mask R-CNN network has enhanced the detection accuracy. Also, despite the long epochs of learning the network, over-fitting has not had a significant influence on the efficiency of the proposed network, which can be attributed to the use of a dropout layer in the middle of fully connected layers in the segmentation branch. Moreover, the comparison with basic deep networks shows that the use of ResNet-101 network instead of ResNet-50 can be more efficient for this task.

As can be observed in Table 4, the outcomes indicate that changing the backbone network architecture in Mask R-CNN affects the final detection accuracy. Furthermore, the ResNet-101 network provided the highest detection accuracy in the experiments.

To further understand the limitations of the proposed method, we analyzed failure cases where the model misclassified or failed to detect defects. These typically occurred in images with highly cluttered backgrounds or low-contrast defects. Fig.7 illustrates several examples of such misclassifications. These insights suggest future work on integrating additional context-aware modules or transformer-based architectures to improve robustness in complex environments.

In addition to detection accuracy, we measured the runtime performance of our improved Mask R-CNN model. On an RTX 4060 GPU, the proposed model achieved approximately 22 frames per second (FPS) during inference, with a model size of around 175 MB. While this is slightly slower

than YOLOv4 (which runs at 30 FPS), our model's ability to generate precise defect masks offers an important trade-off between accuracy and speed for maintenance applications.

While our study focused on comparing with YOLOv4, YOLOv3, and YOLOv7 as baselines, newer models such as YOLOv8, DETR, and Swin Transformer have shown promising results in similar object detection tasks. Integrating these advanced architectures into the defect detection framework or exploring hybrid CNN-transformer approaches could further improve performance in highly cluttered scenes and real-world deployments. We consider this as an exciting direction for future work.

6. Conclusion

This article introduced an enhanced method for defect detection (DD) in power transmission lines, structured in two key stages: insulator segmentation and DD. To enhance accuracy, an optimized version of Mask R-CNN was developed, refining the DD and mask generation branches in the network's head. Experimental results exhibited that the suggested model outperforms existing methods in accurately classifying insulator defects. Unlike traditional machine learning and deep learning (DL) approaches, which mainly focus on classification, the proposed model not only identifies defects but also precisely segments both the insulator and defect bounding boxes. This segmentation capability is particularly valuable in images containing multiple insulators, reducing operational and financial costs for maintenance teams. Furthermore, precise defect localization provides useful insights into the size and nature of the damage, facilitating faster and more efficient repairs. For future research, the enhanced Mask R-CNN framework can be applied to other segmentation and classification tasks, such as detecting surface defects in industrial or manufacturing settings. Its modular architecture including mask generation, bounding box segmentation, and classification makes it suitable for a wide range of real-world applications that require precise defect detection and analysis. Additionally, integrating advanced models like YOLOv8, DETR, or Swin Transformer-based approaches into the defect detection framework may further improve performance, especially in cluttered environments. Exploring hybrid CNN-transformer architectures could also be a promising direction for future research to enhance robustness and generalizability.

Data Availability Statements (DAS)

To evaluate the suggested method of this research, two datasets IDID[1] and COCO[2] available in the footnote links and also exist in references [31] and [32] have been used.

ResNet-101 with 101 layers exists in [28]. ResNet-152 [29] has 58.5 million parameters and ResNet-101 with 42.8 million parameters [29]. The proposed transmission line insulator DD

[1] https://ieee-dataport.org/competitions/insulator-defect-detection, accessed on 14 January 2024.
[2] https://github.com/matterport/Mask_RCNN, accessed on 10 January 2024.

method is implemented in the software environment of Python 3.11, TensorFlow 2.15, Keras 2.15, and Windows 10. The pre-trained weights of Mask R-CNN on the COCO dataset are used [31]. The Insulator Defect Image Dataset (IDID) [32] contains 7568 images of power transmission lines with natural backgrounds, which are categorized into three groups: normal insulation, defective insulation due to flashover and defective insulators due to Shell-Broken. This database contains 2636 images of completely normal insulators and 1140 images of insulators with defective coating and 2004 images of insulators of power transmission lines that have been damaged by Flash over [32]. Examples of both normal images and defective images from IDID are displayed in the Fig.6. In order to verify the performance of the suggested method, three type of deep object detection methods were used, including Faster R-CNN, YOLOv3 and YOLOv4. The Insulator Defect Image Dataset (IDID) [32] was also used for training models and test desired object detection models.

References:

[1] Ayala García, A., Galván Bobadilla, I., Arroyo Figueroa, G., Pérez Ramírez, M., & Muñoz Román, J. (2016). Virtual reality training system for maintenance and operation of high-voltage overhead power lines. *Virtual Reality*, *20*, 27-40.

[2] Liu, J., Hu, M., Dong, J., & Lu, X. (2023). Summary of insulator defect detection based on deep learning. *Electric Power Systems Research*, *224*, 109688.

[3] Mirallès, F., Pouliot, N., & Montambault, S. (2014, October). State-of-the-art review of computer vision for the management of power transmission lines. In *Proceedings of the 2014 3rd international conference on applied robotics for the power industry* (pp. 1-6). IEEE.

[4] Zhou, Z., Zhang, C., Xu, C., Xiong, F., Zhang, Y., & Umer, T. (2018). Energy-efficient industrial Internet of UAVs for power line inspection in smart grid. *IEEE Transactions on Industrial Informatics*, *14*(6), 2705-2714.

[5] Ni, L., Ma, Y., Lin, Q., Yang, J., & Jin, L. (2021, October). Research on Insulator Defect Detection Method Based on Image Processing and Watershed Algorithm. In *2021 International Conference on Advanced Electrical Equipment and Reliable Operation (AEERO)* (pp. 1-6). IEEE.

[6] Xia, H., Yang, B., Li, Y., & Wang, B. (2022). An improved CenterNet model for insulator defect detection using aerial imagery. *Sensors*, *22*(8), 2850.

[7] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[8] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448)

[9] Jabid, T., & Ahsan, T. (2018). Insulator detection and defect classification using rotation invariant local directional pattern. *International Journal of Advanced Computer Science and Applications*, *9*(2).

[10] Iruansi, U., Tapamo, J. R., & Davidson, I. E. (2019). Classification of power-line insulator condition using local binary patterns with support vector machines. *IAENG International Journal of Computer Science*, *46*(2), 300-310.

[11] Siddiqui, Z. A., Park, U., Lee, S. W., Jung, N. J., Choi, M., Lim, C., & Seo, J. H. (2018). Robust powerline equipment inspection system based on a convolutional neural network. *Sensors*, *18*(11), 3837.

[12] Han, J., Yang, Z., Xu, H., Hu, G., Zhang, C., Li, H., ... & Zeng, H. (2020). Search like an eagle: A cascaded model for insulator missing faults detection in aerial images. *Energies*, *13*(3), 713.

[13] Panigrahy, S., Karmakar, S., & Sahoo, R. (2021, July). Condition Assessment of High Voltage Insulator using Convolutional Neural Network. In *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1-6). IEEE.

[14] Chen, W., Li, Y., & Zhao, Z. (2021). InsulatorGAN: A transmission line insulator detection model using multi-granularity conditional generative adversarial nets for UAV inspection. *Remote Sensing*, *13*(19), 3971.

[15] Liu, C., Wu, Y., Liu, J., Sun, Z., & Xu, H. (2021). Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model. *Applied Sciences*, *11*(10), 4647.

[16] Liquan, Z., Mengjun, Z., Ying, C., & Yanfei, J. (2022). Fast detection of defective insulator based on improved YOLOv5s. *Computational Intelligence and Neuroscience*, *2022*.

[17] Dong, C., Zhang, K., Xie, Z., & Shi, C. (2023). An improved cascade RCNN detection method for key components and defects of transmission lines. *IET Generation, Transmission & Distribution*, *17*(19), 4277-4292.

[18] Sui, T., & Wang, J. (2023). PDDD-Net: Defect Detection Network based on Parallel Attention Mechanism and Dual-Channel Spatial Pyramid Pooling. *IEEE Access*.

[19] Elngar, A. A., Arafa, M., Fathy, A., Moustafa, B., Mahmoud, O., Shaban, M., & Fawzy, N. (2021). Image classification based on CNN: a survey. *Journal of Cybersecurity and Information Management*, *6*(1), 18-50.

[20] Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to mask R-CNN: a survey. Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019, 657-668.

[21] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).

[22] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

[23] Wu, M., Yue, H., Wang, J., Huang, Y., Liu, M., Jiang, Y., Zeng, C. (2020). Object detection based on RGC mask R-CNN. IET Image Processing, 14(8), 1502-1508.

[24] Tian, Y., Yang, G., Wang, Z., Li, E., & Liang, Z. (2020). Instance segmentation of apple flowers using the improved mask R–CNN model. Biosystems engineering, 193, 264-278.

[25] Zhang, Y., Liu, Y. L., Nie, K., Zhou, J., Chen, Z., Chen, J. H., ... & Su, M. Y. (2023). Deep learning-based automatic diagnosis of breast cancer on MRI using mask R-CNN for detection followed by ResNet50 for classification. Academic radiology, 30, S161-S171.

[26] Zhang, Y., Liu, Y. L., Nie, K., Zhou, J., Chen, Z., Chen, J. H., ... & Su, M. Y. (2023). Deep learning-based automatic diagnosis of breast cancer on MRI using mask R-CNN for detection followed by ResNet50 for classification. *Academic radiology*, *30*, S161-S171.

[27] Wang, Y., Wang, X., Hao, R., Lu, B., & Huang, B. (2024). Metal Surface Defect Detection Method Based on Improved Cascade R-CNN. *Journal of Computing and Information Science in Engineering*, *24*(4), 041002.

[28] Dash, S. P., Ramadevi, J., Amat, R., Sethy, P. K., Behera, S. K., & Mallick, S. (2023, September). Wafer Defect Identification with Optimal Hyper-Parameter Tuning of Support Vector Machine using the Deep Feature of ResNet 101. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1291, No. 1, p. 012048). IOP Publishing.

[29] Yu, H., Miao, X., & Wang, H. (2022). Bearing fault reconstruction diagnosis method based on ResNet-152 with multi-scale stacked receptive field. *Sensors*, *22*(5), 1705.

[30] Munir, N., Kim, H. J., Song, S. J., & Kang, S. S. (2018). Investigation of deep neural network with drop out for ultrasonic flaw classification in weldments. *Journal of Mechanical Science and Technology*, *32*, 3073-3080.

[31] Waleed Abdulla. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub repository. Available online: https://github.com/matterport/Mask_RCNN (accessed on 10 January 2024).

[32] IEEE Dataport. Insulator Defect Detection Dataset - Competition. Available online: https://ieee-dataport.org/competitions/insulator-defect-detection (accessed on 14 January 2024).

[33] Qiu, Z., Zhu, X., Liao, C., Shi, D., & Qu, W. (2022). Detection of transmission line insulator defects based on an improved lightweight YOLOv4 model. *Applied Sciences*, *12*(3), 1207.

[34] Chen, Y., Liu, H., Chen, J., Hu, J., & Zheng, E. (2023). Insu-YOLO: an insulator defect detection algorithm based on multiscale feature fusion. *Electronics*, *12*(15), 3210.

[35] Zhang, T., Zhang, Y., Xin, M., Liao, J., & Xie, Q. (2023). A light-weight network for small insulator and defect detection using UAV imaging based on improved YOLOv5. *Sensors*, *23*(11), 5249.

[36] Chen, B., Zhang, W., Wu, W., Li, Y., Chen, Z., & Li, C. (2024). ID-YOLOv7: an efficient method for insulator defect detection in power distribution network. *Frontiers in Neurorobotics*, *17*, 1331427.

[37] Chen, Y., Deng, C., Sun, Q., Wu, Z., Zou, L., Zhang, G., & Li, W. (2024). Lightweight Detection Methods for Insulator Self-Explosion Defects. *Sensors*, *24*(1), 290.

[38] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). 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* (pp. 7464-7475).

[39] Moradi, E. (2025). Accuracy Enhancement of Fault Diagnosis for Power Transformers with a Hybrid Approach Integrating Robust and Tree-Based Algorithms. *Majlesi Journal of Electrical Engineering*, 19(1), 1–9.

[40] Ghahraei, O., Kheshtzarrin, M., & Amiri, M. (2023). Development of a Button Mushroom Harvesting Robot Using Expert System and Image Processing in Shelf Cultivation Method. *Majlesi Journal of Electrical Engineering*, 17(2), 45–52.

[41] Rezaee, K., Ahmadi, M. K. N., & Anari, M. S. (2022). Interactive Medical Image Segmentation Using Active Contour with Improved F Energy in Level-Set Tuning. *Majlesi Journal of Electrical Engineering*, 16(4), 210–218.

[42] Kalluri, R. D., & Selvaraj, P. (2024). A Review on Application of Various Deep Learning Techniques and Filtering Approach in Plant Phenotyping. *Majlesi Journal of Electrical Engineering*, 18(2), 103–110.

[43] Jiang, Y., Cheng, T., Dong, J., Liang, J., Zhang, Y., Lin, X., & Yao, H. (2022). Dermoscopic image segmentation based on Pyramid Residual Attention Module. PLoS ONE, 17(9), e0267380. https://doi.org/10.1371/journal.pone.0267380

[44] Sun, Y., Malihi, S., Li, H., & Maboudi, M. (2022). DeepWindows: Windows instance segmentation through an improved Mask R-CNN using spatial attention and relation modules. ISPRS International Journal of Geo-Information, 11(3), 162. https://doi.org/10.3390/ijgi11030162.

[45] He, M., He, K., Huang, Q., Xiao, H., Zhang, H., Li, G., & Chen, A. (2025). Lightweight Mask R-CNN for instance segmentation and particle physical property analysis in multiphase flow. Powder Technology, 449, 120366. https://doi.org/10.1016/j.powtec.2024.120366.