Article

A Method Based on Multi-Network Feature Fusion and Random Forest for Foreign Objects Detection on Transmission Lines

Yanzhen Yu [1], Zhibin Qiu [1,*], Haoshuang Liao [2], Zixiang Wei [3], Xuan Zhu [1] and Zhibiao Zhou [1]

[1] Department of Energy and Electrical Engineering, School of Information Engineering, Nanchang University, Nanchang 330031, China; yyz127@email.ncu.edu.cn (Y.Y.); zx1996@email.ncu.edu.cn (X.Z.); zzb1997@email.ncu.edu.cn (Z.Z.)

[2] State Grid Jiangxi Electric Power Research Institute, Nanchang 330096, China; hasaily@sina.cn

[3] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou 215028, China; zixiang.wei19@student.xjtlu.edu.cn

* Correspondence: qiuzb@ncu.edu.cn

Abstract: Foreign objects such as kites, nests and balloons, etc., suspended on transmission lines may shorten the insulation distance and cause short-circuits between phases. A detection method for foreign objects on transmission lines is proposed, which combines multi-network feature fusion and random forest. Firstly, the foreign object image dataset of balloons, kites, nests and plastic was established. Then, the Otus binarization threshold segmentation and morphology processing were applied to extract the target region of the foreign object. The features of the target region were extracted by five types of convolutional neural networks (CNN): GoogLeNet, DenseNet-201, EfficientNet-B0, ResNet-101, AlexNet and then fused by concatenation fusion strategy. Furthermore, the fused features in different schemes were used to train and test random forest, meanwhile, the gradient-weighted class activation mapping (Grad-CAM) was used to visualize the decision region of each network, which can verify the effectiveness of the optimal feature fusion scheme. Simulation results indicate that the detection accuracy of the proposed method can reach 95.88%, whose performance is better than the model of a single network. This study provides references for detection of foreign objects suspended on transmission lines.

Keywords: transmission lines; foreign object detection; feature fusion; random forest; convolutional neural network

1. Introduction

Overhead transmission lines are the most commonly used method for power delivery, whose safety and stability are the premise of normal power supply, thus the inspection and maintenance of transmission lines are of vital importance [1,2]. With the increase in power demands, the geographical span of transmission lines is also gradually increasing, which may pass through densely populated urban areas and construction sites. Due to the diverse distribution areas of transmission lines and the influences of human and bird activities, etc., the transmission corridors may be invaded by foreign objects such as kites, balloons, plastic and nests, and results in short-circuit and ground discharges [3,4].

Traditional transmission line inspection is usually carried out by manual inspection. Due to the complex terrain environment along the transmission corridors and the possible artificial misjudgment, manual inspection is unable to meet the practical requirements [5]. In recent years, unmanned aerial vehicle (UAV) [6,7] and monitoring equipment installed on transmission towers [8] have been widely used to acquire inspection images. In addition, the classification, recognition and object detection of the massive inspection images using computer vision algorithms have been widely studied.

At present, the object detection methods are generally divided into two main categories. One is traditional object detection that combines image processing, feature extraction

and object identification by classifiers, the other is the detection method based on deep learning. Traditional detection method adopts manual feature extraction and machine learning classifiers, the input images often should be preprocessed to deal with the effects of complex backgrounds. Yao et al. [9] proposed a method based on GMM and K-means to cluster and detect foreign objects. Taskeed et al. [10] proposed a method for rotation invariant insulator detection that extracted the sliding window based on local directional pattern (LDP) feature from the image and used the support vector machines (SVM) for classification of those sliding windows. Lu et al. [11] put forward a detection method for bird's nests on transmission towers based on cascade classifier and combination features, whereas the images were described by the proportion of white area (PWA), ratio of white pixels (RWP), projection feature (PF) and improved burr feature (IBF), and recognized by the cascade classifier. Traditional detection methods require a series of image preprocessing and manual feature extraction to detect the targets [12], the process is complicated, with poor universality and low accuracy in practical complex engineering environment.

In recent years, deep learning algorithms have been widely used in transmission line defects detection and perform well in accuracy [13]. Ni et al. [14] proposed a defect recognition method for key components of transmission lines based on improved Faster R-CNN, which used the concept-ResNetv2 feature extraction network to achieve the network structure adjustment and parameter optimization. Zu et al. [15] proposed a foreign object detection model based on Faster R-CNN using data augmentation method by rotation, adjusting brightness and color saturation, adding Gaussian noise to expand the images, the recognition accuracy of the improved model can reach 93.09%. Xu et al. [16] proposed an efficient foreign objects detection network for power substation (FODN4PS), which consists of a moving object region extraction network and a classification network. The experiment proved that this model has a higher accuracy than Fast R-CNN and Mask R-CNN. Zhai et al. [17] proposed a hybrid knowledge region-based convolutional neural network (HK R-CNN) to detect aerial images, and the results show that it has higher performance than Faster R-CNN, which can accurately detect the multiple fittings on transmission lines. Zhang et al. [18] proposed an automatic detection method based on RetinaNet and the model for nest detection was established by adjusting the appropriate network structure and parameters. The recognition accuracy can reach 94.1%, which is higher than YOLO and SDD, but the recognition is slow. Li et al. [19] adopted an improved YOLOv3 to shorten the time for foreign object detection by replacing Darknet-53 with lightweight MobileNet on the skeleton of the original YOLOv3. Chen et al. [20] proposed an insulator recognition method for power distribution lines based on modified YOLOv3. The K-means algorithm was used to cluster the bounding boxes, and the DenseNet was introduced to replace the DarkNet53. The accuracy was increased by 10% than the original YOLOv3. Song et al. [21] added K-means clustering and DIoU-NMS to YOLOv4, and the average accuracy was 8.39% higher than that of the original YOLOv4, which can provide early warning about the intrusion of foreign objects around transmission lines. Practical applications usually require timely feedback on the detection results, most of the current detection methods sacrifice accuracy to improve speed; thus, to meet the requirements of real-time detection. Moreover, the object detection method based on deep learning requires a large number of input images and high-performance inference equipment, which is obviously a challenge for practical deployment. Due to the automatic feature extraction procedure, the deep learning-based visual trackers have recently achieved a great success in challenging scenarios [22]. Using pre-trained CNN as feature extraction networks (FEN) and fusing deep representations of different CNNs [23] to detect the foreign objects on transmission lines can not only reduce the manual resources, but also have a high detection accuracy in the absence of high-performance equipment.

In order to overcome the problem of insufficient samples and improve the detection performance, this paper proposes a method combining multi-network feature fusion and random forest for detection of foreign objects on transmission lines. A foreign object image dataset composed of balloons, kites, nests and plastic was established. The target region

in the image was extracted by Otus binarization threshold segmentation and morphology processing. Five types of convolutional neural network (CNN) including GoogLeNet, DenseNet-201, EfficientNet-B0, ResNet-101 and AlexNet, were used to extract the features of the target region, and then they were fused by concatenation fusion strategy. The fused features of different schemes were input to train and test the random forest model. The gradient-weighted class activation mapping (Grad-CAM) was applied to offer visual explanations and verify the effectiveness of the optimal scheme. The results show that the detection accuracy reaches 95.88%.

2. Target Region Extraction of Foreign Object Images

Since the performance of the detection model relies on the training result of classifiers, the complexity and diversity of the background in foreign object images obtained by UAV aerial images and video surveillance may affect the feature extraction effect. In order to improve the representativeness of the extracted features, the Otsu adaptive threshold segmentation algorithm [24] and morphological processing were adopted to extract the target region of the foreign object images, so as to remove the complex background and retain the important region with the foreign object. The implementation details and steps are described as follows.

Step 1: Gray the input foreign object images and divide the gray level into N, i.e., (0, 1, 2, . . . , $n-1$). G_i represents the number of pixels with gray level i and the total of gray levels in the image are $G_{sum} = G_0 + G_1 + \ldots + G_{n-1}$;

Step 2: Set the initial threshold T, define the pixels in the $[0, T]$ gray level interval as the target, and define the pixels in the $[T + 1, N - 1]$ gray level interval as the background.

Step 3: Calculate the class variance g between the target and background according to Equation (1) and maximize the class variance by changing the value of T, that is the best threshold T for Otsu adaptive threshold segmentation.

$$g = \omega_0 \omega_1 (\mu_0 - \mu_1)^2 \tag{1}$$

where μ_0 and μ_1 represent the average gray level of target class and background class, respectively; ω_0 and ω_1 represent the proportion of target and background pixels in the whole image and the values can be calculated according to Equation (2).

$$\begin{cases} \omega_0 = \sum\limits_{i=0}^{T} \dfrac{G_i}{G_{\text{sum}}}, & \omega_1 = \sum\limits_{i=T+1}^{N-1} \dfrac{G_i}{G_{\text{sum}}} \\ \mu_0 = \sum\limits_{i=0}^{T} \dfrac{i \times G_i}{\omega_0 \times G_{\text{sum}}}, & \mu_1 = \sum\limits_{i=T+1}^{N-1} \dfrac{i \times G_i}{\omega_0 \times G_{\text{sum}}} \end{cases} \tag{2}$$

Step 4: Morphological open operation is used to process the segmented binary image, which corrodes the image for eliminating the noise and then expands the image to obtain the connected region and colors the different regions.

Step 5: According to the connected region of interest, the centroid and bounding rectangle are obtained, so as to mark the target region. Finally, the target region in the image is cropped and extracted.

Taking a foreign object image of the nest as an example to show the process and effect of target region extraction based on the Otsu adaptive threshold segmentation algorithm, as shown in Figure 1.
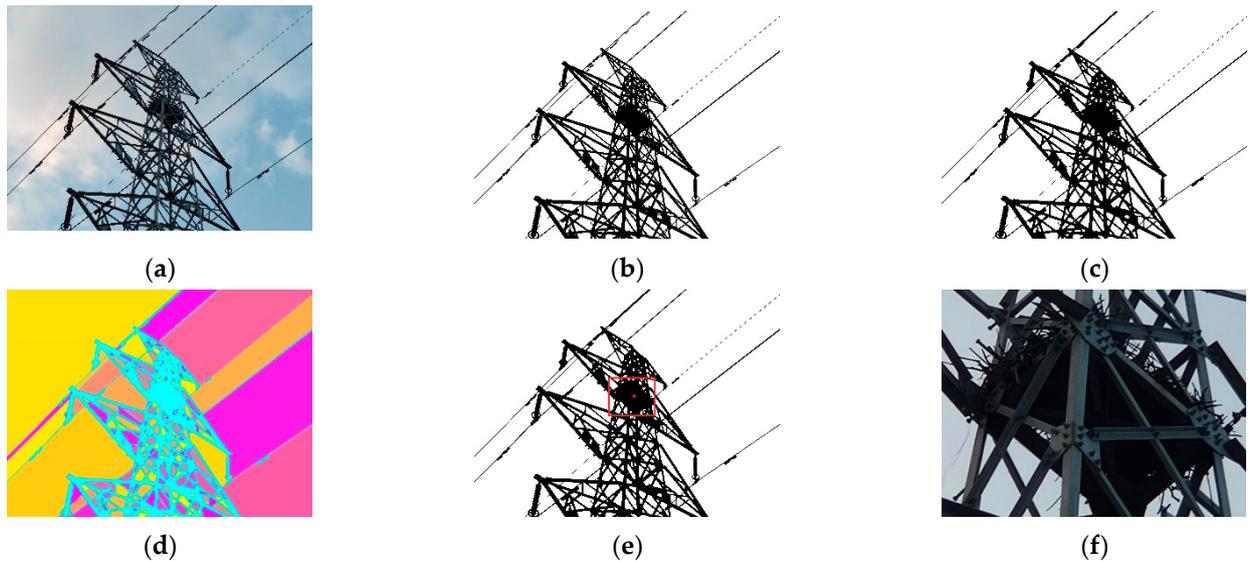
Figure 1. Extraction process of the target region. (**a**) original image, (**b**) binary image, (**c**) open operation, (**d**) rgb image, (**e**) target region, (**f**) processed image.

3. Foreign Objects Detection Algorithm

Compared with the complex feature definition, extraction, and screening processes of traditional detection method, CNN is more convenient and effective to extract the features of foreign object images and has higher detection accuracy. To improve the detection accuracy of the foreign objects on transmission lines, the strategy combining multi-network feature fusion and random forest classifier is proposed. The overall implementation process is shown in Figure 2.
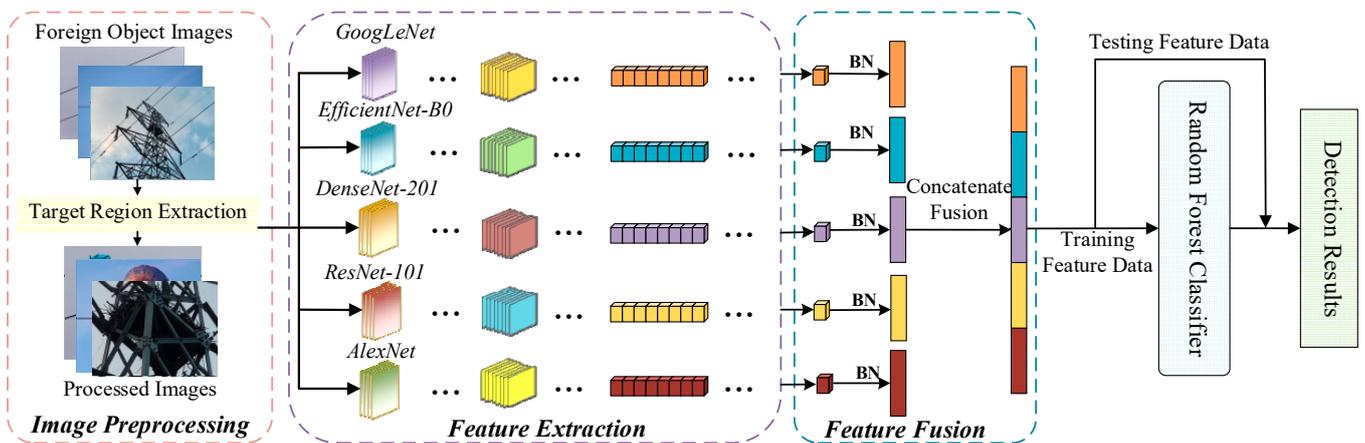
Figure 2. The overall implementation process of the detection method.

3.1. Feature Extraction Network

3.1.1. GoogLeNet

GoogLeNet was proposed by Christian in ILSVRC 2014 competition [25], whose advancement of this network lies in the introduction of the Inception structure, building a sparse network structure and reducing the parameters. GoogLeNet also uses various data augmentation methods to reduce the redundant structure of the network, which solves the problem of overfitting with the small samples and increasing computation as the layers of the network going deeper, so that the depth of GoogLeNet is larger than that of VGG, AlexNet and other networks, but the parameters are much smaller. Therefore,

GoogLeNet has high efficiency and practicability under the condition of the same memory and computing resources.

Inception is the main module proposed by GoogLeNet, whose main idea is to extract multi-scale features of the input images through a variety of convolution kernels with different sizes, and the features are fused to achieve better image representation ability of GoogLeNet. The structure of original Inception module is mainly composed of 1×1 convolution (Conv), 3×3 Conv, 5×5 Conv and 3×3 max pooling, as shown in Figure 3.
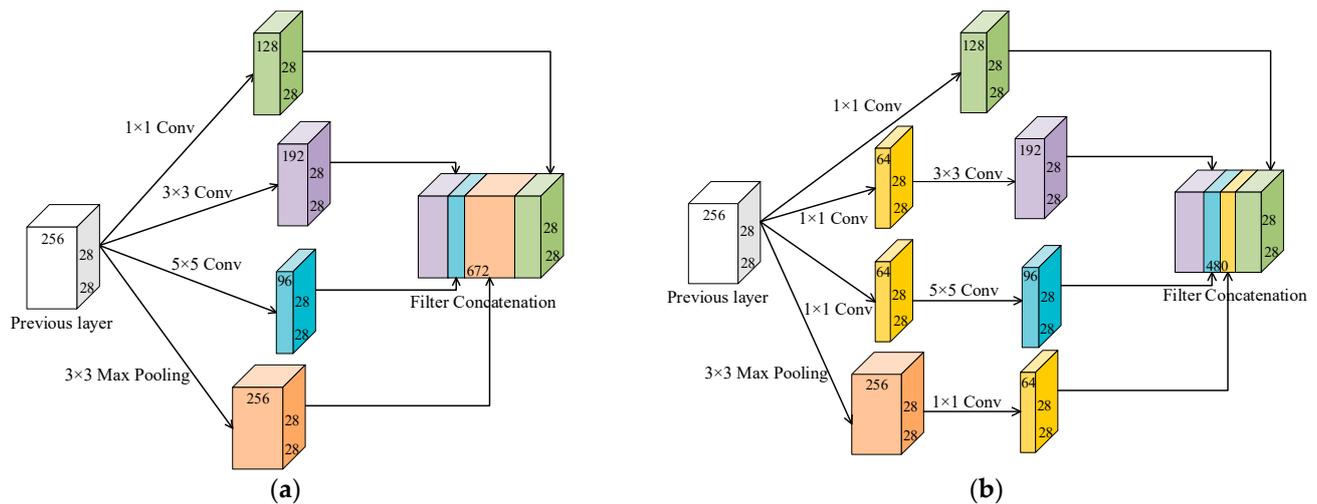
Figure 3. The structure of Inception module in GoogLeNet. (**a**) Original inception module, (**b**) inception module with dimension reduction.

The Inceptionv1 of the GoogLeNet is adopted in this paper which is composed of nine Inception modules. The Inception modules in Inceptionv1 are improved on the basis of the original Inception modules, whose structure is shown in Figure 3b. That is, three 1×1 Convs are added to the original Inception modules to reduce the complexity of the model without losing the feature representation ability of the model. With the introduction of these three 1×1 reduced-dimensional Convs, the number of parameters and computation of the network is greatly reduced.

3.1.2. EfficientNet-B0

EfficientNet is an efficient convolutional neural network proposed by Tan in 2019 [26], which uses a simple and efficient composite coefficient to control the width and depth of the network and the resolution of the input images. Then, the overall model can be scaled through neural network search to determine the optimal composite coefficient, and a series of EfficientNet (B0 to B7) with high accuracy can be obtained by this method. Additionally, with the continuous increase from B0 to B7, the accuracy is growing higher and higher, but the requirements for hardware devices are also growing higher and higher; thus, the EfficientNet-B0 was chosen as the feature extraction network.

The overall structure of EfficientNet-B0 is composed of sixteen mobile inverted bottleneck convolutions (MBConvs), two convolution layers, one average pooling layer and one classification layer, as shown in Table 1. As can be seen, the core structure of the EfficientNet-B0 network is the MBConv module [27], which introduced the squeeze-and-excitation networks (SENet) [28] attention mechanism, as shown in Figure 4. The introduction of MBConv module makes the network have random depth, shortens the training time, and improves the performance of the model. At the same time, the SENet attention mechanism enables the MBConv module to focus on the channel features with the most useful information and suppress the useless features. Therefore, the MBConv module makes EfficientNet-B0 more efficient in feature extraction.

Table 1. The structure of EfficientNet-B0.

Stage	Operation	Size	Class
1	Conv 3×3	224×224	1
2	MBConv1, k3 $\times$ 3	112×112	1
3	MBConv6, k3 $\times$ 3	112×112	2
4	MBConv6, k5 $\times$ 5	56×56	2
5	MBConv6, k3 $\times$ 3	28×28	3
6	MBConv6, k5 $\times$ 5	14×14	3
7	MBConv6, k5 $\times$ 5	14×14	4
8	MBConv6, k3 $\times$ 3	7×7	1
9	Conv 1×1 Pooling and FC	7×7	1
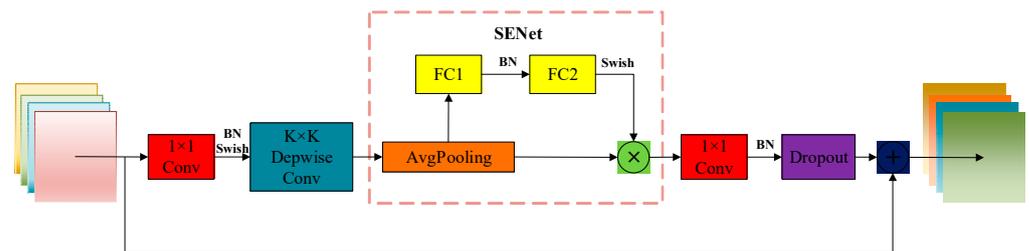
Figure 4. The structure of MBConv with SENet.

3.1.3. DenseNet-201

In order to solve the problems of overfitting and gradient disappearance caused by the complex network structure, DenseNet, a convolutional neural network proposed by Gao et al. [29], proposed a structure of dense connection. The core technology of the DenseNet network is to solve the overfitting problem through the fusion of deep and shallow features, and use cross-connections to alleviate the gradient disappearance caused by too deep network layers.

DenseNet is mainly composed of multiple Dense Blocks and Transition Layers. The structure of the Dense Block is shown in Figure 5. The output features of each layer are integrated with the nonlinear output of all previous layers. For example, the output features of layer l are connected by the features of all previous layers, and the output features of layer l are processed by the nonlinear function as shown in Equation (3).

$$x_l = H_l([x_0, x_1, x_2, ..., x_{l-1}]) \tag{3}$$

where x_l is the output features of layer l; H_l is the nonlinear variation function in front of l layer, which includes the Conv layer, batch normalization layer (BN) and ReLU activation function; $[x_0, x_1, x_2, \ldots, x_{l-1}]$ is the output features of all layers before layer l. In addition, Transition Layers are used for down-sampling at the junction of each Dense Block, which includes the BN layer, ReLU activation function, 1×1 Conv and 2×2 average pooling, so as to avoid the problem of excessive dimension of feature channel caused in the feature fusion and improve the efficiency of the feature transmission indirectly.

At present, there are many versions of DenseNet, including DenseNet-121, DenseNet-169 and DenseNet-201. In this paper, DenseNet-201 with the deepest structure is used as the feature extraction network to achieve higher accuracy, whose structure is shown in Table 2, which contains four Dense Blocks and three Transition Layers. Finally, the depth features of the foreign object images can be extracted through multiple Dense Blocks and Transition Layers.
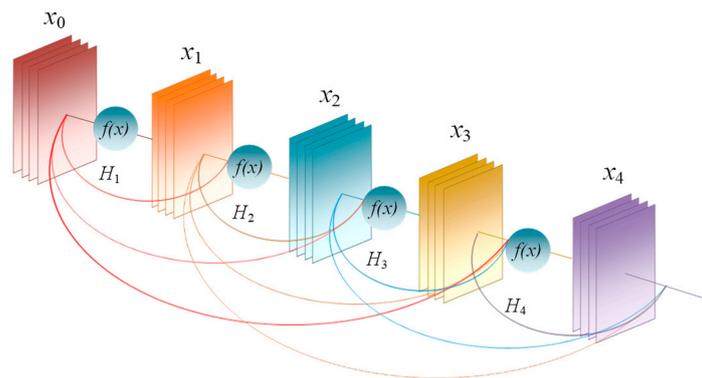
Figure 5. The structure of the Dense Block.

Table 2. The structure of DenseNet-201.

Layer	Size	Operation
Conv	112×112	7×7 conv
pooling	56×56	3×3 maxpool
Dense Block 1	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer 1	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 2 \times 2 \text{ average pool} \end{bmatrix}$
Dense Block 2	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer 2	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 2 \times 2 \text{ average pool} \end{bmatrix}$
Dense Block 3	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Transition Layer 3	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 2 \times 2 \text{ average pool} \end{bmatrix}$
Dense Block 4	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$
Classification Layer	1×1 1000	7×7 global average pool Fully connected, softmax

3.1.4. ResNet-101

ResNet [30] is a convolutional neural network on the basis of residual network. It is able to skip the middle layer, and directly connects the upper layer to the lower layer, which can overcome the gradient disappearance problems caused by the depth of network. The versions of ResNet can be divided into two categories. One is the shallow network based on BasicBlock, such as ResNet-18 and ResNet-34, the other is the deeper network based on Bottleneck, including ResNet-50 and ResNet-101 et al. ResNet-101 has a depth of 101, it is composed of four main layers, and each layer in the network consists of several bottles. The structure of ResNet-101 is shown in Table 3.

Table 3. The structure of ResNet-101.

Layer Name	Output Size	Operation
Conv1	112×112	$7 \times 7, 64$
Conv2_x	56×56	3×3 max pool
Dense Block 1	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$

Table 3. *Cont.*

Layer Name	Output Size	Operation
Conv3_x	28×28	$\begin{bmatrix} 1 \times 1, \ 128 \\ 3 \times 3, \ 128 \\ 1 \times 1, \ 512 \end{bmatrix} \times 4$
Conv4_x	14×14	$\begin{bmatrix} 1 \times 1, \ 256 \\ 3 \times 3, \ 256 \\ 1 \times 1, \ 1024 \end{bmatrix} \times 6$
Conv5_x	7×7	$\begin{bmatrix} 1 \times 1, \ 512 \\ 3 \times 3, \ 512 \\ 1 \times 1, \ 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax

3.1.5. AlexNet

AlexNet [31] is a convolutional neural network that adopts ReLU as the activation function. It has faster convergence. The structure of AlexNet consists of five convolution layers and three fully connected layers. The five convolution layers are composed of 96 convolution kernels with the size of $11 \times 11 \times 3$, 256 convolution kernels with the size of $5 \times 5 \times 48$, 384 convolution kernels with the size of $3 \times 3 \times 256$, 384 convolution kernels with the size of $3 \times 3 \times 192$, and 256 convolution kernels with the size of $3 \times 3 \times 192$, respectively. The three fully connected layers are connected behind the five convolutions layers.

3.2. Multi-Network Feature Fusion

The accuracy and robustness of the deep learning methods rely on a large number of training samples, a small dataset is insufficient to train a deep learning model with high precision and strong robustness [32]. The sizes of the input images will be cut or compressed to meet the requirements of different networks, which may cause the loss or distortion of some image information [33]. Therefore, the features extracted by a single convolutional neural network cannot fully characterize the image information, which may affect the detection accuracy. The fusion of features extracted by different networks is a strategy to improve the accuracy of foreign object detection. The features extracted by different networks express different information of images, so superimposing useful feature vectors of the images can increase the diversity of the features.

Four types of foreign object images collected from UAV aerial images and video surveillance were used to establish the dataset. Due to different pixels of UAV devices, the image resolutions in the dataset are not the same, which are in the range of 640×435 to 2279×2181. Therefore, the target region in the foreign object image was extracted by the Otsu adaptive threshold segmentation algorithm and morphological processing, thus removing the complex background in the image and retaining the region with the foreign object. The size of the input images was transformed into $224 \times 224 \times 3$ to meet the requirements of the employed networks, and they were input to five types of CNN, including GoogLeNet, EfficentNet-B0, DenseNet-201, ResNet-101, and AlexNet, thus extracting the features.

The concatenation fusion strategy is adopted to fuse the features extracted by different networks. The features extracted by GoogLeNet, EfficientNet-B0, DenseNet-201, ResNet-101 and AlexNet are denoted as F_G, F_E, F_D, F_R, F_A, respectively, and then the fused features F are concatenated according to Equation (4), which is shown as follows:

$$F = \text{concatenate}(F_G, F_E, F_D, F_R, F_A) \tag{4}$$

In order to verify the effectiveness of feature fusion and the complementarity of features extracted by different networks, the gradient-weighted class activation mapping (Grad-CAM) [34] was applied to visualize the decision region of each network, which is a technique for producing visual explanations for the decision region of CNN models.

Grad-CAM uses the gradient of classification and the last convolution layer to produce a coarse localization map that highlights the decision region of the image.

In order to verify the complementarity of features extracted by different networks, the saliency maps of five networks obtained by Grad-CAM were used to visualize the decision regions of each network. Taking an image of a kite hanging on transmission lines as an example, the visualization results are shown in Figure 6, where the deeper color corresponds to the area that the selected network pays more attention to. Different CNN has diverse decision region, it can be seen from Figure 6 that the coverage of the decision region of the proposed method is larger than that of a single convolutional neural network, which means that the extracted features contain more image information.
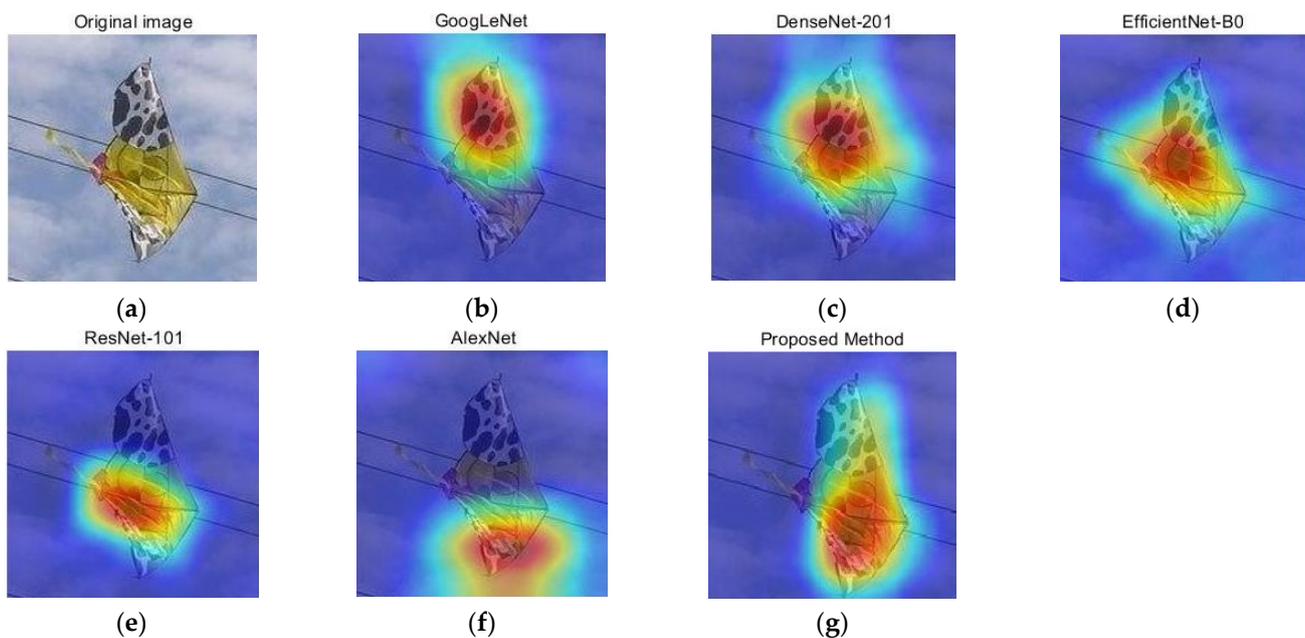
Figure 6. Saliency maps of different convolutional neural networks. (**a**) Original image, (**b**) GoogLeNet, (**c**) DenseNet-201, (**d**) EfficientNet-B0, (**e**) ResNet-101, (**f**) AlexNet, (**g**) proposed method.

3.3. Random Forest Classifier

Random forest is a machine learning classification model proposed by Ho in 1995 [35], which is mainly composed of multiple decision trees (DT) based on the idea of Bagging and is often used in classification and recognition. In general, DT is often used in classification to achieve fast classification performance, but a single DT has the problems of weak generalization and low classification accuracy. However, random forest improves the generalization ability by combining multiple DT, so as to avoid the overfitting of DT algorithm and improve the detection accuracy.

The classification process of the random forest classification model is shown in Figure 7, which can be concluded as the following steps: (1) random forest adopts Bootstrapping re-sampling technology to select samples from the random input training features, which were used to train DTs; (2) these well-trained DTs are combined to form random forest classifier, and each sub-DT will classify the testing samples and give the classification results; (3) the final result is determined by majority voting.
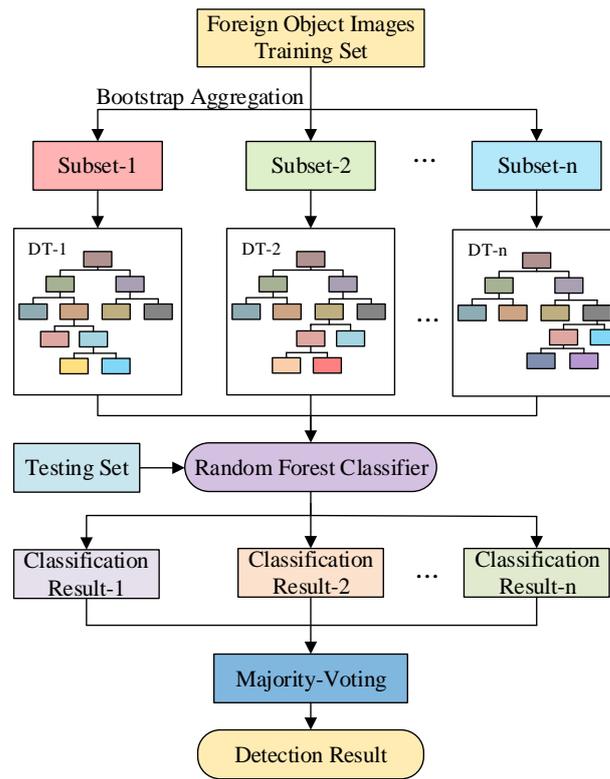
Figure 7. The schematic diagram of random forest.

4. Simulation and Analysis of Transmission Lines Foreign Object Detection

4.1. Simulation Environment and Evaluation Indexes

The experimental configuration of foreign object detection consists of the software environment of MATLAB 2021a and the hardware equipment with CPU of AMD (A) Ryzen 5600H and GPU of NVIDIA GeForce RTX 3050Ti with 4GB video memory, memory 16 GB and the Windows10 operation system.

The confusion matrix was introduced to analyze the experimental results, which was obtained by the classification of the random forest. Meanwhile, in order to evaluate the performance of the proposed method, the commonly used indexes: precision (P_i), recall (R_i) and accuracy were introduced, where i = (1, 2, 3, 4) represents different categories of the foreign objects. The specific definitions of the indexes are

$$P_i = \frac{TP}{TP + FP} \times 100\%, \quad R_i = \frac{TP}{TP + FN} \times 100\% \tag{5}$$

$$Accuracy = \frac{TP}{TP + FP + FN + TN} \times 100\% \tag{6}$$

where *TP* (true positive) represents the number of positive samples correctly predicted by the random forest; *TP* + *FP* represent the number of positive samples predicted by the random forest. *TP* + *FN* represent the number of actual positive samples. *TP* + *TN* + *FP* + *FN* represent the total number of samples.

In addition, the *macro_F*$_1$ also was introduced to evaluate the performance of the foreign object detection model, which is commonly used for multiple classification problems and suits the multi-classification problems with unbalanced samples. The calculation equation is

$$macro\_P = \frac{1}{m}\sum_{i=1}^{m} P_i \times 100\%, \quad macro\_R = \frac{1}{m}\sum_{i=1}^{m} R_i \times 100\% \tag{7}$$

$$macro\_F_1 = \frac{2 \times macro\_P \times macro\_R}{macro\_P + macro\_R} \times 100\% \tag{8}$$

where m equals four, i.e., the total number of foreign object categories. P_i is the precision of each category. R_i is the recall of each category.

4.2. Simulation Results and Analysis

4.2.1. Results under Different Foreign Object Samples

The foreign object image dataset of balloons, kites, nests and plastic was constructed through transmission line inspection images and public images. To verify the effectiveness of target region extraction and overcome the problem caused by small samples, three types of samples including samples A, B and C were prepared. Sample A was the original images collected by inspection equipment and public images dataset; Sample B was the target region images extracted from the images in Sample A by the image preprocessing; Sample C added synthetic images on the basis of Sample B, the synthetic images blended the foreign object target with the transmission lines by Photoshop, and each type of foreign object images adds an additional nearly 50% of its original image number. The details of three types of samples are shown in Table 4.

Table 4. The distribution of foreign object samples.

Sample	Type of Foreign Object				Total
	Balloons	Kites	Nests	Plastic	
Sample A	40	67	95	44	246
Sample B	40	67	95	44	246
Sample C	60	100	143	66	369

In order to verify the effect of image preprocessing and artificial synthetic images on improving the accuracy, the prepared samples were randomly divided into training set and testing set according to the proportion of 7:3. Then, the divided samples were transformed into the size of 224 × 224 × 3 and input to the selected networks such as GoogLeNet, EfficientNet-B0, DenseNet-201, ResNet-101 and AlexNet for feature extraction and the extracted features were used to train and test the random forest classifier. The detection results of three kinds of samples are compared and analyzed, as shown in Figure 8.
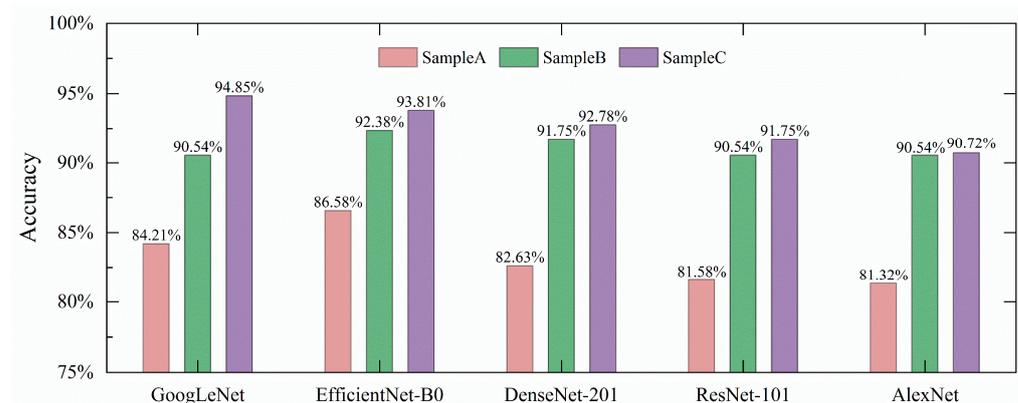
Figure 8. Detection accuracy of different samples based on different networks.

According to the detection accuracy in Figure 8, the accuracy of each network adopting Sample B was much higher than that of Sample A, which can be concluded that the image of target region extracted from original images had reduced the influence of complex background factors in the inspection images and effectively improved the accuracy. Moreover, according to the comparative analysis of samples B and C, it can be seen that

adding synthetic images can further improve the fitting effect and generalization ability of random forest classifier, so as to deal with the actual complex background of the detection.

4.2.2. Results under Different Feature Extraction Layers

In order to obtain the optimal fused features, it is necessary to determine the feature extraction layer of each network. The features were extracted from different layers of five types of CNN between shallow layer and deep layer and then sent to the random forest for training and testing. The following experimental analysis takes sample C as the input samples, and the training set and testing set were divided according to the proportion of 7:3.

In the simulation experiment, four max pooling layers and the deepest layer (global average pooling layer) of the GoogLeNet were selected as the feature extraction layers, a total of five groups. The global average pooling layers behind each MBconv module in the EfficientNet-B0 were selected as the feature extraction layer, with a total of ten groups. Four groups of features behind each Dense Block and Transition Layer module and six groups of features from shallow layers to deep layers in the DenseNet-201 were selected as the feature extraction layer, with a total of ten groups. Three pooling layers and two drop layers in the AlexNet were selected as the feature extraction layer, a total of five groups. Two pooling layers and eight ReLU activation function layers in the transition layers of Residual Blocks in ResNet-101 were selected as the feature extraction layers, a total of ten groups. Finally, these features were used to train and test the random forest classifier and the accuracy of different networks' layers can be shown in Figure 9.
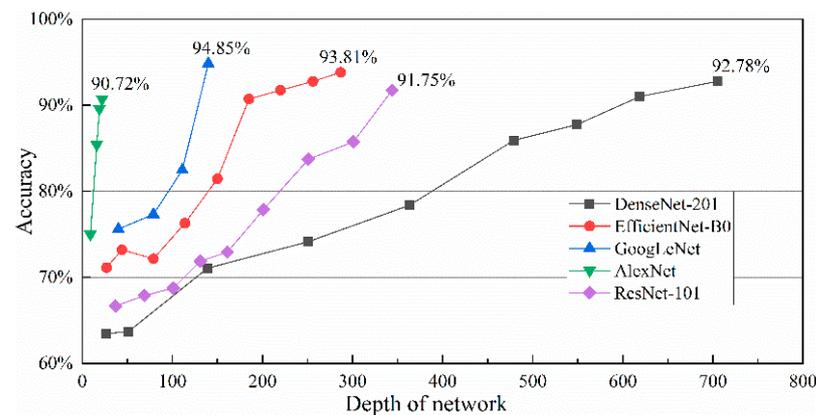
Figure 9. The accuracy under different feature extraction layers of different networks.

It can be found from Figure 9 that with the deepening of feature extraction layer, the detection accuracy of random forest classifier is higher; thus, the deep feature extraction layer is adopted generally. Therefore, GoogLeNet, EfficientNet, and DenseNet-201, respectively took Layer-140, Layer-287, and Layer-705 global average pooling layer as the feature extraction layer. Both AlexNet and ResNet101 took the previous layer of the fully connected layer as the feature extraction layer. The selected layers were used in the following experiments.

4.2.3. Comparison of Different Feature Extraction Networks

In order to determine the CNN model for feature extraction, experiments on five networks including DenseNet-201, GoogLeNet, AlexNet, ResNet-101, and EfficientNet-B0 were conducted to compare their learning performance. After the simulation experiment, the performance indexes including the accuracy, *macro_P*, *macro_R* and *macro_F*$_1$ of the above models are obtained, as shown in Table 5.

Table 5. The performance indexes of different network models.

Network	Indexes/%			
	Accuracy	*macro_P*	*macro_R*	*macro_F$_1$*
DenseNet-201	92.78	92.22	91.21	91.71
GoogLeNet	94.85	94.83	93.79	94.31
AlexNet	90.72	91.50	88.97	90.22
ResNet-101	91.75	92.33	89.69	90.99
EfficientNet-B0	93.81	94.81	92.41	93.59

It can be seen from Table 5 that GoogLeNet has the highest accuracy and *macro_F$_1$*, which are 94.84% and 94.31%. The accuracy of EfficientNet-B0 is 93.81% and the *macro_F$_1$* is 93.59%, which is the second place behind GoogLeNet. The network with the lowest performance is AlexNet, with the accuracy of 90.72% and the *macro_F$_1$* of 90.22%. The above results are sufficient to show that GoogLeNet is the optimal network in feature extraction, followed by EffecentNet-B0 and DenseNet-201.

4.2.4. Comparison and Analysis of Different Feature Fusion Schemes

In order to obtain the optimal feature fusion scheme and verify the effectiveness of multi-network feature fusion in actual simulation, the networks were sorted according to the detection performance obtained by experiments in Section 4.2.3 and named them as N1 to N5, namely GoogLeNet, EfficientNet-B0, DenseNet-201, ResNet-101 and AlexNet. Then, the features extracted from different networks were concatenated separately and then sent to train and test the random forest classifier. Finally, the optimal feature fusion scheme was determined for the transmission line foreign object detection based on the results under different fusion schemes, as shown in Table 6.

Table 6. The results under different fusion schemes.

Baseline Network					Indexes/%			
N1	N2	N3	N4	N5	Accuracy	*macro_P*	*macro_R*	*macro_F$_1$*
√	√				93.81	92.52	94.31	93.41
√	√	√			95.88	96.33	95.19	95.76
√	√	√	√		92.78	95.40	90.43	92.85
√	√	√	√	√	93.81	95.02	92.39	93.69

From the results of the above fusion schemes, it can be concluded that the accuracy can obtain the optimal 95.88% by fusing the features of GoogLeNet, EfficientNet-B0, DenseNet-201. The reason why the other schemes are not superior to the feature fusion performance of the above three networks is that the features extracted from the other two networks do not contain enough information or redundant features resulting in lower accuracy after fusion. Therefore, the concatenation fusion of features extracted from GoogLeNet, Efficientnet-B0 and DenSenet-201 can further improve the detection accuracy of foreign objects on transmission lines.

5. Discussion

The methods for foreign objects detection on transmission lines can be classified into two categories. One is the traditional detection models, such as the method combining GMM and K-means clustering in [9] and that based on combination features and cascade classifier in [11], etc. These methods should extract artificially designed features from the images, which were greatly affected by image processing results and the complex backgrounds in the images. Therefore, the traditional detection models are usually not suitable for practical engineering applications. The proposed method in this paper adopts the Otus binarization threshold segmentation and morphology processing to extract the

target region in the images, thus eliminating the influence of the complex backgrounds on the image feature extraction. Meanwhile, the CNNs are used to extract the features automatically, which is able to overcome the shortcomings of manual feature extraction.

The other is the detection models based on deep learning algorithms, such as SSD [13], Faster R-CNN [15], improved YOLOv3 [19] and YOLOv4 [21], etc. These methods require high-performance equipment to train the detection models, which are usually time-consuming and need a large number of training samples. The proposed method uses five pre-trained networks to extract the image features, and adopts the random forest to establish a machine learning model, thus achieving classification of foreign objects. This strategy is useful to overcome the disadvantages of object detection algorithms based on deep learnings.

The detection accuracy was taken as an evaluation index to compare the proposed method with other foreign object detection methods in previous studies. The detection accuracies of these methods are summarized in Table 7. It can be seen that the proposed method has a higher accuracy compared with most of the published foreign object detection methods, only a little bit lower than that based on combination features and cascade classifier. The latter should extract the features manually, and only verified by bird's nest detection cases, whereas the proposed method can achieve satisfied accuracy with automatic feature extraction. The comparison results verify the superiority of the proposed method based on multi-network feature fusion and random forest, and it is useful for foreign objects detection on transmission lines.

Table 7. Comparison of different detection methods.

Detection Method	Accuracy (%)
GMM + K-means clustering [9]	78.00
Combination features + cascade classifier [11]	97.33
SSD [13]	85.20
Faster R-CNN [15]	93.09
FODN4PS Net [16]	85.00
RetinaNet [18]	94.10
Improved YOLOv3 [19]	92.60
YOLOv4 [21]	81.72
The proposed method	95.88

6. Conclusions

In order to detect the foreign objects suspended on transmission lines, this paper proposes a method combining multi-network feature fusion and random forest. A case is carried out on foreign object detection and the influences of different samples, feature extraction layers, CNN models and feature fusion schemes are compared and analyzed. The conclusions are drawn as follows:

1. Target region extraction based on Otus binarization threshold segmentation and morphology processing can reduce the influence and distraction of complex background factors in the inspection images and improve the detection accuracy.
2. The optimal feature fusion scheme by fusing the features extracted by GoogLeNet, EfficientNet-B0 and DenseNet-201 can achieve the highest accuracy of 95.88%. It can be seen through Grad-CAM that the fused features can compensate for the differences of different networks and fully reflect the features of the foreign object images, thus improving the generalization ability of the detection model.

The proposed detection method combining multi-network feature fusion and random forest can realize accurate detection of foreign objects on transmission lines, which can avoid problems such as insufficient samples, low performance of equipment and long time consuming, and assist transmission line inspection personnel to detect and remove foreign objects suspended on transmission lines.

References

1. Zhang, W.; Liu, X.; Yuan, J.; Xu, L.; Sun, H.; Zhou, J.; Liu, X. RCNN-based foreign object detection for securing power transmission lines (RCNN4SPTL). *Procedia Comput. Sci.* **2019**, *147*, 331–337. [CrossRef]
2. Chen, W.; Li, Y.; Li, C. A Visual detection method for foreign objects in power lines based on Mask R-CNN. *Int. J. Ambient. Comput. Intell.* **2020**, *11*, 34–47. [CrossRef]
3. Pei, X.; Kang, Y. Short-Circuit fault protection strategy for High-Power Three-Phase Three-Wire inverter. *IEEE Trans. Industr. Inform.* **2012**, *8*, 545–553. [CrossRef]
4. Qin, X.; Wu, G.; Lei, J.; Fei, F.; Ye, Y.; Mei, Q. A novel method of autonomous inspection for transmission line based on cable inspection robot LiDAR data. *Sensors* **2018**, *18*, 596. [CrossRef]
5. Xia, P.; Yin, J.; He, J.; Gu, L.; Yang, K. Neural detection of foreign objects for transmission lines in power systems. *J. Phys. Conf. Ser.* **2019**, *1267*, 012043. [CrossRef]
6. Li, J.; Yan, D.; Luan, K.; Li, Z.; Liang, H. Deep learning-based bird's nest detection on transmission lines using UAV imagery. *Appl. Sci.* **2020**, *10*, 6147. [CrossRef]
7. Liu, Y.; Huo, H.; Fang, J.; Mai, J.; Zhang, S. UAV Transmission line inspection object recognition based on Mask R-CNN. *J. Phys. Conf. Ser.* **2019**, *1345*, 062043. [CrossRef]
8. Guo, S.; Bai, Q.; Zhou, X. Foreign object detection of transmission lines based on Faster R-CNN. In *Information Science and Applications*; Springer: Berlin, Germany, 2019; pp. 269–275. [CrossRef]
9. Yao, N.; Zhu, L. A novel foreign object detection algorithm based on GMM and K-means for power transmission line Inspection. *J. Phys. Conf. Ser.* **2020**, *1607*, 012014. [CrossRef]
10. Taskeed, J.; Md, Z.U. Rotation invariant power line insulator detection using local directional pattern and support vector machine. In Proceedings of the 2016 International Conference on Innovations in Science, Engineering and Technology, Dhaka, Bangladesh, 28–29 October 2016; pp. 1–4. [CrossRef]
11. Lu, J.; Xu, X.; Li, X.; Li, L.; Chang, C.; Feng, X.; Zhang, S. Detection of bird's nest in high power lines in the vicinity of remote campus based on combination features and cascade classifier. *IEEE Access* **2018**, *6*, 39063–39071. [CrossRef]
12. Huang, X.; Shang, E.; Xue, J.; Ding, H.; Li, P. A multi-feature fusion-based deep learning for insulator image identification and fault detection. In Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chongqing, China, 12–14 June 2020; pp. 1957–1960. [CrossRef]
13. Wang, B.; Wu, R.; Zheng, Z.; Zhang, W.; Guo, J. Study on the method of transmission line foreign body detection based on deep learning. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–5. [CrossRef]
14. Ni, H.; Wang, M.; Zhao, L. An improved Faster R-CNN for defect recognition of key components of transmission line. *Math. Biosci. Eng.* **2021**, *18*, 4679–4695. [CrossRef]
15. Zu, G.; Wu, G.; Zhang, C.; He, X. Detection of common foreign objects on power grid lines based on Faster R-CNN algorithm and data augmentation method. *J. Phys. Conf. Ser.* **2021**, *1746*, 012039. [CrossRef]
16. Xu, L.; Song, Y.; Zhang, W.; An, Y.; Wang, Y.; Ning, H. An efficient foreign objects detection network for power substation. *Image Vis. Comput.* **2021**, *109*, 104159. [CrossRef]
17. Zhai, Y.; Yang, X.; Wang, Q.; Zhao, Z.; Zhao, W. Hybrid knowledge R-CNN for transmission line multifitting detection. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5013312. [CrossRef]
18. Zhang, H.; Zhao, J.; Chen, Y.; Jiang, S.; Wang, D.; Du, H. Intelligent bird's nest hazard detection of transmission line based on RetinaNet model. *J. Phys. Conf. Ser.* **2021**, *2005*, 012235. [CrossRef]
19. Li, J.; Nie, Y.; Cui, W.; Liu, R.; Zheng, Z. Power transmission line foreign object detection based on improved YOLOv3 and deployed to the chip. In Proceedings of the MLMI'20: 2020 the 3rd International Conference on Machine Learning and Machine Intelligence, Hangzhou, China, 18–20 September 2020; pp. 100–104. [CrossRef]

20. Chen, Z.; Xiao, Y.; Zhou, Y.; Li, Z.; Liu, Y. Insulator recognition method for distribution network overhead transmission lines based on modified YOLOv3. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020; pp. 2815–2820. [CrossRef]

21. Song, Y.; Zhou, Z.; Li, Q.; Chen, Y.; Xiang, P.; Yu, Q.; Zhang, L.; Lu, Y. Intrusion detection of foreign objects in high-voltage lines based on YOLOv4. In Proceedings of the 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 9–11 April 2021; pp. 1295–1300. [CrossRef]

22. Marvasti-Zadeh, S.M.; Ghanei-Yakhdan, H.; Kasaei, S. Adaptive exploitation of pre-trained deep convolutional neural networks for robust visual tracking. *Multimed. Tools Appl.* **2021**, *80*, 22027–22076. [CrossRef]

23. Marvasti-Zadeh, S.M.; Ghanei-Yakhdan, H.; Kasaei, S.; Nasrollahi, K.; Moeslund, T.B. Effective fusion of deep multitasking representations for robust visual tracking. *Vis. Comput.* **2021**, 1–21. [CrossRef]

24. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [CrossRef]

25. Christian, S.; Liu, W.; Jia, Y.; Pierre, S.; Scott, R.; Dragomir, A.; Dumitru, E.; Vincent, V.; Andrew, R. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [CrossRef]

26. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946.

27. Mark, S.; Andrew, H.; Zhu, M.; Andrey, Z.; Chen, L. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520. [CrossRef]

28. Hu, J.; Shen, L.; Samuel, A.; Sun, G.; Wu, E. Squeeze-and-Excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]

29. Huang, G.; Liu, Z.; Laurens, V.D.M.; Kilian, Q.W. Densely connected convolutional networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269. [CrossRef]

30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]

31. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]

32. Wang, Y.; Lei, B.; Elazab, A.; Tan, E.; Wang, W. Breast cancer image classification via multi-network features and dual-network orthogonal low-rank learning. *IEEE Access* **2020**, *8*, 27779–27792. [CrossRef]

33. Chen, B.; Yuan, D.; Liu, C.; Qian, W. Loop closure detection based on multi-scale deep feature fusion. *Appl. Sci.* **2019**, *9*, 1120. [CrossRef]

34. Ramprasaath, R.S.; Michael, C.; Abhishek, D.; Remakrishna, V.; Devi, P.; Dhruv, B. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626. [CrossRef]

35. Tin, K.H. Random decision forests. In Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, Canada, 14–16 August 1995; pp. 278–282. [CrossRef]