energies

MDPI

Article

An Improved U-Net Segmentation Model That Integrates a Dual Attention Mechanism and a Residual Network for Transformer Oil Leakage Detection

Xuxu Li [1], Xiaojiang Liu [2], Yun Xiao [1], Yao Zhang [3], Xiaomei Yang [4] and Wenhai Zhang [4,*]

[1] State Grid Sichuan Electric Power Company, Chengdu 610041, China; lixuxu1981@126.com (X.L.); xiaoyun15881295007@163.com (Y.X.)
[2] State Grid Sichuan Electric Power Institute, Chengdu 610041, China; xjsccd@yeah.net
[3] State Grid Sichuan Ultra High Voltage Company, Chengdu 610041, China; zhangyao_cgy@163.com
[4] College of Electrical Engineering, Sichuan University, Chengdu 610065, China; yangxiaomei@scu.edu.cn
* Correspondence: zhangwh2079@scu.edu.cn

Abstract: Accurately detecting oil leakage from a power transformer is important to maintain its normal operation. Deep learning (DL) methods have achieved satisfactory performance in automatic oil detection, but challenges remain due to the small amount of training data and oil targets with large variations in position, shape, and scale. To manage these issues, we propose a dual attention residual U-net (DAttRes-Unet) within a U-net architecture that extensively uses a residual network as well as spatial and channel-wise attention modules. To overcome the vanishing gradient problem due to deeper layers and a small amount of training data, a residual module from ResNet18 is used to construct the encoder path in the U-net framework. Meanwhile, to overcome the issue of training difficulty for the network, inspired by the advantage of transfer learning, initial network parameters in the encoder are obtained from the pre-trained ResNet18 on the ImageNet dataset. Further, in the decoder path, spatial attention and channel attention are integrated to highlight oil-stained regions while suppressing the background or irrelevant parts/channels. To facilitate the acquisition of the fluorescence images of the transformer, we designed a portable acquisition device integrating an ultraviolet light source and a digital camera. The proposed network is trained on the amount of fluorescence images after data augmentation is used and tested on actual fluorescence images. The experiment results show that the proposed DAttRes-Unet network can recognize oil-stained regions with a high accuracy of 98.49% for various shapes and scales of oil leakage.

Keywords: residual block; spatial attention; channel-wise attention; U-net segmentation; oil leakage detection

Citation: Li, X.; Liu, X.; Xiao, Y.; Zhang, Y.; Yang, X.; Zhang, W. An Improved U-Net Segmentation Model Integrating Dual Attention Mechanism and Residual Network for the Transformer Oil Leakage Detection. *Energies* **2022**, *15*, 4238. https://doi.org/10.3390/en15124238

Academic Editor: Dimitris Drikakis

Received: 6 May 2022
Accepted: 7 June 2022
Published: 9 June 2022

1. Introduction

Mineral oil is commonly used as an insulator and coolant for power transformers, which are critical pieces of equipment in power substations. Mineral oil leaks need to be detected effectively and/or quickly. A reduced amount of mineral oil reduces the insulation strength of the power transformer and can even lead to breakdown, short circuits, fires, explosions, and other faults [1,2]. Thus, detecting the leaked oil of a power transformer has become a routine task to ensure the safe operation of a substation [3,4].

The traditional method of detecting oil leakages from a power transformer commonly relies on human operators visually inspecting transformers in a prearranged inspection cycle, which might result in untimely detection and inefficient troubleshooting. With the wide application of inspection robots and video surveillance in substations, many surveillance images can be captured quickly, which helps to detect oil leakages early and maintain the normal operation of power transformers [5–7]. However, recognizing an oil leak from many captured images is difficult. Manual visual inspection yields poor performance because it

requires time and effort by skilled technicians. Thus, it is important to automatically detect transformer oil leakage from surveillance images.

With the development and application of machine vision technology, image recognition methods have been used to detect transformer oil leaks. Commonly, oil leaks should be detected as quickly as possible because early detection helps power operators take action to repair the transformer as soon as possible. While the leaked oil appears yellow and transparent under natural light, it is difficult to automatically recognize it using image processing. Thus, to effectively detect oil leakages, most researchers use the fluorescence features of mineral oil when the oil is irradiated with ultraviolet (UV) light [8,9]. Using fluorescence imagery, image processing is typically used to recognize oil leaks [3]. Based on change detection and greyscale histogram double Gaussian fit analysis, current transformer oil leakages have been detected [10]. However, traditional image processing methods have limitations, such as a low accuracy of images with complex backgrounds and detection thresholds that are difficult to adaptively set according to varied illumination.

In recent years, with the rapid development of deep learning (DL) in the computer vision field, DL has provided a new idea for the detection of oil leakage. Because DL can automatically learn and extract higher-level latent features from complex image backgrounds by constructing certain deep neural networks, DL methods have been widely used in object detection, such as pedestrian detection, vehicle detection, and medical imaging and tumour detection [11–13]. The DL-based methods of object detection can mainly be categorized as two-stage or single-stage. The two-stage detectors first generate region proposals and then identify whether objects exist in each potential region or not. The representative methods of two-stage detectors include the region-based convolutional neural network (R-CNN) [14], Fast R-CNN [15], and Faster R-CNN [16]. A long detection time is the disadvantage of the two-stage detectors, since they operate in two stages. Contrarily, single-stage detectors implement detection once, i.e., achieving the task of categories and locations directly within one stage. The representative methods of single-stage detectors include the Single Shot Detector (SSD) [17] and the series of YOLO (You Only Look Once) [18], e.g., YOLOv1∼YOLOv4 [19–21].

Although there are many DL networks for object detection, recently, a few networks, e.g., the U-net network [4,22], YOLO [23], and LSTM combined with a Genetic Algorithm [24], have been applied to detect the leaked mineral oil of transformers. Among these networks, the U-net network has achieved the highest accuracy of leaked oil detection in the experimental research. Additionally, U-net networks can be used to segment the target object (i.e., oil), which is conducive to evaluating the level of oil leakage, providing a basis for additional research. However, when U-Net is used in a practical scenario, we face the following issues. First, generally, the detection performance of DL methods relies on a large amount of labeled training datasets and constructed network models. However, in practice, the probability of transformer oil leakage is low (i.e., the amount of surveillance images containing insulating oil leakage is limited). Second, due to the various environments of installing transformers and the fluidity of insulating oil, the areas stained by leaked oil exhibit a variety of shapes in the images. These issues make it difficult for conventional U-net networks to completely focus on the characteristics of oil pixels, making them misjudge non-oil portions of an image as leaked oil or miss oil-stained areas, which leads to a low detection accuracy.

To accurately detect leaked oil and further evaluate the seriousness of stained oil in the near future, we propose a dual attention residual U-net (DAttRes-Unet) network that includes an embedded residual model and dual attention modules from both spatial and channel-wise perspectives in the existing U-Net architecture. The primary contributions of this study are as follows: (1) integrating spatial attention and channel attention in the decoder path is conducive to recalibrating the feature responses towards the most informative and important oil component; (2) a residual learning mechanism is introduced in the standard U-net framework, and the residual module is used in the encoder path of the proposed DAttRes-Unet network to overcome the problems of gradient vanishing,

partly caused by the small amount of training data; (3) a transfer learning strategy is used, and ResNet18 weights obtained by pretraining on the ImageNet dataset are used as the initial weights for training the DAttRes-Unet network to improve the performance of the proposed DAttRes-Unet in terms of the amount limitation of the training dataset. Experimental results with fluorescence images of transformers captured in field substations show that the proposed DAttRes-Unet can detect and segment leaked oil with better performance than other improved U-Nets (VGG16-Unet [25] and Res18-Unet) and the variant of DAttRes-Unet with only one attention module.

The remainder of this paper is organized as follows. Section 1 briefly introduces the framework of the classical U-net. Section 3 describes the proposed method, including the design of the device that acquires relevant fluorescent images and the proposed DAttResU-Net architecture with a residual block and two attention modules. Section 4 describes the experiments and results of this study, and conclusions are given in Section 5.

2. Basic Principles of the U-Net Network

In this section, we briefly introduce the classical U-net network [22]. Its structure is shown in Figure 1. The topology of the U-net is symmetrical, mainly consists of the encoder on the left and the decoder on the right. The encoder is the backbone network to extract abstracting features and information, composed of four basic convolutional units. Each basic convolutional unit consists of a 3×3 traditional convolutional layer, a batch normalization layer, and a Rectified Linear Unit (ReLU) activation layer, which is then followed by a pooling layer for downsampling. Thus, the encoder uses downsampling to gradually increase the image depth to implement pixel classification, while the decoder uses upsampling to restore the image size and implement pixel positioning.

The decoder is the feature enhancement network that precisely positions and gradually restores the original size of the image, consisting of four upsampling layers followed by a ReLU and four skip connection layers. The upsampling layer expands the deep feature map as a larger feature. With the downsampling performed in the encoder four times, symmetrically, the upsampling is performed four times in the decoder, in order to restore the feature map extracted in the encoder into images with the same size of the original images. The skip connection layer is used to fuse the different levels of features from the encoder and the decoder. This operation is conducive to supplementing the missing pixel position information during the downsampling process and improving the accuracy of the segmentation. Finally, a softmax classifier is utilized for pixel-by-pixel classification to achieve semantic segmentation.

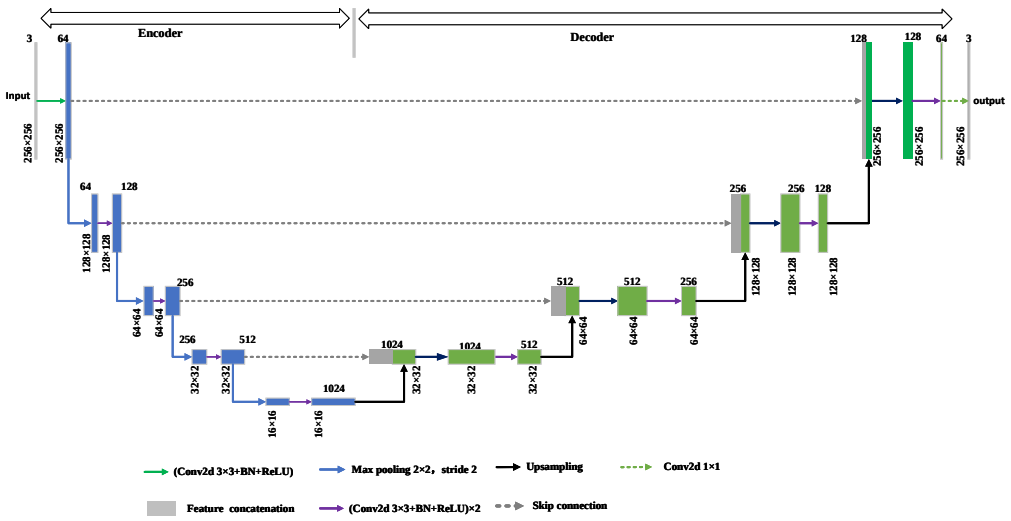

Figure 1. Structure of the classical U-net network.

3. The Proposed Method

Although the classical U-net has achieved high accuracy in segmentation, stacking multiple traditional convolutional layers causes the network to suffer from the problem of gradient vanishing, which impedes the optimization of the network weights. To overcome this problem, four residual blocks replace the original basic convolutional unit in the encoder to facilitate the training of the deep networks. Additionally, considering that the downsampling operation in the encoder would cause information loss, dual attention modules, i.e., the attention gate (AE) and squeeze-and-excitation (SE) blocks, are integrated into the decoder, in order to enhance the richer low-and high-level information and provide a higher weight coefficient for more relevant channels. For these purposes, the DAttRes-Unet network is proposed in this work.

Figure 2 shows the procedure of detecting leaked oil from a transformer, including the acquisition device for fluorescence images, data processing, the proposed DAttRes-Unet network, and the training and testing of the proposed networks, described below.

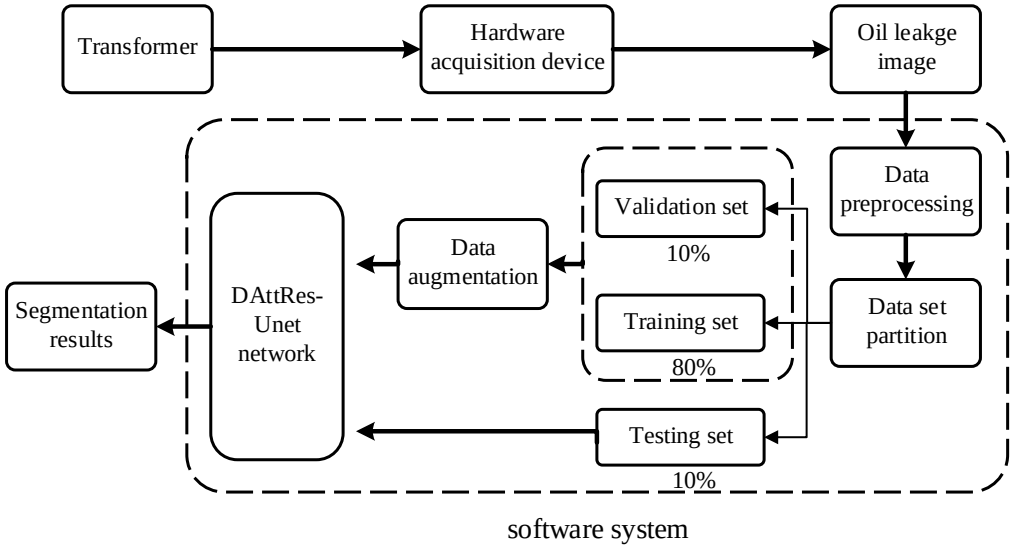

Figure 2. Procedure of detecting leaked oil from a transformer through fluorescent images.

3.1. Image Acquisition Device

To facilitate the acquisition of fluorescent images, we designed a portable acquisition device that integrated a light source that emits UV light and a digital camera. The appearance of the designed device is shown in Figure 3a. The rated power of the UV light source is 15 W, which can produce approximately 5000 $\mu W/cm^2$ UV light at most. Different light levels with a lithium battery as the supply can be adjusted according to practical situations. In our experiments, the focus of UV light source was fixed, and the rated wavelength was designed for 365 nm UV light. The cooperative operation of the light source and the digital camera was implemented through a control unit, and its corresponding simplified circuit is shown in Figure 3b, where the symbols "S" and "KT" denote the primary switch and time relay, respectively; the symbols "L" and "Camera" denote the UV light source and digital camera, respectively.

To collect oil-stained images, the primary switch "S" was first turned on to concurrently provide power to the UV lamp and the coil of the time relay. To guarantee that the UV light source provided stable UV light, the camera was automatically turned on after the timing in the time relay reached a predetermined delay (10 s in this study). The researchers then operated the camera to acquire fluorescent images under UV light. After multiple fluorescent images were collected, the acquired images were stored on a computer via a USB interface and processed by the constructed DAttRes-Unet network to automatically recognize the transformer oil leakage area.

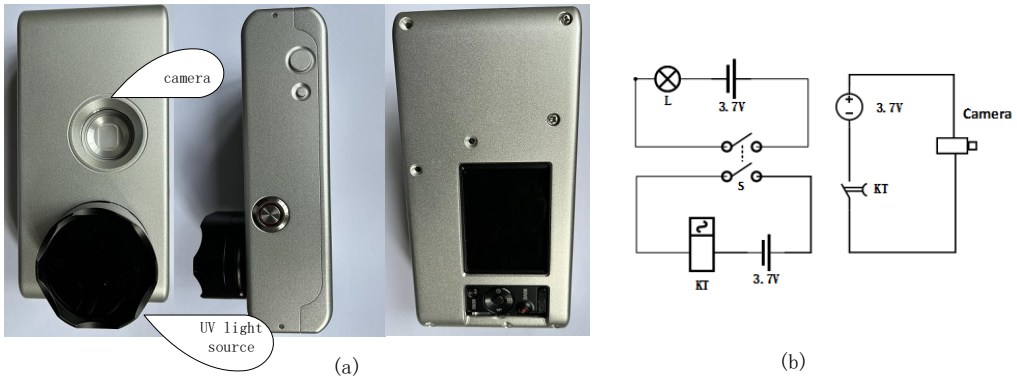

Figure 3. Device designed to acquire fluorescent images: (**a**) the appearance of the device; (**b**) simplified circuit diagram.

3.2. DAttRes-Unet Architecture

The proposed DAttRes-Unet model is an end-to-end network for leaked oil detection, and its architecture is similar to that of a classical U-net, symmetrically consisting of the encoder and the decoder, as shown in Figure 4. In the encoder, the basic convolutional unit, composed of a 3×3 convolutional layer, batch normalization, and a ReLu activation layer, are the same as in the classical U-net. A max-pooling layer with a 2×2 kernel and a stride of 2 was used to select the most powerful features. Different from the classical U-net shown in Figure 1, four residual blocks are embedded in the encoder to overcome the vanishing gradient problem. In the decoder, different from the classical U-net, an AG block is embedded to enhance the richer low- and high-level information by weighting oil pixels. Followed by a feature concatenation layer, an SE block is embedded as the channel attention to yield a higher weight coefficient for more relevant channels. Finally, a Softmax layer is used to obtain the final segmentation results. The details of the residual block and the SE block are described in the following subsection.

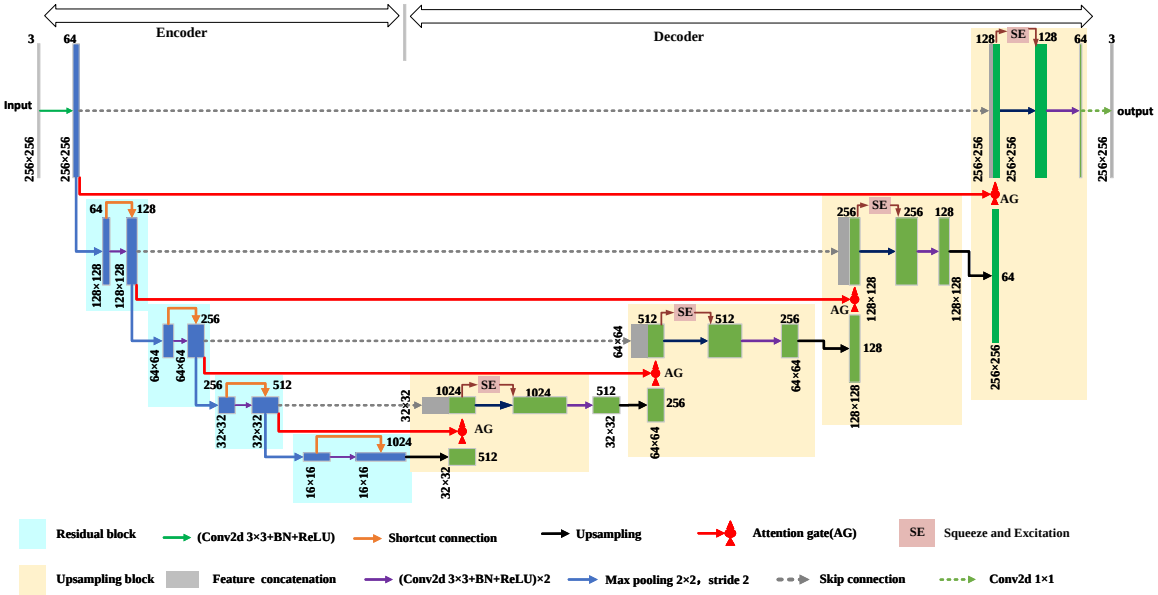

Figure 4. The architecture of the DAttRes-Unet network.

3.2.1. Residual Block

To solve the network degradation and realize the feature reuse, the residual block introduces a shortcut connection channel and an element-wise addition operation, making the network more accurate without any extra parameters. By stacking the residual

blocks [26], different types of residual networks [27], e.g., ResNet18, ResNet34, ResNet50, and ResNet101, can be constructed as the encoder within the U-net architecture.

Considering that the amount of acquired fluorescence images of transformer oil leakage is limited, if we adopt a deep residual network model, which implies that the number of training parameters increases dramatically, then the network becomes difficult to train, and the trained model is prone to over-fitting. Secondly, considering that ResNet18 has been well trained with good performance with the ImageNet dataset, we can adopt the transfer learning strategy to utilize the trained weights of ResNet18 as initial weights for leakage oil images. This will increase the training speed and improve the performance of the network. Thus, as a trade-off between computation costs and accuracy, we mainly adopted ResNet18 as the encoder path to extract features in our DAttRes-Unet network. Further, in order to be applicable to the U-Net architecture, we removed the full connection layer in the original ResNet18 network and only used a basic convolutional unit to construct the residual block.

The structure of the residual block is shown in Figure 5 and primarily consists of two convolution units and identity mapping. Each convolution unit is composed of a 3×3 convolutional layer, batch normalization, and a ReLU activation layer. The identity mapping connects the input and output of the residual block, and the residual operation is denoted as

$$y_l = x_l + F(x_l, W_l) \tag{1}$$

where x_l and y_l are the input and output vectors of the $l-$th residual unit, respectively; $F(x_l, W_l)$ is the mapping function for the residual path; W_l is the trainable weight. From (1), y_l of the residual unit can be mapped as x_l (i.e., $y_l = x_l$), even when $F(x_l, W_l) = 0$. Taking the derivative of Equation (1), we have $\partial y_l / \partial x_l = 1 + \partial F(x_l, W_l) / \partial x_l$. $\partial y_l / \partial x_l \geq 1$ is always satisfied, which implies that the residual operation avoids gradient vanishing and accelerates network convergence. By stacking the residual blocks, we can obtain hierarchical latent feature maps of the input fluorescence image, including the low-level detailed information from shallow blocks and the high-level semantic information from deep blocks.

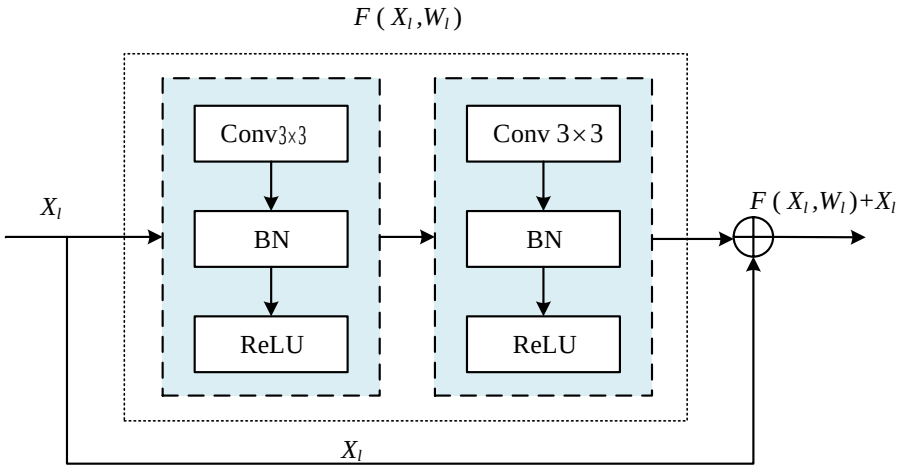

Figure 5. The architecture of the residual block.

3.2.2. The combination of AG and SE Attention Blocks

The proposed DAttRes-UNet network has four spatial attention modules and four channel attention modules, as shown in Figure 4. We use the attention gate [28] to calculate the spatial attention to highlight the oil-stained region on the feature maps while suppressing the background or irrelevant parts. The SE block [29] is also used as the channel attention to yield a higher weight coefficient for more relevant channels [30]. These attention modules are described in the following subsections.

(1) The AG attention block

Four AG blocks in the decoder learn attention at four different resolution levels, and the structure of each AG module is shown in Figure 6. Given that x_l is the feature map of the l-th layer, the attention coefficient $\alpha_i \in [0, 1]$ is used to determine focus regions and to curb useless feature information. The output of the AG x_o is the multiplication of x_l with α_i as follows:

$$x_o = \alpha_i \cdot x_l. \tag{2}$$

where the operator "$\cdot$" denotes element-wise multiplication. With multiplication, the AG attention module can overlook non-oil background information by giving more weight to feature maps with higher semantic information.

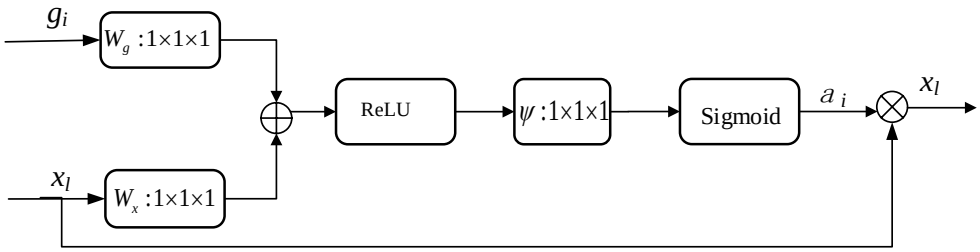

Figure 6. The architecture of the AG spatial attention module.

α_i is obtained using additive attention, and its diagram is shown in Figure 6. As the result after upsampling, the attention gating sign vector g_i is used for each pixel to determine focus regions. The additive is formulated as follows:

$$\begin{aligned} Att(x_l, g_i; \theta) &= \psi^T \left(\sigma_1 \left(W_x^T x_l + W_g^T g_i + b_g \right) \right) + b_\psi \\ \alpha_i &= \sigma_2(Att(x_l, g_i; \theta)) \end{aligned} \tag{3}$$

σ_1 is often chosen as the ReLu activation function [31] (i.e., $\sigma_1(x) = \max(0, x)$), and σ_2 is often chosen as the sigmoid function (i.e., $\sigma_2(x) = 1/(1 + \exp(-x))$). b_g and b_ψ are bias terms, and W_x, W_g, and ψ are linear transformations implemented using the $1 \times 1 \times 1$ convolution in the channel direction of the input tensor. The parameter set θ consists of W_x, W_g, ψ, b_g, and b_ψ, which must be updated by the backpropagation algorithm in the model training.

(2) The SE attention block

Typically, convolutional blocks treat channel-wise features equally and thus have difficulty managing various shapes and sizes of the leaked oil region in the images. This issue encourages us to use the SE attention block to selectively emphasize informative feature maps by calculating channel-wise summary statistics. Figure 7 shows the structure of the SE attention module.

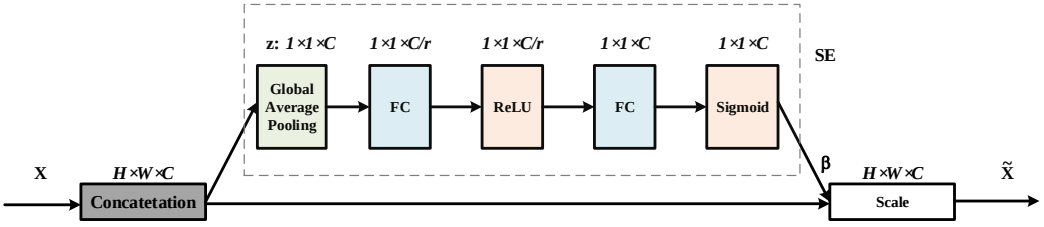

Figure 7. The architecture of the SE channel-wise attention module.

We let $X = [x_1, x_2, ..., x_c] \in R^{H \times W \times c}$ be the feature map input of the SE module, where c is the channel number, and H and W are the height and width of the feature maps, respectively. The SE module is composed of global average pooling, two fully connected (FC) layers, and their corresponding activation functions. To obtain the global information of each channel $z \in R^{1 \times 1 \times c}$, global average pooling is performed on individual feature

channels along spatial dimensions so that the global spatial information is squeezed into a channel descriptor. The c-th element of z is calculated as the global information of each channel:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i,j), \tag{4}$$

where (i,j) is the position of the c-th channel x_c.

To obtain the channel attention coefficient $\beta \in R^{1\times1\times c}$, a multiple layer perception (MLP) is constructed by two fully connected (FC) layers around the nonlinearity. The first FC layer is a dimensionality-reducing layer with reduction ratio r (i.e., the output channel number is c/r), which is then followed by a ReLU. In this study, r is set to 16. The second FC layer is a dimensionality-increasing layer with the same channel number of c as the input X, and its result is fed into a sigmoid to obtain β, given by

$$\beta = \sigma_2(W_2\sigma_1(W_1 z)), \tag{5}$$

where $W_1 \in R^{\frac{c}{r}\times c}$, and $W_2 \in R^{c\times\frac{c}{r}}$. Finally, the final output $\tilde{X}$ of the SE module is obtained by scaling the feature map input X with β:

$$\tilde{x_c} = \beta_c x_c, \tag{6}$$

where $\tilde{X} = [\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_c]$.

3.2.3. Loss Function

To update the parameters of the proposed model, we used a binary cross entropy as a loss function, defined as:

$$\mathcal{L}_p(q) = -\frac{1}{N} \sum_{i=1}^{N} r_i log(p(r_i)) + (1 - r_i) log(1 - p(r_i)) \tag{7}$$

where N is the number of data, r_i is the class of classification that has the value of 0 or 1, and $p(r_i)$ is the probability of r_i.

4. Experiments and Results

4.1. Data Pre-Processing and Augmentation

Because there is no publicly available fluorescence image dataset of transformer leaked oil, we used the custom device shown in Figure 3 to capture 307 fluorescence images at local stations. Among the acquired images, 245 and 30 images were used as training data and validating data, respectively, and the remaining 32 images were used as testing data. Each image has 565 × 584 pixels. To evaluate the performance of the model, Labelme software was used to manually label the oil-stained area for the acquired images. Considering that inputting the whole image with a larger size into the deep network can easily cause memory overflow and increase networking training time, the cropping is automatically operated on the acquired images and the corresponding labels, the size of which is consistent with the size of the targeted image in the proposed DAttResU-net model, i.e., 256 × 256 pixels.

In order to make more target features contained in the training samples and avoid over-fitting due to insufficient training samples, we increased the number of training and validating set samples, including horizontal flipping, vertical flipping, rotating, and zooming. Thus, a total of 1400 normal samples were obtained; meanwhile, 1225 and 150 images with leaked oil are also obtained for training and validating data, respectively.

4.2. Model Training

In the training procedure, initial values of network parameters in the DAttResU-net model were borrowed from that of ResNet18 and pre-trained on ImageNet dataset. Using the enhanced training data, the network parameters were constantly updated by the Adam

optimizer, to reduce the loss value of the network to the convergence value. In the Adam optimizer, two hyperparameters were set as $\beta_1 = 0.9$ and $\beta_2 = 0.99$ to ensure the convergence of the training networks. A batch size of 24 was used due to memory constraints, and the learning rate was set as 10^{-3}. When the minimum of loss value had not changed within 10 epochs or when the network had been trained for more than 50 epochs, the network training was completed, and the network parameters were saved. The test images were input into the trained model, and the image results were output after being calculated by the model. All the experiments in this paper were implemented in Python using a PyTorch backend on a single GPU with 8 GB of RAM and an NVIDIA GeForce GTX 2080.

4.3. Evaluation Indicators

In this study, the detection of leaked oil was considered a binary classification of either oil or background (non-oil). Thus, the performance was evaluated using several performance indices, including the F_1 score, Accuracy (Acc), and IOU (Intersection Over Union). Accuracy is the proportion of samples with correct predictions for entire pixels. Because the F_1 score comprehensively considers the value of both precision and recall, we used the F_1 score as one evaluation criterion. The confusion matrix, as shown in Table 1, was used to calculate ACC and F_1 as follows:

$$\begin{aligned} Acc &= \frac{TP + TN}{TP + TN + FP + FN} \\ F_1 &= 2 \times \frac{Pre \times Rec}{Pre + Rec} \end{aligned} \tag{8}$$

respectively, where *Pre* and *Rec* denote precision and recall, respectively, which are given by

$$Pre = \frac{TP}{TP + FP}\ , \qquad Rec = \frac{TP}{TP + FN}$$

respectively.

Table 1. Confusion matrix.

Actual	**Predict Value**	
	True	**False**
True	TP (True Positive)	FN (False Negative)
False	FP (False Positive)	TN (True Negative)

As a similarity measure between the predicted result A and the ground truth B, IOU is defined as

$$IOU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}. \tag{9}$$

For an image consisting of discrete objects, IoU can be rewritten by adapting the last expression in (9) as

$$IOU = \frac{1}{N} \sum_{i=1}^{N} \frac{y_i \hat{y}_i}{y_i + \hat{y}_i - y_i \hat{y}_i}, \tag{10}$$

where y_i is the number of a binary value (label) of the corresponding pixel i, and $\hat{y}_i$ is the number of predicted labels for the pixel. The larger the above metrics are, the higher the performance of the detection models.

4.4. Performance Comparison

Considering that the proposed DAttRes-Unet model inherits the framework of the classical U-net and uses the pretrained ResNet18 [32] as the encoder, we compared the proposed model with VGG16-Unet and Res18-Unet, commonly used in segmentation, where VGG16-Unet uses the pretrained VGG16 as an encoder due to VGG16 with good

performance, while Res18-Unet uses ResNet18 as an encoder without embedding any attention modules in the decoder path. VGG16 and ResNet18 were selected for use in this study because their layer depths are similar or equal to that of the proposed model, providing a fair comparison of their results.

VGG16-Unet and Res18-Unet also use the binary cross entropy in (7) as a loss function and the Adam optimizer with the same $\beta_{1,2}$, learning rate, and epoch number as the proposed model. The loss values of the three models for the training and validation data are shown in Figure 8a. Because we primarily focused on oil segmentation, and experimental results showed that there is no marked difference in the performance indices for the background (non-oil region) among the comparative models, we only show the "F_1" score and "IOU" values for oil pixels versus the training epoch in Figure 8b,c. Figure 8 shows that the models with ResNet18 as the encoder can achieve lower loss values and higher "F_1" scores and "IOU" values than those with VGG16.

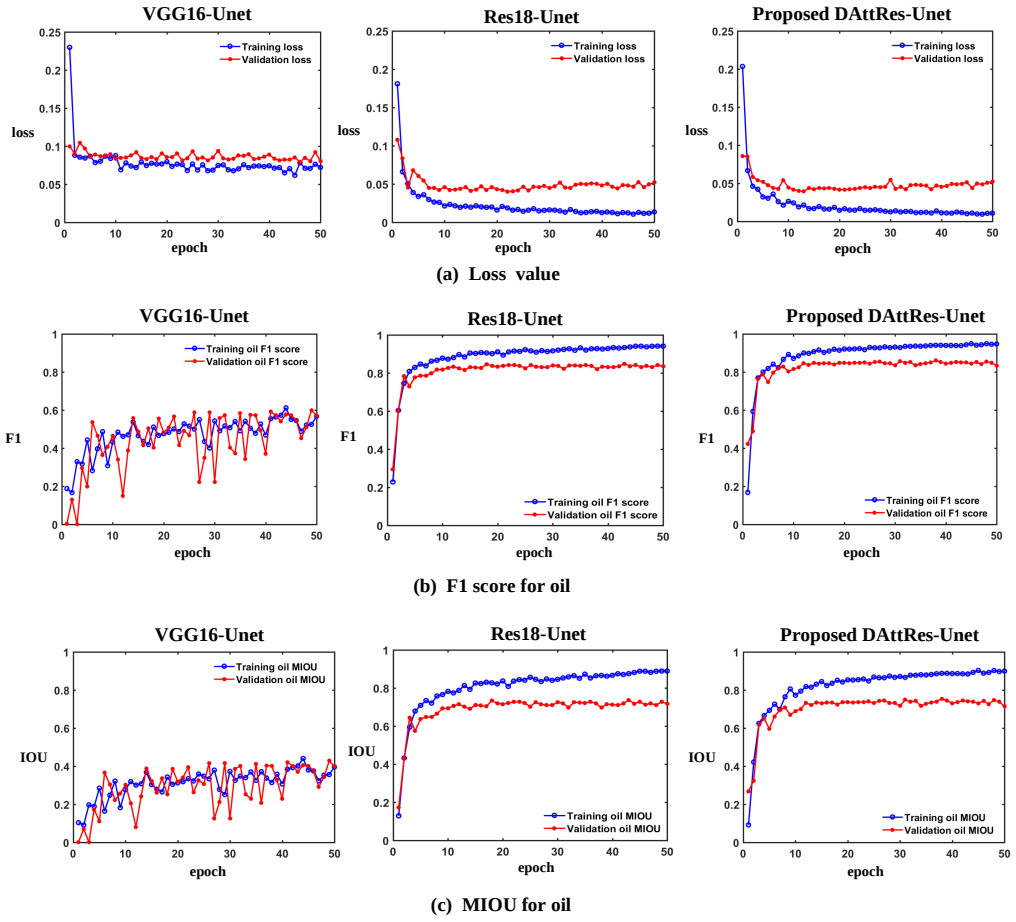

Figure 8. Loss, F_1, and IOU curves of VGG16-Unet, Res18-Unet, and the proposed DAttRes-Unet.

To compare the performance of the three models on segmented oil regions in more detail, we show the statistically calculated "F_1" and "IOU" for oil pixels and the entire "Acc" using training, validation, and testing data in Table 2. Using the ResNet18 structure as a decoder achieves better performance than that of VGG16. As an example, the proposed DAtteRes-Unet model improves ("F_1", "IOU", and "Acc") by 17.14%, 23.10%, 1.77% compared to VGG16-Unet on the testing data. Due to the dual attention modules embedded in the decoder path, the proposed DAttRes-Unet model increases ("F_1", "IOU", and "Acc") by 4.04%, 5.98%, and 0.34% compared to Res-Unet on the testing data. These results imply that the attention module is conducive to the location of the oil region.

Table 2. Performance indices of VGG16-Unet, Res18-Unet, and the proposed DAttRes-Unet on testing, validating, and training data.

Models	Indices	Testing (%)	Validating (%)	Training (%)
VGG16-Unet	F_1 (oil)	69.43	55.91	60.30
	IOU (oil)	53.24	38.80	43.16
	Acc	96.59	95.95	96.82
Res18-Unet	F_1 (oil)	82.53	83.63	94.13
	IOU (oil)	70.36	**71.91**	88.91
	Acc	98.02	**98.50**	**99.54**
Proposed DAttRes-Unet	F_1 (oil)	**86.57**	**83.66**	**94.22**
	IOU (oil)	**76.34**	71.87	**89.08**
	Acc	**98.36**	98.42	99.50

The bold in the table represents the more outstanding and excellent performance.

Figure 9 shows the segmented results of three models on several fluorescence images. Ground truth images are manually labeled using Labelme software, where regions in red denote the oil-stained regions, and regions in black denote background regions. Compared with the other two models, the proposed model can more accurately segment the oil region with fewer false detections.

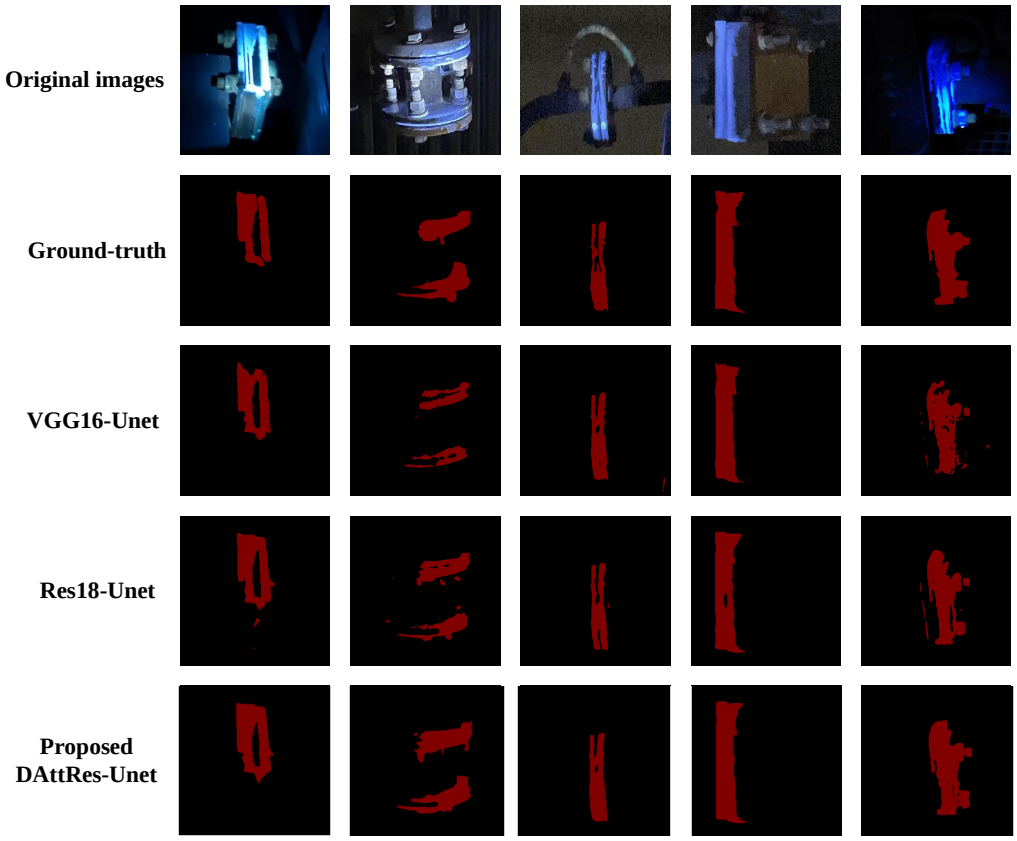

Figure 9. The segmented results of VGG16-Unet, Res-Unet, and the proposed DAttRes-Unet.

4.5. Ablation Studies

As described in Section 3.2, the proposed DAtteRes-Unet model contains two primary attention components: a spatial AG module and a channel-wise SE module. To verify the effectiveness of the two modules, we compared the DAtteRes-Unet model with AGRes-Unet and SERes-Unet, where AGRes-Unet only embeds the AG module without the SE module, and SERes-Unet only embeds the SE module without the AG module.

Table 3 shows "F_1" and "IOU" for oil pixels and the entire "Acc" with training, validating, and testing data. Compared with the results in Table 2 from the Res18-Unet model without AE and SE attention modules, embedding one of two modules improves these three performance indices with the testing data. Integrating AG and SE modules also increases "F_1" and "IOU" with the testing data; for example, the proposed DAtteRes-Unet model improves "F_1" by 1.91% and 0.92% compared to AGRes-Unet and SERes-Unet, respectively, and increases "IOU" by 2.77% and 1.45% compared to AGRes-Unet and SERes-Unet, respectively.

Table 3. Performance indices of AGRes-Unet, SERes-Unet, and the proposed DAttRes-Unet on testing, validating, and training data.

Models	Indices	Testing (%)	Validating (%)	Training (%)
AGRes-Unet	F_1 (oil)	84.66	**85.27**	94.74
	IOU (oil)	73.57	**74.33**	90.01
	Acc	98.14	**98.61**	**99.56**
SERes-Unet	F_1 (oil)	85.65	83.31	**95.22**
	IOU (oil)	74.89	71.39	**90.88**
	Acc	98.25	98.42	99.50
Proposed DAttRes-Unet	F_1 (oil)	**86.57**	83.63	94.13
	IOU (oil)	**76.34**	71.87	88.91
	Acc	**98.36**	98.42	99.50

The bold in the table represents the more outstanding and excellent performance.

Figure 10 shows the segmented results of three models on several fluorescence images. The proposed DAtteRes-Unet model with dual attention modules can thus achieve results that are more consistent with the ground truth than the model with only one attention module.

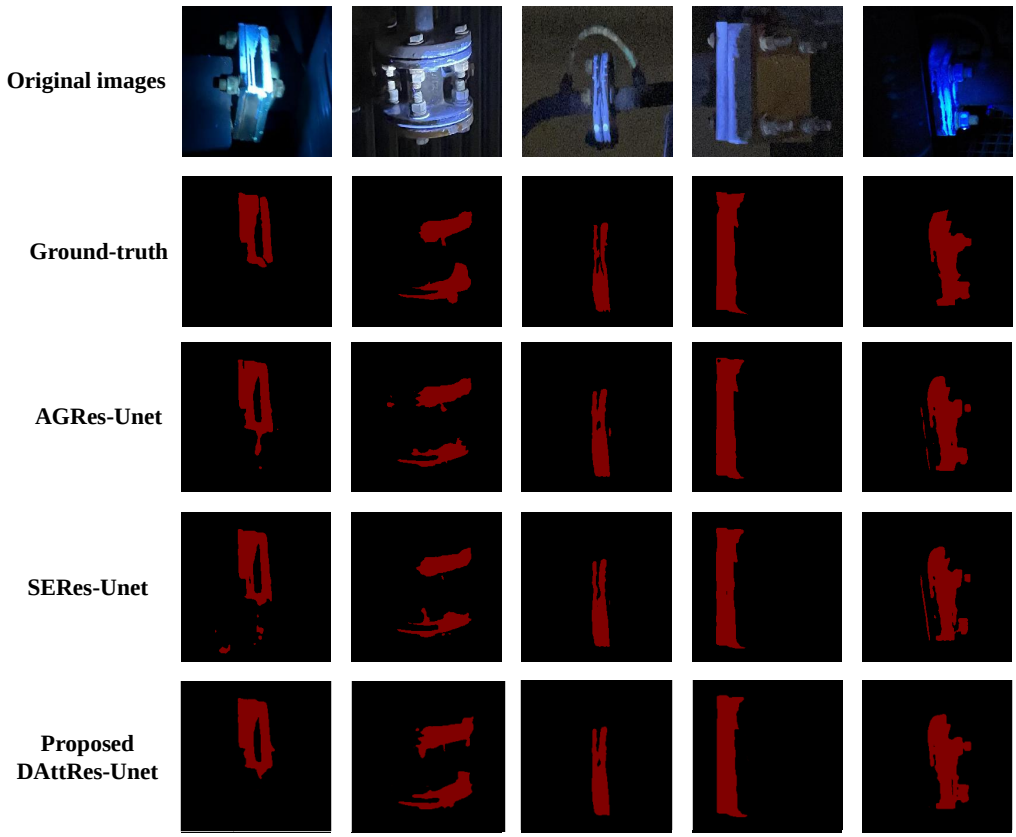

Figure 10. The segmented results of AGRes-Unet, SERes-Unet, and the proposed DAttRes-Unet.

4.6. Parameters and Testing Time

Table 4 shows the number of parameters, required memory, and testing time for each image of the five networks. Compared with VGG16-Unet, the proposed DAttRes-Unet model is lightweight. Due to the embedding of dual attention modules, compared with ResU-net, AGRes-Unet, and SERes-Unet, the proposed DAttRes-Unet model needs more parameters and memory. Although more parameters of the proposed model leads to the longest testing time, the model can satisfy requirements in practical applications. The performance of the proposed DAttRes-Unet model improves at the cost of more parameters and memory.

Table 4. Parameters, memory, and testing time of the comparative networks.

Models	Parameter	Memory (MB)	Testing Time (s)
VGG16U-net	28,142,530	972.36	0.0936
ResU-net	15,273,538	483.51	0.0876
AGRes-Unet	15,365,454	524.61	0.0912
SERes-Unet	15,318,594	506.71	0.1090
Proposed DAttRes-Unet	15,410,510	547.81	0.1140

5. Conclusions

In this paper, a new DAttRes-Unet network that is designed to segment stained oil regions by integrating ResNet and dual attention modules is proposed. The proposed model uses the pretrained ResNet18 as an encoder to mitigate issues due to few training data and gradient vanishing during backpropagation. Combining spatial and channel attention modules in the decoder improves feature representation for oil-stained regions while suppressing the background. Experimental results showed that the proposed model outperforms the commonly used VGG16-Unet and Res18-Unet, and the accuracy reached 98.49%. Extensive ablation studies also confirm the effectiveness of the dual attention modules.

In spite of the good performance of the DAttRes-Unet network, some aspects are still open for improvement in the future. For example, the SE attention module or other attention modules are applied in the encoder to emphasize more features in the image, the influence of the embedding attention modules in the encoder or in the decoder needs to be further studied. The hyperparameters (e.g., β_1 and β_2) of the Adam optimizer have considerable influence on the performance of networks. Sensitivity needs to be analyzed. The proposed network trained with fluorescence images has not been tested by the dataset of other categories. The generality of the proposed network needs to be verified in other segmentation tasks. Additionally, the generality of the proposed network needs to be improved by using some strategies, e.g., dropout by stopping feature detectors with a random probability at each training epoch, and a selection loss function. Further, to avoid the problem of over-fitting, we will continuously collect more types of fluorescent images with leaked oil to increase the amount of training data and to improve the architecture of the network by the selection of learning mode and optimization method.

Author Contributions: Conceptualization, X.L. (Xuxu Li) and X.L. (Xiaojiang Liu); methodology, X.Y.; software, W.Z.; validation, Y.X. and Y.Z.; writing—original draft preparation, X.L. (Xuxu Li) and X.Y.; writing—review and editing, W.Z.; visualization, X.L. (Xiaojiang Liu); supervision, W.Z.; funding acquisition, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of Sichuan Electric Power Company, State Grid, grant number 52199720003D.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Zhang, X.; Shi, M.; He, C.; Li, J. On Site Oscillating Lightning Impulse Test and Insulation Diagnose for Power Transformers. *IEEE Trans. Power Deliv.* **2020**, *35*, 2548–2550. [CrossRef]
2. Pahlavanpour, B.; Wilson, A. Analysis of transformer oil for transformer condition monitoring. In Proceedings of the IEE Colloquium on An Engineering Review of Liquid Insulation (Digest No. 1997/003), London, UK, 7 January 1997; pp. 1/1–1/5.
3. Lu, L.; Ichimura, S.; Moriyama, T.; Yamagishi, A.; Rokunohe, T. A system to detect small amounts of oil leakage with oil visualization for transformers using fluorescence recognition. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 1249–1255. [CrossRef]
4. Li, L.; Ichimura, S.; Yamagishi, A.; Rokunohe, T. Oil Film Detection Under Solar Irradiation and Image Processing. *IEEE Sens. J.* **2020**, *20*, 3070–3077.
5. Aljohani, O.; Abu-Siada, A. Application of digital image processing to detect transformer bushing faults and oil degradation using FRA polar plot signature. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 428–436. [CrossRef]
6. Duan, J.; He, Y.; Du, B.; Ghandour, R.M.R.; Wu, W.; Zhang, H. Intelligent Localization of Transformer Internal Degradations Combining Deep Convolutional Neural Networks and Image Segmentation. *IEEE Access* **2019**, *7*, 62705–62720. [CrossRef]
7. Aljohani, O.; Abu-Siada, A. Application of Digital Image Processing to Detect Short-Circuit Turns in Power Transformers Using Frequency Response Analysis. *IEEE Trans. Ind. Inform.* **2016**, *12*, 2062–2073. [CrossRef]
8. Jiagui, T.; Jinggang, Y.; Gaoxiang, Y.; Ju, T. A system using fluorescent fiber for partial discharge detection in transformer. In Proceedings of the 2016 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Chengdu, China, 19–22 September 2016; pp. 1–4.
9. Blue, R.; Uttamchandani, D.; Farish, O. The determination of FFA concentration in transformer oil by fluorescence measurements. *IEEE Trans. Dielectr. Electr. Insul.* **1998**, *5*, 892–895. [CrossRef]
10. Ma, M.; Chen, T.; Yang, T.; Guo, L.; Liao, J. Current Transformer Oil Leak Detection Algorithm Based on Change Detection and Grayscale Histogram Double Gaussian Fitting. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021.
11. Asadi, P.; Beckingham, L.E. Integrating Machine/Deep Learning Methods and Filtering Techniques for Reliable Mineral Phase Segmentation of 3D X-ray Computed Tomography Images. *Energies* **2021**, *14*, 4595. [CrossRef]
12. Junaid, M.; Szalay, Z.; Torok, A. Evaluation of Non-Classical Decision-Making Methods in Self Driving Cars: Pedestrian Detection Testing on Cluster of Images with Different Luminance Conditions. *Energies* **2021**, *14*, 7172. [CrossRef]
13. Hensel, S.; Marinov, M.B.; Koch, M.; Arnaudov, D. Evaluation of Deep Learning-Based Neural Network Methods for Cloud Detection and Segmentation. *Energies* **2021**, *14*, 6156. [CrossRef]
14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
15. Girshick, R. Fast r-cnn. In Proceedings of the IEEE international Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 1440–1448.
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [CrossRef] [PubMed]
17. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 21–37.
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
20. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
21. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
22. Ronneberger, O.; Fischer, P.; Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation;* Springer International Publishing: Berlin/Heidelberg, Germany, 2015.
23. Chen, L.; Xiong, W.; Yang, S.; Zhang, Z. Research on Recognition Technology of Transformer Oil Leakage Based on Improved YOLOv3. In Proceedings of the 2020 International Conference on Computer Information and Big Data Applications (CIBDA), Guiyang, China, 17–19 April 2020; pp. 454–458.
24. Gao, M.; Zhang, C.; Xu, C.; Gao, Q.; Gao, J.; Yan, J.; Liu, W.; Fan, X.; Tu, H. Electric Transformer Oil Leakage Visual Detection as Service Based on LSTM and Genetic Algorithm. In *International Conference on Internet of Things*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 90–101.
25. Balakrishna, C.; Dadashzadeh, S.; Soltaninejad, S. Automatic detection of lumen and media in the IVUS images using U-Net with VGG16 Encoder. *arXiv* **2018**, arXiv:1806.07554.
26. Brain tumor image segmentation via asymmetric/symmetric UNet based on two-pathway-residual blocks. *Biomed. Signal Process. Control* **2021**, *69*, 102841. [CrossRef]
27. Targ, S.; Almeida, D.; Lyman, K. Resnet in resnet: Generalizing residual architectures. *arXiv* **2016**, arXiv:1603.08029.

28. Zhang, J.; Jiang, Z.; Dong, J.; Hou, Y.; Liu, B. Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation. *IEEE Access* **2020**, *8*, 58533–58545. [CrossRef]
29. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018, pp. 7132–7141.
30. Roy, A.G.; Navab, N.; Wachinger, C. Recalibrating Fully Convolutional Networks With Spatial and Channel Squeeze and Excitation Blocks. *IEEE Trans. Med Imaging* **2019**, *38*, 540–549. [CrossRef] [PubMed]
31. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; Omnipress: Madison, WI, USA; 2010; pp. 807–814.
32. Yu, X.; Wang, S.H. Abnormality diagnosis in mammograms by transfer learning based on ResNet18. *Fundam. Inform.* **2019**, *168*, 219–230. [CrossRef]