HSP-UNet: An Accuracy and Efficient Segmentation Method for Carbon Traces of Surface Discharge in the Oil-Immersed Transformer

Hongxin Ji [1],*, Xinghua Liu [2], Peilin Han [1], Liqing Liu [3] and Chun He [3]

1 School of Electrical Engineering, China University of Mining and Technology, Xuzhou 221116, China; ts22230107p31@cumt.edu.cn
2 College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an 271018, China; lxh9357@163.com
3 State Grid Tianjin Electric Power Research Institute, Tianjin 300180, China; liulq328@126.com (L.L.); hechuntjsgcc@163.com (C.H.)
* Correspondence: jihongx816@163.com

Abstract: Restricted by a metal-enclosed structure, the internal defects of large transformers are difficult to visually detect. In this paper, a micro-robot is used to visually inspect the interior of a transformer. For the micro-robot to successfully detect the discharge level and insulation degradation trend in the transformer, it is essential to segment the carbon trace accurately and rapidly from the complex background. However, the complex edge features and significant size differences of carbon traces pose a serious challenge for accurate segmentation. To this end, we propose the Hadamard production-Spatial coordinate attention-PixelShuffle UNet (HSP-UNet), an innovative architecture specifically designed for carbon trace segmentation. To address the pixel over-concentration and weak contrast of carbon trace image, the Adaptive Histogram Equalization (AHE) algorithm is used for image enhancement. To realize the effective fusion of carbon trace features with different scales and reduce model complexity, the novel grouped Hadamard Product Attention (HPA) module is designed to replace the original convolution module of the UNet. Meanwhile, to improve the activation intensity and segmentation completeness of carbon traces, the Spatial Coordinate Attention (SCA) mechanism is designed to replace the original jump connection. Furthermore, the PixelShuffle up-sampling module is used to improve the parsing ability of complex boundaries. Compared with UNet, UNet++, UNeXt, MALUNet, and EGE-UNet, HSP-UNet outperformed all the state-of-the-art methods on both carbon trace datasets. For dendritic carbon traces, HSP-UNet improved the Mean Intersection over Union (MIoU), Pixel Accuracy (PA), and Class Pixel Accuracy (CPA) of the benchmark UNet by 2.13, 1.24, and 4.68 percentage points, respectively. For clustered carbon traces, HSP-UNet improved MIoU, PA, and CPA by 0.98, 0.65, and 0.83 percentage points, respectively. At the same time, the validation results showed that the HSP-UNet has a good model lightweighting advantage, with the number of parameters and GFLOPs of 0.061 M and 0.066, respectively. This study could contribute to the accurate segmentation of discharge carbon traces and the assessment of the insulation condition of the oil-immersed transformer.

Keywords: surface discharge; discharge carbon trace; image segmentation; semantic segmentation; UNet; HPA; SCA

1. Introduction

Large oil-immersed transformers play a critical role in the power system [1]. Their failure would affect the power supply of the entire system and even have serious social consequences. Preventive overhaul and maintenance of the transformers are essential for the stable operation of the power system [2–4]. Due to the metal-enclosed shell and the complex internal structure, internal defects of large transformers are difficult to detect.

Commonly used methods, such as manually drilling into the transformer and lifting the shell, face the problems of low efficiency, poor accuracy, high risk and high cost. With the rapid development of robotics and artificial intelligence, micro-robots will be an efficient tool for the inspection and detection of transformer internal defects. The basic technique, how to accurately and intelligently determine the degree of insulation degradation based on the captured images, is the key to inspection with the micro-robots.

Surface discharge is one of the most common causes of insulation degradation that occurs inside the oil-immersed transformer. It refers to the discharge along the interface of the oil and paper. Figure 1 shows surface discharge near the upper clamp and the tap-connecting lugs on the internal body structure of the transformer. It results in insulation degradation, electric leakage, or even an explosion [5,6]. In recent years, a number of transformer failures caused by surface discharge had serious impacts on power supply and new energy consumption [7,8]. Since surface discharge leads to the carbonization damage of the oil and paper composite insulation, the carbon trace is an important visual characteristic of surface discharge. The area, morphology, and edge features of carbon trace have important reference value for judging and analyzing the cause, degree, and development trend in the surface discharge [9]. Therefore, the accurate semantic segmentation of carbon traces is the premise and foundation for the micro-robot to successfully detect the discharge degree and insulation degradation trend. However, there is no relevant research at present.
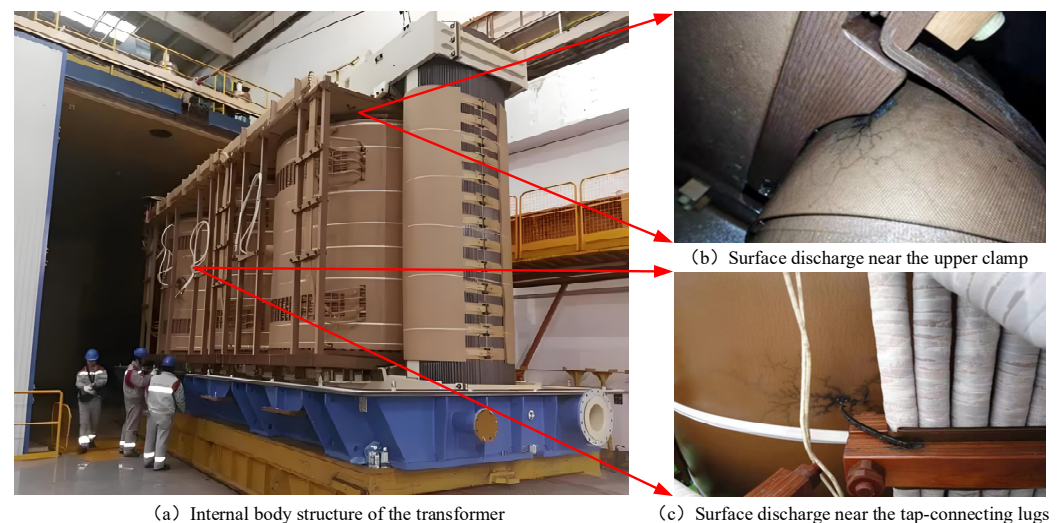

（a）Internal body structure of the transformer　　　（b）Surface discharge near the upper clamp

（c）Surface discharge near the tap-connecting lugs

Figure 1. Surface discharge and carbon traces of different parts inside the transformer.

As one of the research hotspots in computer vision, semantic segmentation aims to recognize and understand the specific meaning of each pixel in an image. Traditional semantic segmentation is mainly based on low-level features such as texture, color, and shape and then segments the image by clustering or graph cutting, etc. Typical networks include Efficient Graph-Based Image Segmentation, TextonBoost, etc. [10,11]. With the rapid development of deep learning technology, significant breakthroughs have been made in the field of semantic segmentation, and there are some novel and efficient segmentation networks, such as Fully Convolutional Network (FCN) [12], UNet [13], DeepLab [14], etc. To meet the needs of medical image segmentation, UNet was designed to perform pixel accurate localization and segmentation by feature fusion with its special encoding-decoding structure and jump connections [13]. In order to overcome the limitations of UNet's ordinary convolutional module and achieve a better perception of global features and long-range semantic information, Cao et al. proposed a U-shaped encoder-decoder architecture based on a transformer mechanism. By using a Shifted-Window module to extract contextual features, a Swin-Transformer decoder was designed for the accurate segmentation of the heart and other organ images [15]. Based on the Swin-Transformer, Atek et al. constructed the Transformer Interactive Fusion (TIF) module to realize the fusion

of different-scale features and built the dual-scale coding U-type segmentation network SwinT-Unet [16]. Although the transformer module improves the network performance, it also increases the network parameters and decreases the training and inference speed. By using regular convolution in the shallow stage and Tok-MLP module to label and project the convolutional features in the deep stage, UNeXt effectively reduced the network parameters and complexity while achieving a better segmentation performance [17]. Meanwhile, the attention mechanism has provided new insights for segmentation performance improvement. A variety of UNet structures with different attention mechanisms, such as Nested UNet [18], Resnet Coordinate Hardswish UNet (RCH-UNet) [19], and Spatial-Coordinate Attention UNet (SPCA-UNet) [20], have emerged. The above networks perform well in segmenting edge-regular targets such as medical images and traffic road images but face severe challenges in segmenting carbon traces. We put forward a much higher requirement for the segmentation model to perceive the global information and local detail features.

- Challenges of carbon trace segmentation:

① Inside the metal-enclosed shell of the transformer, the micro-robot needs supplemental light to properly acquire image data. Changes in the intensity of supplemental light lead to large differences in the overall brightness and contrast of the captured images, as shown in Figure 2a,b.

② Changes in the degree of surface discharge result in significant differences in the size of carbon traces, as shown in Figure 2c,d.

③ For the surface discharge, there is spatial randomness of the arc ablation site and local complexity of the dendritic development of carbon traces, resulting in extremely complex edge features, as shown in Figure 2e,f. This trait is the main challenge of carbon trace segmentation, which dramatically reduces the segmentation accuracy.
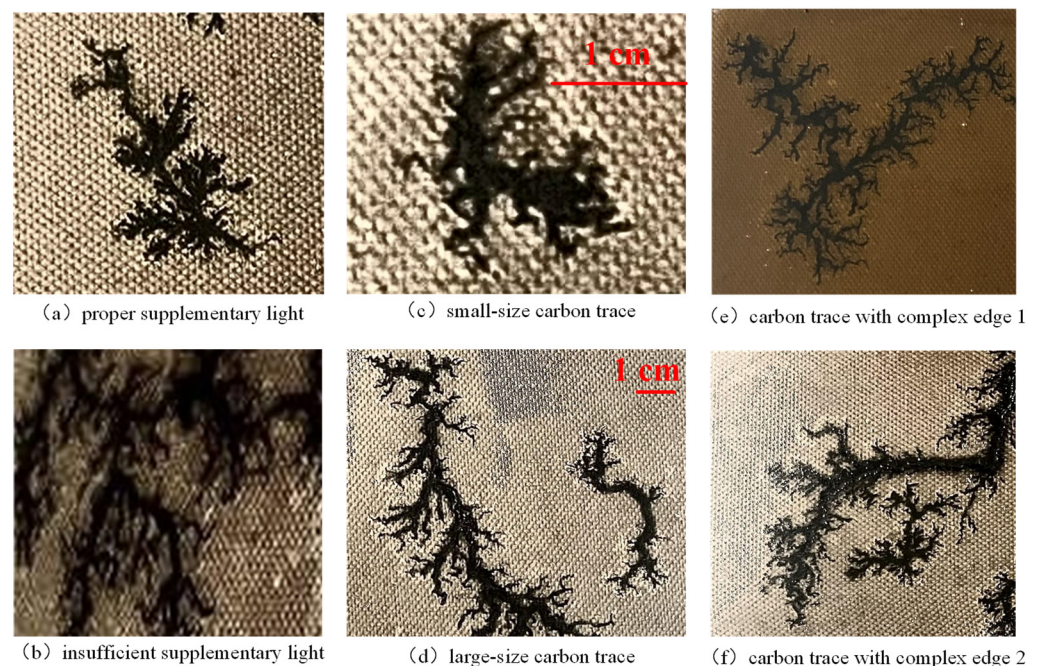

(a) proper supplementary light   (c) small-size carbon trace   (e) carbon trace with complex edge 1

(b) insufficient supplementary light   (d) large-size carbon trace   (f) carbon trace with complex edge 2

Figure 2. Significant contrast, size differences, and complex edges of the samples.

For the accurate segmentation of carbon traces, an HSP-UNet semantic segmentation network based on the UNet architecture was proposed in this paper. The proposed HSP-UNet model could achieve good semantic segmentation effect for carbon traces with complex edge features and reduce the model parameters and computational overhead.

Main contributions of this paper:

① The grouped HPA module is designed for the high-dimensional feature extraction, which reduces the algorithm complexity while realizing the effective fusion of carbon trace feature maps with different scales.

② To alleviate the semantic gap between encoder and decoder and improve the segmentation completeness of carbon trace, the SCA mechanism is designed to replace the original jump connection.

③ To improve the parsing ability of carbon trace edge features, the PixelShuffle up-sampling module with better adaptability for feature maps is used to replace the original Bilinear Interpolation module.

2. Brief Introduction of Our Inspection Micro-Robot

For efficient and convenient inspection of the transformer internal defects, an inspection micro-robot was developed, as shown in Figure 3. The micro-robot mainly consists of a body shell, an ultrasonic emission module, an ultrasonic range module, an image acquisition module, propeller propulsion modules, and a manipulation platform. The body shell of the micro-robot is an elliptical sealed structure used to mount and protect the following functional modules. The ultrasonic emission module is installed at the top of the body, which is mainly used for the three-dimensional positioning of the robot. The image acquisition module is installed on the upper part of the body, which is used to inspect the internal structure of the transformer and collect carbon trace images at the same time. Ultrasonic range modules are installed around the body shell, which are used to detect the distance between the robot and nearby objects. Propeller propulsion modules are used to control the movement of the micro-robot inside the transformer. Meanwhile, there is a manipulation platform for the micro-robot, which can remotely control the robot and store the collected images of carbon traces. The dimensions of the micro-robot are $15 \times 15 \times 26$ cm (Length $\times$ Width $\times$ Height), which is determined for high throughput of the micro-robot in the narrow space of the transformer.

1. Ultrasonic localization module
2. Communication wireless antenna
3. Body shell of the micro-robot
4. Supplementary light modules
5. Image acquisition module
6. Ultrasonic ranging module
7. Horizontal propeller propulsion
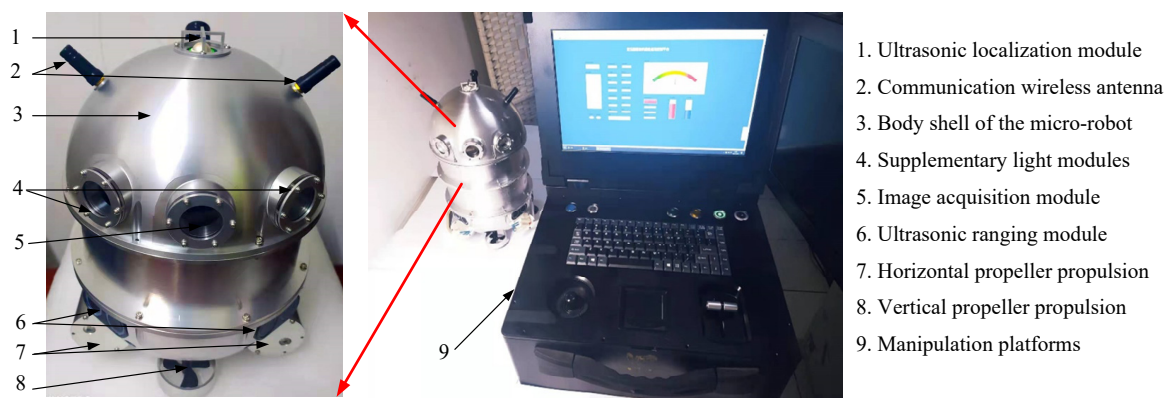8. Vertical propeller propulsion
9. Manipulation platforms

Figure 3. The micro-robot for transformer internal inspection.

As the micro-robot inspects the internal structure of the transformer, the image acquisition module will continuously capture the internal environment, and the captured images will be transmitted to the manipulation platform.

3. Carbon Trace Image Dataset

3.1. Acquisition of Carbon Trace Images

The transformer enclosure is a common site of surface discharge. Due to the difficulty of obtaining carbon trace images of surface discharge inside the actual operating transformer, there are not enough samples. Therefore, an oil–paper insulation discharge test platform was constructed to restore the transformer internal scene and artificially generated carbon trace samples. The test platform mainly consists of a specimen model, a boosting platform, and an image acquisition module. The specimen model consists of a nylon

screw, acrylic board, nylon bracket, front electrode, voltage equalization ring, connecting rod, and oil-immersed cardboard, as shown in Figure 4. The size of the oil-immersed cardboard is 25 cm × 15 cm, which is attached to the acrylic plate with a nylon clamp. The tilt angle of the oil-immersed cardboard could be changed by adjusting the angle of the clamp. The boosting platform adopts a transformer (SB-10KVA/100KV, Haotai Technology, Yangzhou, China) to provide the discharge voltage, and the specimen container consists of a transparent acrylic sheet for easy observation and the collection of carbon trace images. The transformer oil used in the test is Keramay # 25 oil. An industrial camera (HTSUA134GC/M, Huateng Vision, Shenzhen, China, 1.3 megapixels, frame rate 211FPS) was used to collect carbon trace images at 25 cm from the oil-immersed cardboard.
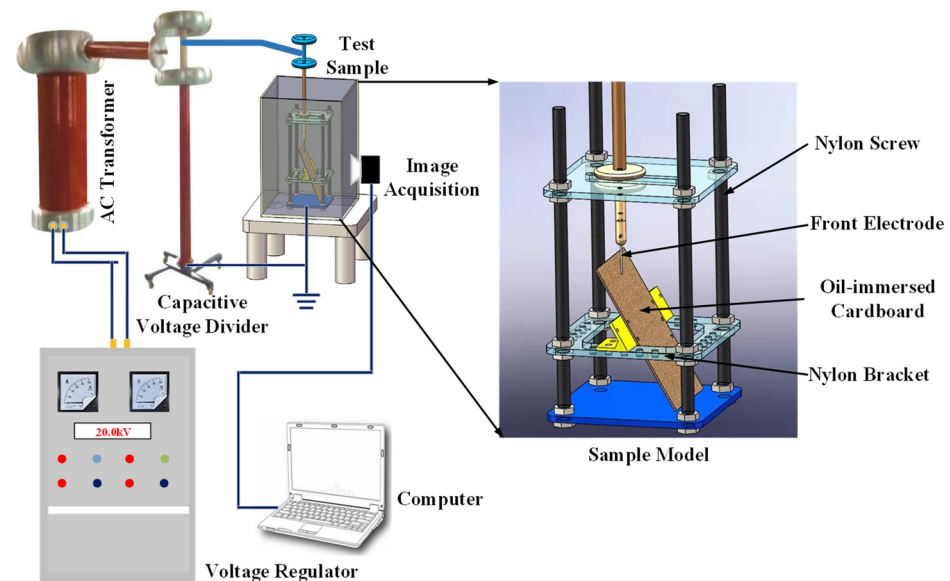

Figure 4. Test platform for carbon trace image acquisition.

Generally, surface discharge will produce two kinds of carbon traces, i.e., dendritic carbon trace and clustered carbon trace. When the oil–paper insulation stays dry, dendritic carbon trace will appear with surface discharge. When the oil-paper insulation is damp, clustered carbon trace appears. In this paper, a total of 499 images of dendritic carbon trace and 565 images of clustered carbon trace were collected. As shown in Figure 5, dendritic carbon trace has a very complex edge, which is the main challenge for accurate semantic segmentation. In contrast, the edge of clustered carbon trace is much smoother, which is relatively much easier to segment.

3.2. Image Enhancement Based on the AHE Algorithm

Restricted by the metal-enclosed shell of the oil-immersed transformer, the acquisition of carbon trace often suffers from the problem of insufficient complementary light, resulting in the carbon trace images showing an over-concentration of pixel values, weak contrast, and other problems. At the same time, the oil stains on the oil-immersed cardboard also tend to cause local reflections, which reduces the clarity of carbon trace images. In order to improve the quality of carbon trace images and reduce the difficulty of extracting carbon trace features by the semantic segmentation model, the AHE algorithm was adopted in this paper [21]. Using the distribution function of the cumulative probability of image gray level as the transformation function, the AHE algorithm focuses on the local region of the carbon trace and performs a pixel-by-pixel localized histogram equalization. It can alleviate the over-concentration of pixel values and effectively enhance the contrast of carbon trace images. Due to insufficient supplemental light, the original carbon trace image was dark and low contrast, as shown in Figure 6a. The corresponding gray level probability distribution of the original image was too concentrated, as shown in Figure 6b. After

processing with the AHE algorithm, the overall brightness of the image was significantly improved, and the edges of the carbon trace were clearer, as shown in Figure 6c. The processed image showed a uniform distribution of gray levels, indicating that the quality of carbon trace images could be greatly improved with the AHE algorithm, as shown in Figure 6d.
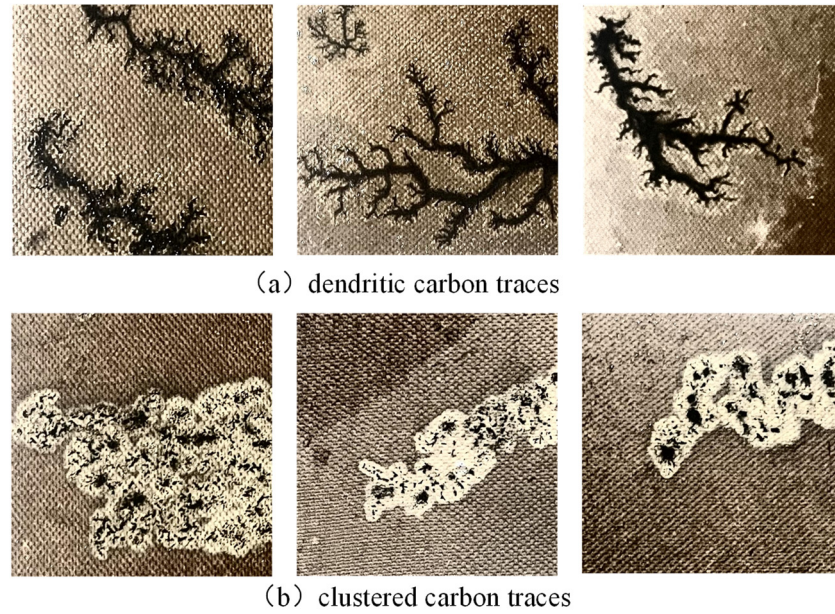

（a）dendritic carbon traces

（b）clustered carbon traces

Figure 5. Examples of two kinds of discharge carbon traces.

(**a**) Original carbon trace image

(**b**) Gray level probability distribution of (a)

(**c**) Improved carbon trace image

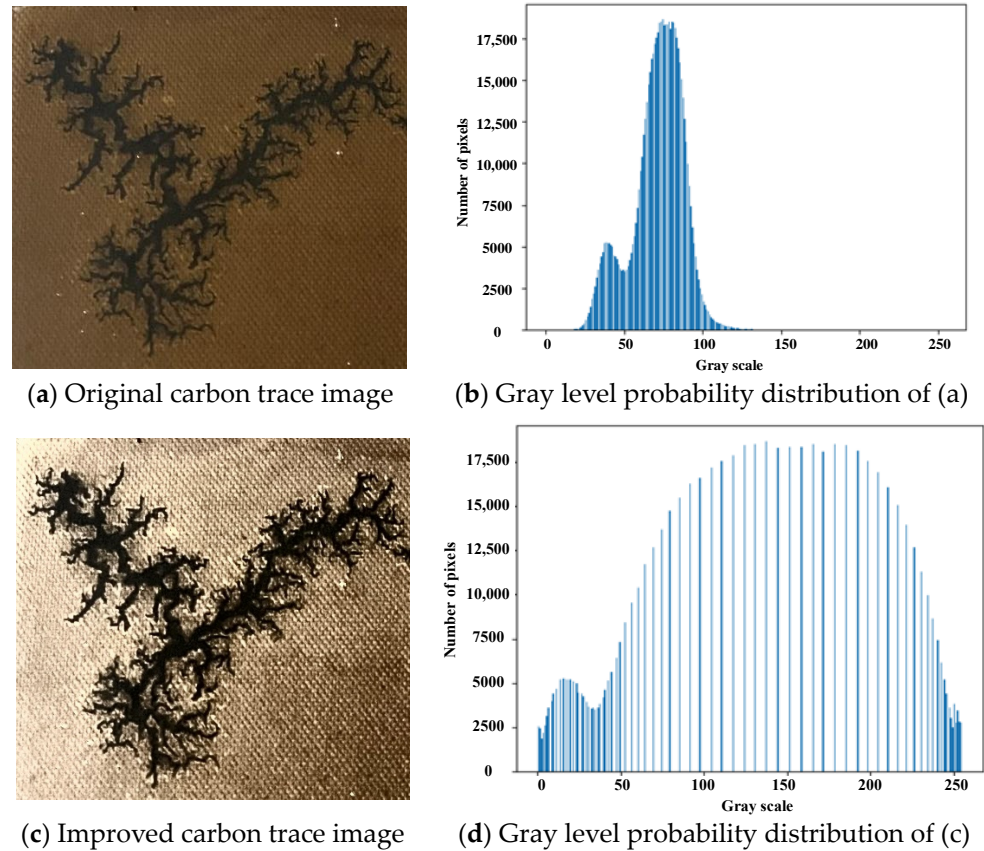(**d**) Gray level probability distribution of (c)

Figure 6. Comparison of carbon trace image with and without the AHE.

3.3. Construction of Carbon Trace Dataset

The acquisition of carbon trace samples inside the oil-immersed transformer is difficult and costly, resulting in the inadequacy of carbon trace samples. In order to improve the segmentation performance and generalization ability of the proposed semantic segmentation model, Gaussian fuzzy process, horizontal and vertical flip, image scale, and horizontal and vertical translation were used to augment the original carbon trace samples. Then, a dataset of the dendritic carbon trace $Set_{dentric}$ was constructed, which contained 2495 samples, and a dataset of the clustered carbon trace $Set_{cluster}$ was constructed, which contained 2825 samples. Furthermore, these two datasets were divided into a training set, validation set, and test set in the ratio of 8:1:1. For the dataset of dendritic carbon traces, the sample sizes in the training set, validation set, and test set were 1996, 250, and 249, respectively. For the dataset of clustered carbon traces, the sample sizes in the training set, validation set, and test set were 2260, 283, and 282, respectively.

4. Proposed Network

4.1. Network Structure of HSP-UNet

In order to improve the perception of complex edge features and realize the accurate segmentation of carbon trace image, a high-precision semantic segmentation model HSP-UNet was designed based on the structure of a UNet network, introducing the grouped HPA module, SCA attention mechanism, and PixelShuffle up-sampling module. The proposed network consisted of an encoder and a decoder, which were composed of 6-layer down-sampling and 6-layer up-sampling modules, respectively, as shown in Figure 7. The specific design process was as follows:

① Design the grouped HPA module to replace the conventional Conv2d module in the Stage 4~6 layers, which can reduce the number of model parameters and complexity while effectively integrating carbon trace features from different perspectives.

② Design the SCA mechanism to replace the original jump connection, which can alleviate the semantic gap between the encoder and decoder and improve the perception ability for complex edge features of carbon traces, improving the completeness and accuracy of carbon trace segmentation.

③ Use the PixelShuffle module to replace the Bilinear Interpolation (BI) up-sampling module of the decoder, which can help parse the deep semantic features of carbon traces, improving the segmentation accuracy of the complex edge of carbon traces.

4.2. Grouped HPA Module

In order to reduce the model parameters and improve the perception of multi-view features, the HPA module with linear complexity was adopted in this paper. According to the size parameters [B, C, H, W] of the input feature map, a tensor p is randomly initialized and adjusted by using the BI algorithm. In order to extract the multi-view features of the input feature map, a grouped HPA module was constructed, which was inspired by the multi-head self-attention (MHSA) mechanism [22], as shown in Figure 8. The input feature map was divided into four groups {X1, X2, X3, X4} along the Channel dimension. The HPA operations were performed on the H-W, C-H, and C-W axes for the first three groups {X1, X2, X3}, respectively. And the depthwise separable convolution (DW) is performed on the fourth group X4. Then, the four groups of feature maps along the Channel dimension were concatenated using the Concat instruction. The merged feature maps were processed by using LN and DW instructions to obtain the final output feature map.
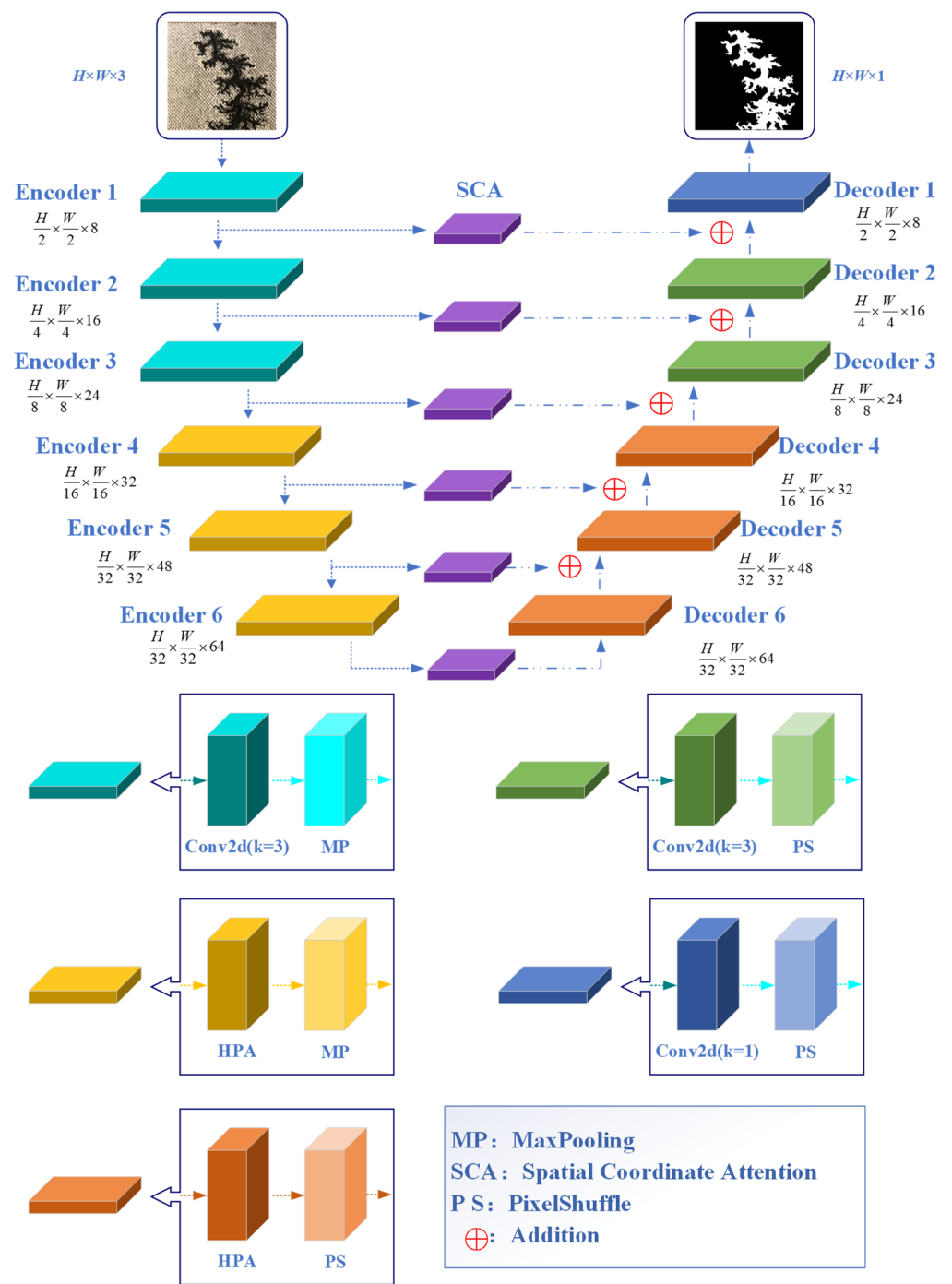
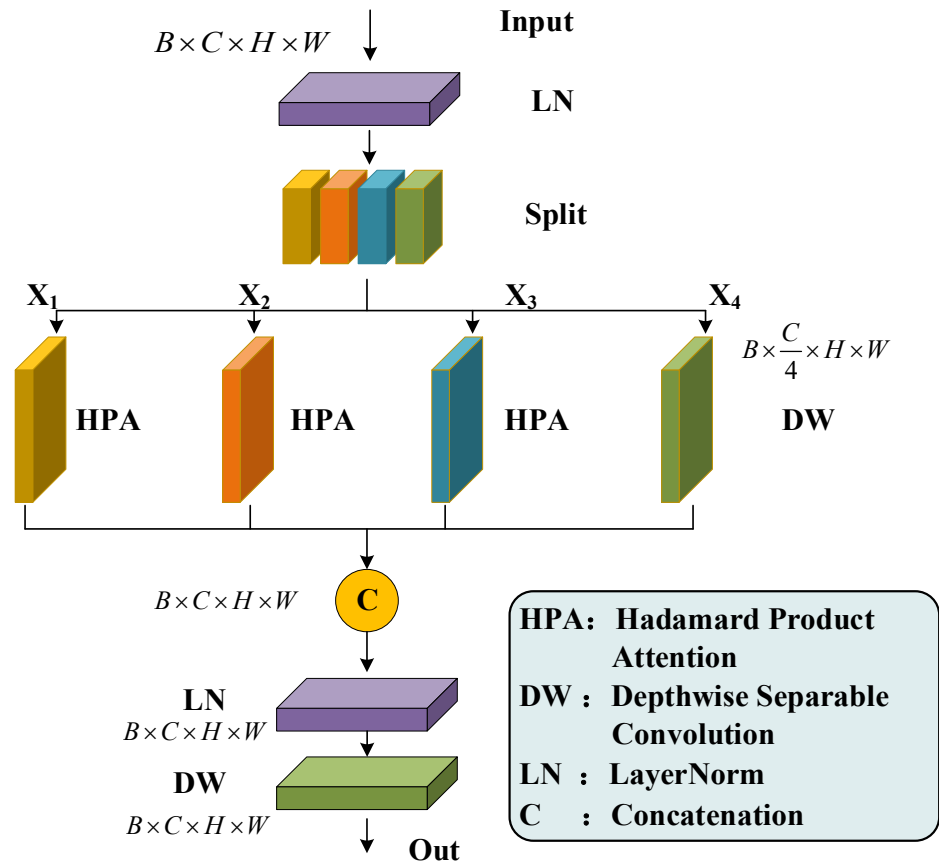**Figure 7.** Network structure of the proposed HSP-UNet.

Figure 8. Structure of the grouped HPA module.

4.3. SCA Attention Mechanism

Compared with other attention modules such as SENet [23] and CBAM [24], Coordinate Attention [25] (CA) could better accurately localize and identify the target of interest in the global range, which reduces the loss of spatial positional information as well as the module parameters. The structure of the CA is shown in Figure 9. The input feature maps are averagely pooled in the height direction and width direction, respectively, to obtain the feature maps in two directions, as shown in Equation (1). Then, to obtain the feature maps of $1 \times (W + H) \times C/r$, the processed feature maps are concatenated and sequentially processed with Conv2d, BatchNorm, and Sigmoid instructions, as shown in Equation (2). Next, the feature map is split into two tensors f^h and f^w along the Channel direction and processed with Conv2d and Sigmoid instructions. As a result, the attention weights g^h and g^w on the height and width direction were obtained, as shown in Equation (3). Finally, the g^h and g^w were weighted with the input feature map to obtain the output feature map, as shown in Equation (4) [25].

$$\begin{cases} Z^h{}_c(h) = \dfrac{1}{W} \sum_{0 \le i < W} |x_c(h,i) \\ Z^w{}_c(w) = \dfrac{1}{H} \sum_{0 \le i < H} |x_c(j,w) \end{cases} \tag{1}$$

$$f = \delta(F_1([z^h, z^w])) \tag{2}$$

$$\begin{cases} g^h = \sigma(F_h(f^h)) \\ g^w = \sigma(F_w(f^w)) \end{cases} \tag{3}$$

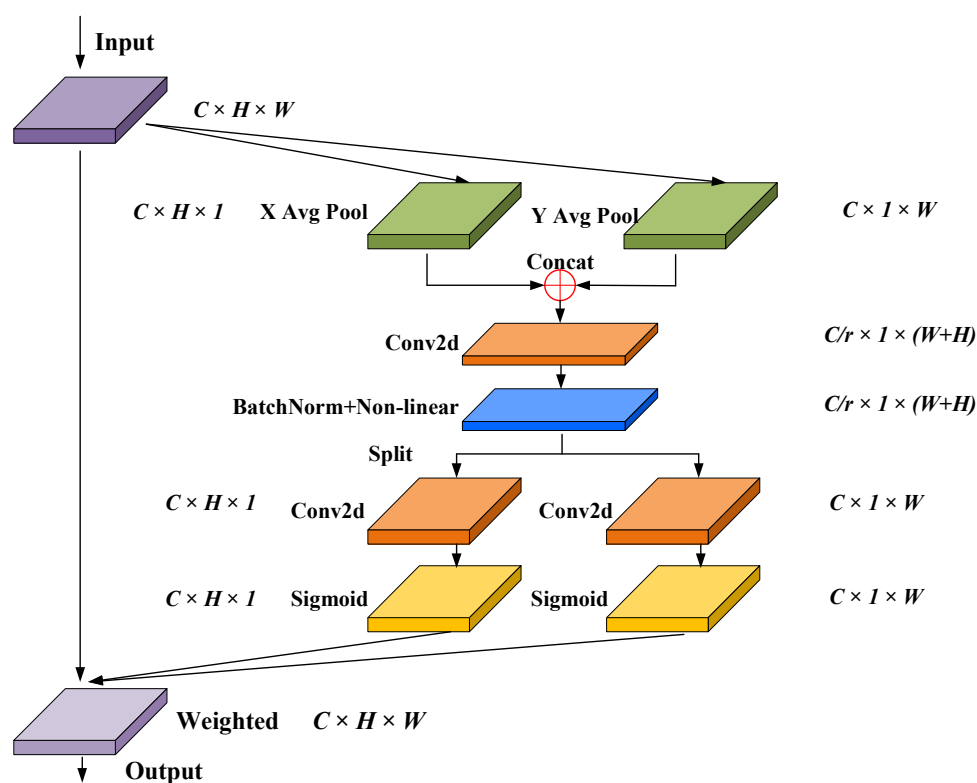$$y_c(i,j) = x_c(i,j) \times g^h{}_c(i) \times g^w{}_c(j) \tag{4}$$

Figure 9. Structure of the CA module.

The discharge carbon trace has obvious spatial edge features, presenting complex edges and rich local information, while its color, texture, and other features are not obvious. So, it is necessary to improve the ability of the CA to perceive the spatial location information. Therefore, the CA-based SCA is proposed to further improve the ability to extract the detailed edge features of carbon traces. First, the input feature layer is globally pooled with the average pool operation and maximum pool operation in the channel direction, respectively. So, the feature layers of the maximum value and the average value at the spatial level are obtained, with the shape of [1, H, W]. Then, the two feature layers of the maximum and the average value are concatenated and fed into the Conv2d layer with the channel number of one, achieving the fusion of spatial location information of the carbon trace. After the operation of activation function, the spatial feature parameters of the input feature layer are obtained. Finally, the above spatial features are weighted with the original input feature layer to enhance the spatial feature of the input layer, which is then sent to the CA mechanism. The structure of the proposed SCA is shown in Figure 10.

4.4. PixelShuffle Upsampling Module

Compared with the original BI module in the UNet, the PixelShuffle module is able to learn and optimize its own up-sampling parameters independently, which has better adaptability to the feature map and a better pixel reconstruction effect [26]. For the feature maps of dendritic carbon traces, the PixelShuffle module could better retain the detailed features and boundary information, which helps to improve the semantic segmentation accuracy. In the PixelShuffle module, a L-layer convolutional network is used to process the low-resolution feature maps of carbon trace, whose $L - 1$ layers are shown in Equation (5). For the Lth layer, a convolution with a step size of $\frac{1}{r}$ is used to up-sample the feature maps of carbon trace from the low-resolution space to the high-resolution space, as shown in Equations (6) and (7). And the loss function of the above up-sampling module is the pixel-wise MSE, as shown in Equation (8) [26].

$$\begin{cases} f^1\left(I^{LR}; W_1, b_1\right) = \phi(W_1 * I^{LR} + b_1) \\ f^l\left(I^{LR}; W_{1:l}, b_{1:l}\right) = \phi(W_l * f^{l-1}(I^{LR}) + b_l) \end{cases} \tag{5}$$

where W_l, b_l, and $l \in (1, L-1)$ are the learnable network weights and bias parameters, respectively; W_1 is a 2D convolution tensor of size $N_{L-1} \times N_1 \times K_l \times K_l$, where N_l is the number of features in the lth layer, the value of N_0 is C, and K_l is the size of the filter in the lth layer; and the bias parameter b_l is a vector with the length of N_l.

$$I^{SR} = f^L\left(I^{LR}\right) = PS\left(W_L * f^{L-1}\left(I^{LR}\right) + b_L\right) \tag{6}$$

$$PS(T)_{x,y,c} = T_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, C \cdot r \cdot mod(y,r) + C \cdot mod(x,r) + c} \tag{7}$$

where PS is a concatenating operator for the periodic pixels that can transform a tensor with the size of $H \times W \times C \cdot r^2$ into a tensor with the size of $rH \times rW \times C$, as shown in Equation (7); and W_L is a convolution operator with the size of $n_{L-1} \times r^2C \times k_L \times k_L$.

$$\ell(W_{1:L}, b_{1:L}) = \frac{1}{r^2 HW} \sum_{x=1}^{rH} \sum_{x=1}^{rW} (I_{x,y}^{HR} - f_{x,y}^L(I^{LR}))^2 \tag{8}$$
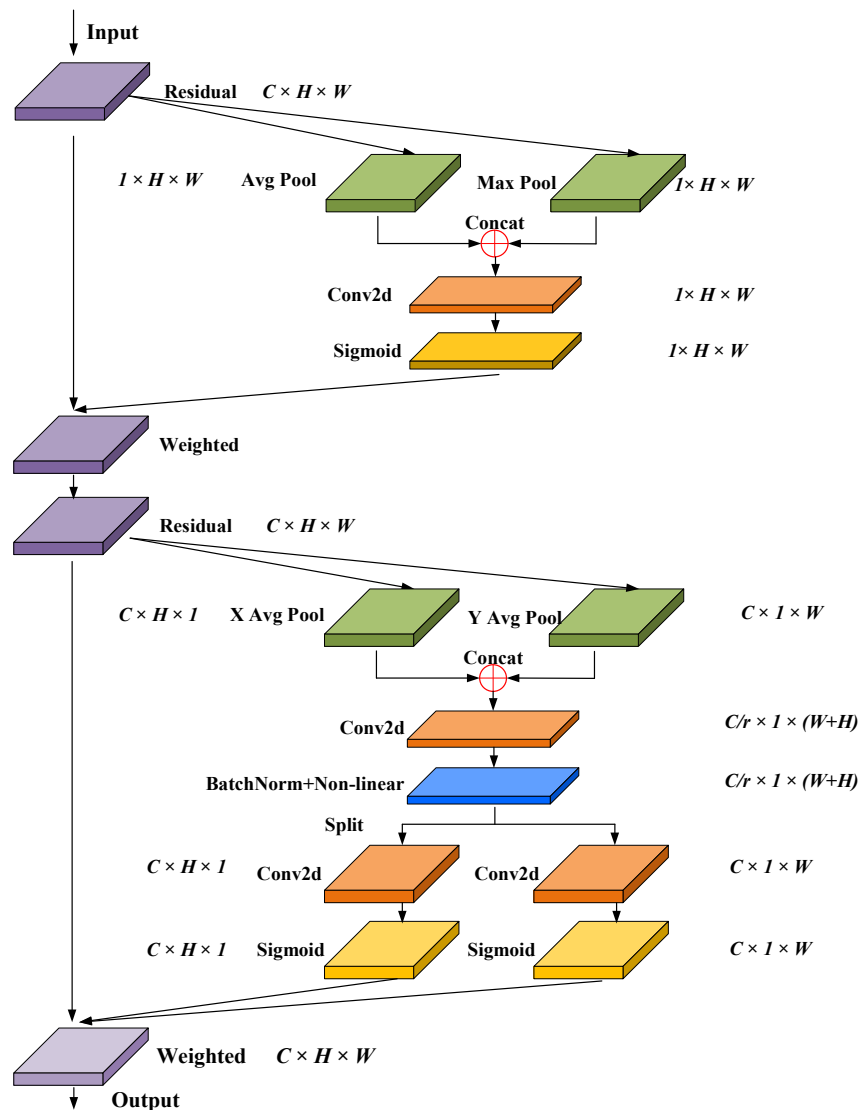

Figure 10. Structure of the SCA.

5. Results and Discussion

5.1. Training Setup

The training environment is Windows 11 × 64, and the hardware parameters are as follows: CPU Intel(R) Core (TM) i5-12500H, RAM 16 GB, GPU Nvidia GeForce RTX2050, video memory 4 GB; and the software parameters are as follows: Python 3.8.17, training framework PyTorch 2.0.1, CUDA version 11.8, CUDNN 8.9.3. The model is trained using an AdamW optimizer, with an initial learning rate of 1×10^{-3}, a learning rate adjustment strategy of CosineAnnealingLR, a weight decay coefficient of 1×10^{-2}, the epochs of 300, and a batch size of eight.

5.2. Evaluation Metrics

In order to verify and evaluate the performance of the proposed HSP-UNet model, three evaluation metrics, Mean Intersection over Union (MIoU), Pixel Accuracy (PA), and Class Pixel Accuracy (CPA) based on the confusion matrix are used in this paper. The calculations of these three metrics are shown in Equations (9)–(11) [19].

$$I_{\mathrm{mIoU}} = \left(\frac{T_\mathrm{p}}{T_\mathrm{p} + F_\mathrm{p} + F_\mathrm{n}} + \frac{T_\mathrm{n}}{T_\mathrm{n} + F_\mathrm{n} + F_\mathrm{p}} \right) \times 100\% \tag{9}$$

$$P_\mathrm{A} = \frac{T_\mathrm{p} + T_\mathrm{n}}{T_\mathrm{p} + T_\mathrm{n} + F_\mathrm{p} + F_\mathrm{n}} \times 100\% \tag{10}$$

$$C_{\mathrm{PA}} = \frac{T_\mathrm{p}}{T_\mathrm{p} + F_\mathrm{p}} \times 100\% \tag{11}$$

where T_p is the correctly identified real sample of carbon trace, F_p is the incorrectly identified real sample of carbon trace, T_n is the correctly identified sample of the background, and F_n is the incorrectly identified sample of the background.

5.3. Validation of the HSP-UNet

In order to verify the effectiveness of the proposed HSP-UNet, the models UNet [13], UNet++ [27], UNeXt [17], MALUNet [28], and EGE-UNet [22] were used to carry out the comparative analysis. The dataset of dendritic carbon trace $Set_{dentritic}$ and the dataset of clustered carbon trace $Set_{cluster}$ were used to train the above models, respectively. As shown in Table 1, among the six segmentation models involved in the comparison, the model parameters and computational GFLOPs of the HSP-UNet were only 0.061 M and 0.066, respectively, which shows a better lightweighting advantage than the other models. The model complexity and arithmetic power demand of the HSP-UNet were comparatively low, which is conducive to the practical deployment of the HSP-UNet. The segmentation effects of all the models on the clustered carbon trace samples were better than that on the dendritic carbon trace samples. The main reason is that the edges of the clustered carbon traces are smoother and easy to segment, while the edges of the dendritic carbon traces are much more complex, which substantially increases the segmentation difficulty. Specifically, HSP-UNet demonstrated obvious performance advantages on both datasets of carbon traces. For the dataset of dendritic carbon traces, HSP-UNet improved the MIoU, PA, and CPA of the benchmark model UNet by 2.13, 1.24, and 4.68 percentage points, respectively. Compared with the other four segmentation models, it also showed a better segmentation performance. For the dataset of clustered carbon traces, the MIoU, PA, and CPA of the six segmentation models were all higher than 90%, 97%, and 94%, respectively. Since the benchmark UNet already had a good segmentation effect on the clustered carbon trace, the MIoU, PA, and CPA of the HSP-UNet are improved by 0.98, 0.65, and 0.83 percentage points, respectively. The enhancement of segmentation performance with the dataset of clustered carbon traces was smaller than that with the dataset of dendritic carbon traces.

Table 1. Segmentation comparison of carbon traces with 6 models.

Dataset	Model	Params↓	GFLOPs↓	I_{mIoU} (%)	P_{A} (%)	C_{PA} (%)
	UNet (Base)	31.2 M	13.76	73.53	93.17	84.42
	UNet++	9.2 M	34.86	73.47	93.07	84.32
$Set_{dendritic}$	UNeXt	1.5 M	0.57	73.93	93.58	85.47
	MALUNet	0.177 M	0.085	74.15	93.79	86.23
	EGE-UNet	0.053 M	0.072	74.71	94.10	86.30
	HSP-UNet	**0.061 M**	**0.066**	**75.66**	**94.41**	**89.10**
	UNet (Base)	31.2 M	13.76	90.41	97.42	94.56
	UNet++	9.2 M	34.86	90.43	97.53	94.77
$Set_{cluster}$	UNeXt	1.5 M	0.57	90.54	97.44	94.53
	MALUNet	0.177 M	0.085	91.14	97.58	95.13
	EGE-UNet	0.053 M	0.072	91.21	98.01	95.24
	HSP-UNet	**0.061 M**	**0.066**	**91.39**	**98.07**	**95.39**

The segmentation effects of the dendritic carbon traces and the clustered carbon traces were comparatively analyzed in Figures 11 and 12, respectively. For the dendritic carbon traces, the four models, UNet, UNet++, UNeXt, and MALUNet, were weak in perceiving the edge features and failed to accurately segment the local complex edge, resulting in low values of the MIoU. EGE-UNet adopted the GAB module to replace the jumping connection between the encoder and the decoder, which increased the segmentation accuracy for the edges of carbon trace. However, compared with the GroundTruth, EGE-UNet had poor perception accuracy of local features and lost a large amount of carbon trace details. Compared with the above five models, the HSP-UNet proposed in this paper achieved the best segmentation effect. The SCA mechanism effectively improved the perception of carbon trace edge features, contributing to a better segmentation completeness of carbon trace. And the PixelShuffle upsampling module helped to resolve the carbon trace details, resulting in the refined segmentation of carbon traces, and retained enough detailed information. Due to the smooth edges of clustered carbon trace, the segmentation effects of the six models were less different. However, when focusing on the segmentation effect of the local area with larger curvature (the red circle in Figure 12), the HSP-UNet model still showed a better segmentation performance.

5.4. Generalization Performance of the HSP-UNet

To validate the segmentation performance of the HSP-UNet with different light conditions, carbon trace samples with sufficient and insufficient supplementary light were selected from the dendritic and clustered trace datasets. As shown in Figure 13a,b, the HSP-UNet segmented two dendritic carbon traces completely and accurately, with the MIoU of 0.736 and 0.775 for the light-sufficient one and the light-insufficient one. The reason for the relatively lower MIoU value of the light-sufficient sample is that the edge features of this sample are too complex to segment the detail boundaries. With respect to the clustered carbon traces of sufficient and insufficient light, the HSP-UNet also showed a steady segmentation performance, with the MIoU of 0.922 and 0.907, respectively. Similarly, to validate the segmentation performance of the HSP-UNet with samples of different sizes, four carbon traces were selected from the dendritic and clustered datasets. The carbon traces in Figure 14a,c are much larger than that in Figure 14b,d. The segmentation results indicated a good generalization performance of the proposed HSP-UNet with the dendritic and clustered samples, with the MIoU values of 0.774, 0.741, 0.918, and 0.913 from subplot (a) to (d) in Figure 14.
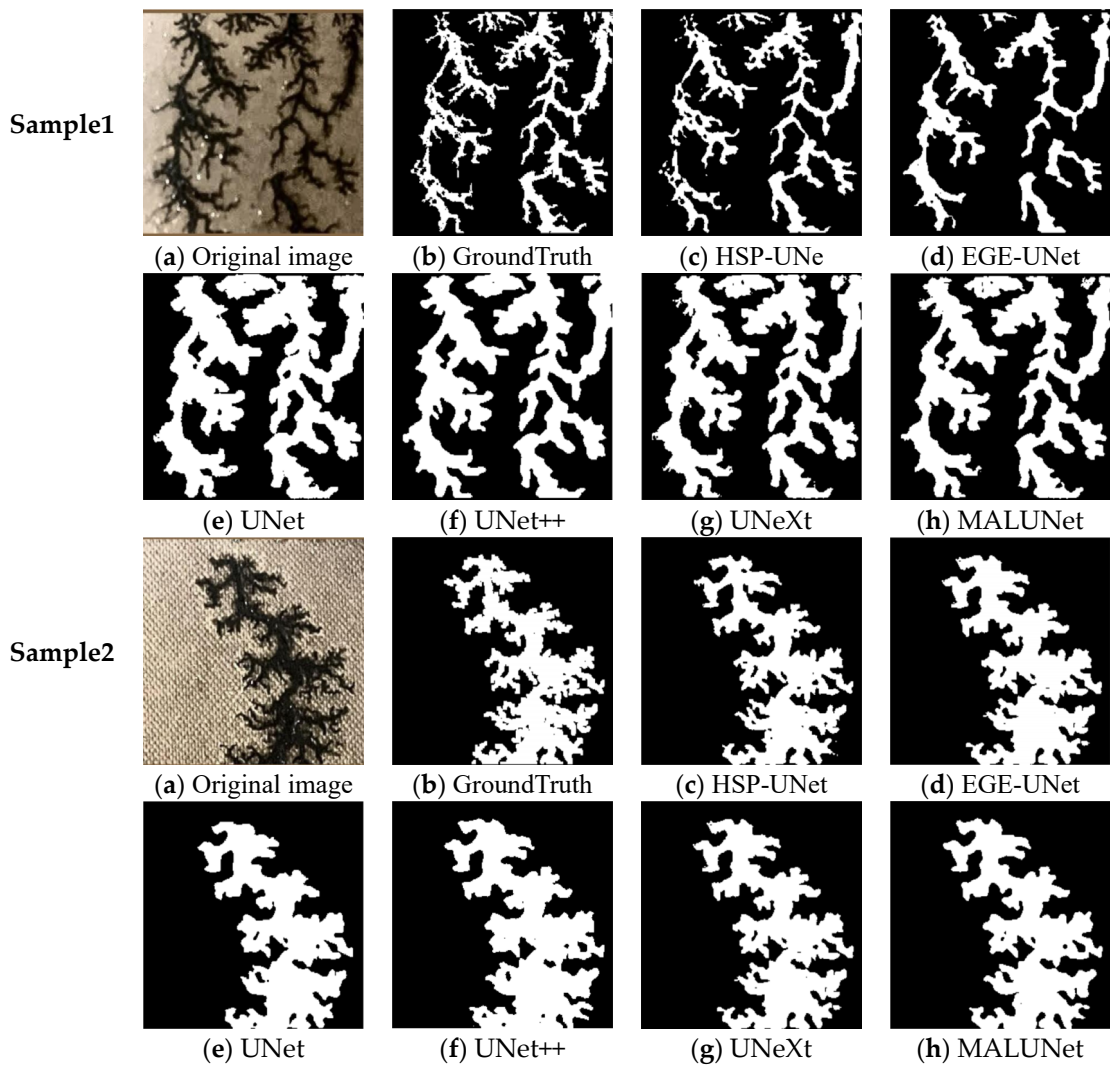
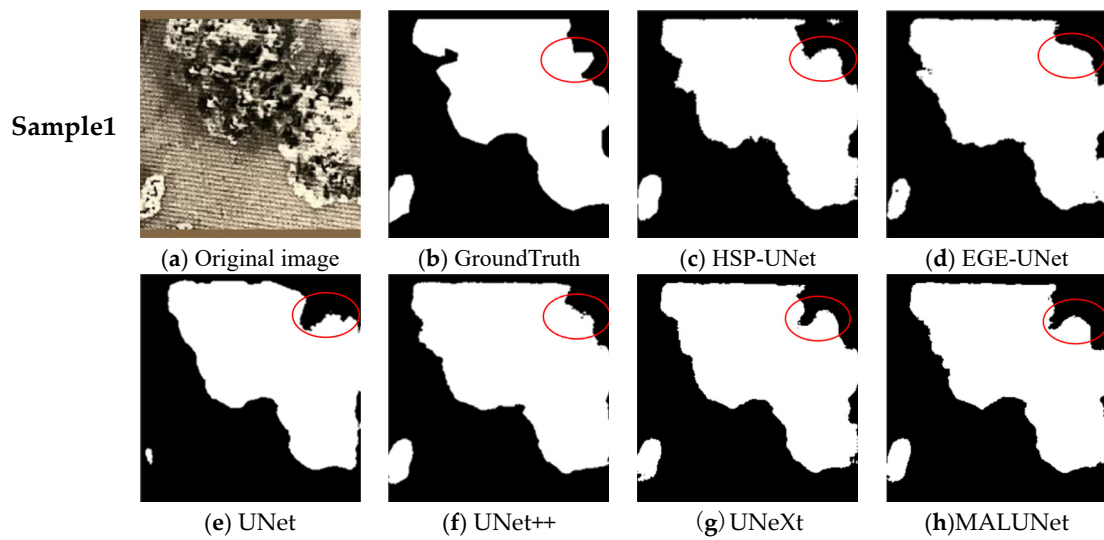**Figure 11.** Segmentation comparison of the dendritic carbon traces.
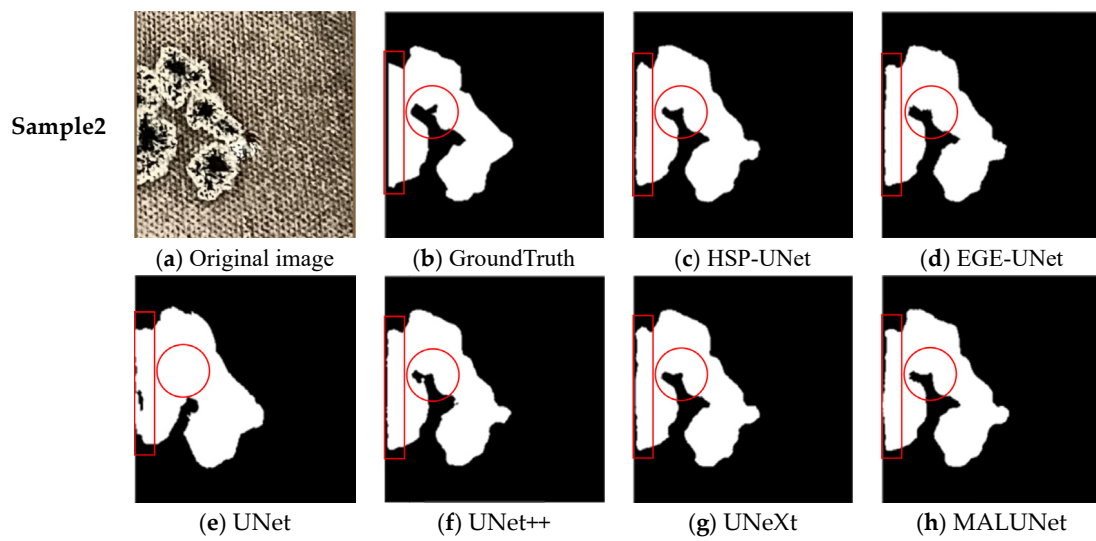

Figure 12. *Cont*.

(**a**) Original image   (**b**) GroundTruth   (**c**) HSP-UNet   (**d**) EGE-UNet

(**e**) UNet   (**f**) UNet++   (**g**) UNeXt   (**h**) MALUNet

Figure 12. Segmentation comparison of the clustered carbon traces.

（a）dendritic carbon trace with sufficient light   （b）dendritic carbon trace with insufficient light   （c）clustered carbon trace with sufficient light   （d）clustered carbon trace with insufficient light
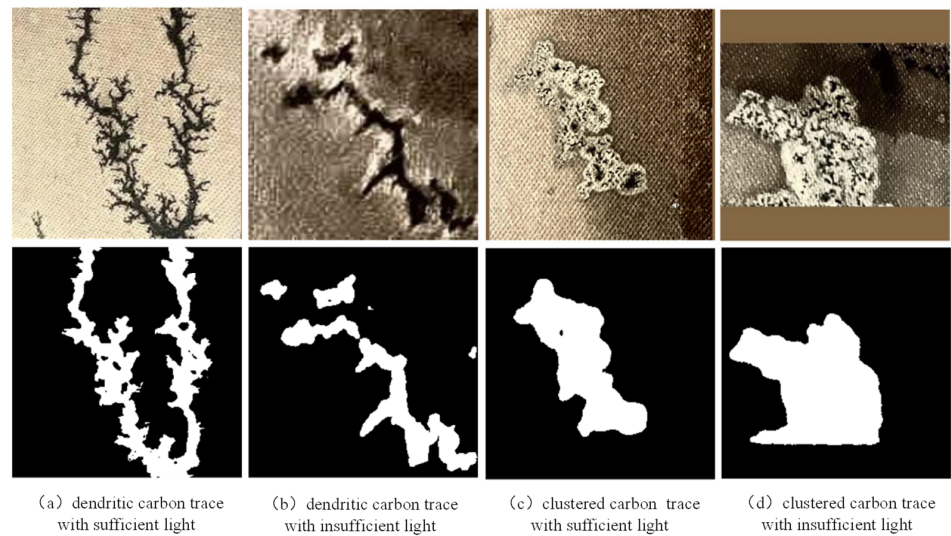
Figure 13. Segmentation performance with samples in different light conditions.

（a）large-size dendritic carbon trace   （b）small-size dendritic carbon trace   （c）large-size clustered carbon trace   （d）small-size clustered carbon trace
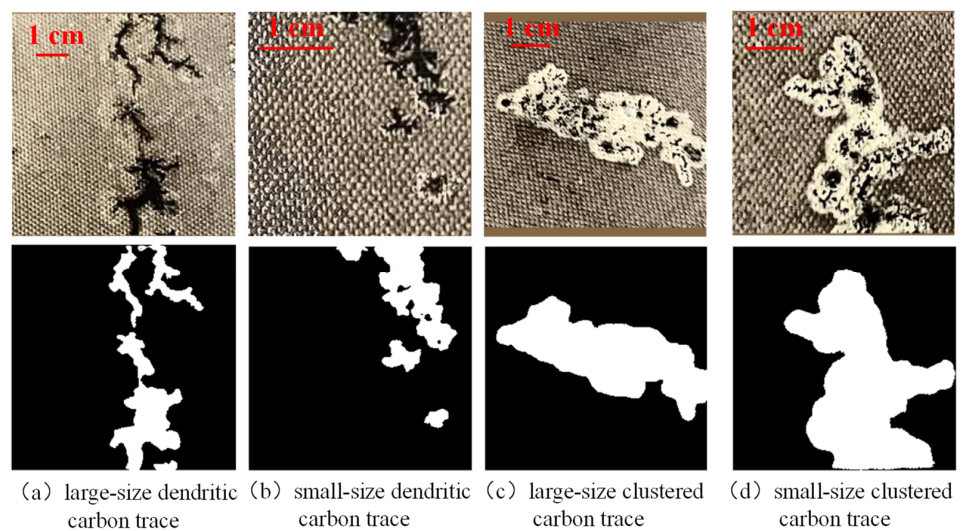
Figure 14. Segmentation performance with samples of different sizes.

5.5. Comparison of Different Attention Mechanism

After verifying the segmentation performance of the HSP-UNet, four attention mechanisms, SCA, SENet [23], CBAM [24], and ECA [29] were selected to compare the segmentation performance for carbon traces. Using the HP-UNet with an addition of the grouped HPA module and the PixelShuffle up-sampling module as the backbone network, the above four attention mechanisms were, respectively, added to analyze the segmentation effect, as shown in Table 2. For dendritic carbon traces with rich detailed features, all four attention mechanisms could improve the segmentation effect, indicating the feasibility of using attention mechanisms to improve the perception of detailed features. Among them, the SCA mechanism achieved the best improvement, with the MIoU, PA, and CPA improved by 2.19, 1.34, and 4.78 percentage points, respectively. For clustered carbon traces, the improvements of the four attention mechanisms were relatively small. The reason was that for the clustered carbon traces with smoother edges and fewer detailed features, the attention mechanisms could not give full play to their detail perception capability. Considering the segmentation needs of the dendritic and clustered carbon traces, the SCA mechanism was adopted in this paper.

Table 2. Segmentation comparison of four attention mechanisms.

Types	$Set_{dentritic}$			$Set_{cluster}$		
	I_{mIoU} (%)	P_A (%)	P_E (%)	I_{mIoU} (%)	P_A (%)	P_E (%)
HP-UNet	74.37	94.07	86.32	90.55	97.73	94.97
HP-UNet+SEnet	74.54	94.34	85.06	90.74	97.84	94.73
HP-UNet+CBAM	75.02	94.30	87.70	91.25	**98.12**	95.24
HP-UNet+ECA	75.34	94.41	87.72	91.35	98.01	**95.41**
HP-UNet+SCA	**75.66**	**94.41**	**89.10**	**91.39**	98.07	95.39

5.6. Ablation Tests for the HSP-UNet

The model validation and the comparative analysis of four attention mechanisms showed that the HSP-UNet proposed in this paper had the best segmentation performance of carbon traces. Taking the dendritic carbon trace with a higher segmentation difficulty as the object, the ablation test was carried out to analyze the contributions of the grouped HPA module, SCA, and the PixelShuffle. The ablation test would provide a reference for the subsequent model improvement and design. Using the UNet model as the benchmark, the test results were shown in Table 3, where $\sqrt{}$ indicated that the corresponding module was used. Compared with the benchmark UNet, the adoption of the grouped HPA module improved the MIoU, PA, and CPA of carbon trace segmentation by 0.61, 0.12, and 1.02 percentage points, respectively; the addition of the SCA could improve the MIoU, PA, and CPA by 0.79, 0.11, and 1.97 percentage points, respectively; and the adoption of the PixelShuffle module improved the MIoU, PA, and CPA by 0.20, 0.01, and 0.69 percentage points, respectively. Therefore, the SCA mechanism contributed the most to the performance improvement, the grouped HPA module the second, and the PixelShuffle module the least.

Table 3. Ablation results of the HSP-UNet.

Num	UNet	HPA	SCA	PixelShuffle	I_{mIou} (%)	P_A (%)	P_E (%)
1	$\sqrt{}$				73.53	93.17	84.42
2	$\sqrt{}$	$\sqrt{}$			74.47	94.29	86.44
3	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$		75.46	94.40	88.41
4	$\sqrt{}$	$\sqrt{}$		$\sqrt{}$	75.37	94.31	87.92
5	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$	75.66	94.41	89.10

Meanwhile, the Grad-CAM was used to visually compare and analyze the contribution of different modules to the segmentation effect of carbon traces. The multi-dimensional

feature map output from the last convolutional layer was utilized to generate the heat maps of the ablation test. With respect to the original image of Sample 1 in Figure 15a, the benchmark UNet had poor ability to perceive carbon traces in Figure 15b. The activation of the upper right carbon trace was low, resulting in an incomplete segmentation of the carbon trace region. The addition of the HPA module improved the segmentation completeness of the carbon trace, but the activation of the carbon trace region was still low, as shown in Figure 15c. The SCA mechanism dramatically improved the segmentation completeness and activation intensity of the carbon trace. However, the activation of the background region in the left part of the image was high, which may induce the model to misclassify the background as the carbon trace, as in Figure 15d. The PixelShuffle module had a lower activation intensity than the SCA mechanism, but it also reduced the activation of the background region, which was conducive to reducing the misclassification probability of the background, as in Figure 15e. The combined use of the above three modules could effectively improve the activation intensity of the carbon trace while reducing the activation intensity of the background region, which is conducive to improving the segmentation integrity and accuracy of carbon traces, as in Figure 15f.

Similar to Sample 1, there were obvious differences in the Grad-CAMs with different modules for Sample 2. For the benchmark UNet, the activation intensity of the lower part of the carbon trace was low, resulting in an incomplete segmentation of the carbon trace, as in Figure 15h. As shown in Figure 15i, the HPA module improved the perception completeness of the carbon trace, but the activation intensity of carbon trace edges was still low. This can lead to significant loss of detailed edge information and make the network misclassify the pixels near the carbon trace edges. In Figure 15j, the SCA module significantly improved the activation intensity and completeness of the carbon trace, but the activation intensity of the background near the carbon trace edge was a little high. It indicated that the perception of the carbon trace edge for the network with the SCA module needs to be further improved. As shown in Figure 15k, the PixelShuffle module can be a good complementary for the SCA module, because the activation intensity of the background near the carbon trace edges was lower than that with the SCA module, obtaining a much clearer boundary. Meanwhile, the overall activation intensity of the carbon trace with the PixelShuffle module was lower than that with the SCA module. Finally, by combining the three modules mentioned above into the benchmark UNet, the Grad-CAM obtained a good activation intensity and completeness with a clear boundary between the carbon trace and the background, as shown in Figure 15l. The above Grad-CAM results were in good agreement with the evaluation indexes in Table 3.

5.7. Discussion

Through the forementioned analysis, the proposed HSP-UNet outperformed over five State-of-the-Art segmentation models. But the segmentation performance on the dendritic carbon traces needs to be further improved. In subsequent studies, the following optimizations may be worth carried out: (1) The conventional convolution kernel in the grouped HPA module has a fixed rectangular inception field, which shows an insufficient adaptation to multi-scale complex edge features of the dendritic carbon traces. Owing to the deformable inception field, deformable convolution [30] may have a better feature extraction ability. (2) The U-shaped architecture is difficult to balance shallow spatial features and deep semantic features. Spatially detailed features are usually sacrificed to ensure the overall accuracy requirements of semantic segmentation, resulting in the need to improve the segmentation performance of the U-shaped model on the dendritic carbon traces. New model architectures, such as Bisenet series [31], may be an effective way to improve the segmentation performance with carbon traces.
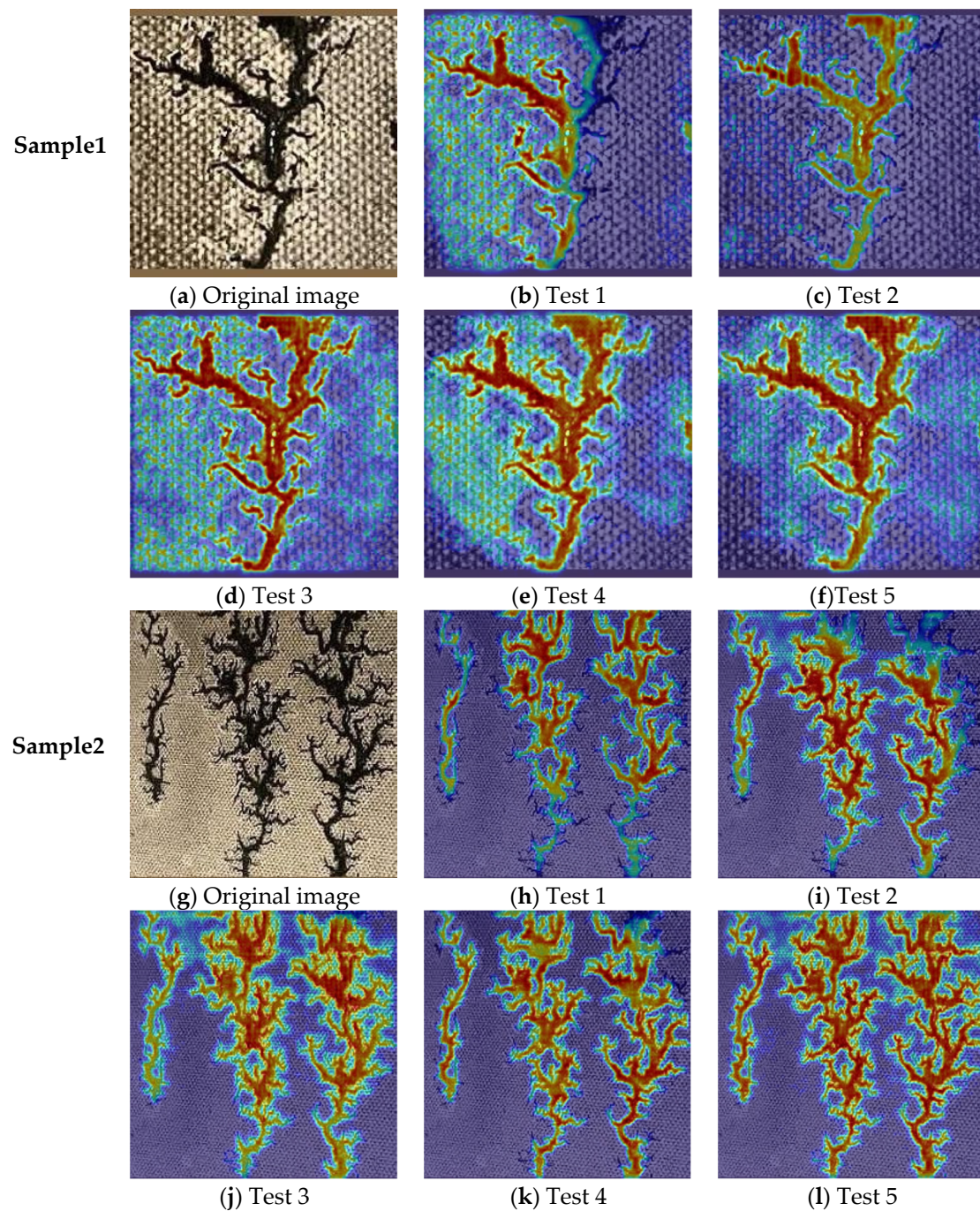
Sample1

(**a**) Original image   (**b**) Test 1   (**c**) Test 2

(**d**) Test 3   (**e**) Test 4   (**f**)Test 5

Sample2

(**g**) Original image   (**h**) Test 1   (**i**) Test 2

(**j**) Test 3   (**k**) Test 4   (**l**) Test 5

Figure 15. Grad-CAM comparison of the HSP-UNet ablation test.

6. Conclusions

Aiming at the accurate assessment of surface discharge inside the transformer, this paper constructed the HSP-UNet semantic segmentation network by means of an AHE-based image enhancement and network structure design and optimization based on the UNet, which achieved a good semantic segmentation of carbon traces with complex edge features. It would provide technical support for an accurate assessment of the transformer insulation condition.

(1) Aiming at the over-concentration of pixel values and the weak contrast of carbon trace images collected inside the transformer, the AHE algorithm was used for image enhancement, which effectively reduced the extraction difficulty of carbon trace

features. At the same time, four data augmentation methods were used to construct the dataset of dendritic carbon trace containing 2495 samples and the dataset of clustered carbon trace containing 2825 samples.

(2) With the goal of model lightweighting and accurate segmentation, the HSP-UNet model was constructed by integrating the grouped HPA module, SCA mechanism, and PixelShuffle module. Experimental results showed that the model parameter and GFLOPs were only 0.061 M and 0.066, respectively, which showed a good lightweighting advantage. Meanwhile, compared with the existing models, HSP-UNet had better segmentation on both carbon trace datasets. For dendritic carbon traces, HSP-UNet improved the MIoU, PA, and CPA of the benchmark UNet by 2.13, 1.24, and 4.68 percentage points, respectively. For clustered carbon traces, HSP-UNet improved the MIoU, PA, and CPA by 0.98, 0.65, and 0.83 percentage points, respectively. Similarly, the validation experiments with the samples of different light conditions and different size demonstrated a good generalization performance of the proposed HSP-UNet.

(3) Ablation experiments for dendritic carbon traces showed that the grouped HPA module, the SCA mechanism, and the PixelShuffle module adopted in the proposed HSP-UNet can all improve the segmentation effect. Due to the improvement in the ability to perceive detailed features, the SCA mechanism contributed the most to the model performance, improving the MIoU, PA, and CPA by 0.79, 0.11, and 1.97 percentage points, respectively.

References

1. Qi, B.; Ji, M.; Zheng, Y.P.; Zhu, K.H.; Pan, S.Y.; Zhao, L.J.; Li, C.R. Application status and development prospect of power internet of things technology in condition assessment of power transmission and transformation equipment. *High Volt. Eng.* **2022**, *48*, 3012–3031.
2. Wada, J.; Nakajima, A.; Miyahara, H.; Takuma, T.; Yanabu, S.; Okabe, S.; Kohtoh, M. Surface breakdown characteristics of silicone oil for electric power apparatus. *IEEE Trans. Dielectr. Electr. Insul.* **2006**, *13*, 830–837. [CrossRef]
3. Shroff, D.; Stannett, A. A review of paper aging in power transformers. *IEE Proc. C Gener. Transm. Distrib.* **1985**, *132*, 312–319. [CrossRef]
4. Li, S.; Huang, M.; Su, Y.X.; Li, S.R.; Shi, S.; Qi, B. Optimization of preparation parameters for cooperative objective of dielectric and breakdown properties of insulating paperboard. *High Volt. Eng.* **2023**, *49*, 1015–1025.
5. Huang, M.; Li, Y.R.; Wu, Y.Y.; Chen, L.J.; Qi, B. Equivalent nonlinear circuit model with interface charge and polarity effect for oil-paper composite insulation. *Trans. China Electrotech. Soc.* **2023**, *1*, 3422–3432.
6. *GB/T 1094.3—2003*; Power Transformers Part 3: Insulation Levels, Insulation Tests and External Insulation Air Gaps. China Electric Power Press: Beijing, China, 2003.
7. Liao, R.J.; Yan, J.M.; Yang, L.J.; Zhu, M.Z.; Sun, C. Characteristics of partial discharge-caused surface damage for oil-impregnated insulation paper. *Proc. CSEE* **2011**, *31*, 129–137.
8. Wechsler, K.; Riccitiello, M. Electric breakdown of a parallel solid and liquid dielectric system. *Trans. Am. Inst. Electr. Eng.* **1961**, *80*, 365–368. [CrossRef]
9. Wei, Y.H.; Yang, L.J.; Xu, Z.R. The rapid-development-type discharge failure and its damage characteristics to oil-paper insulation. *Trans. China Electrotech. Soc.* **2022**, *37*, 1020–1030.

10. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]

11. Shotton, J.; Winn, J.; Rother, C.; Criminisi, A. Texton boost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comput. Vis.* **2009**, *81*, 2–23. [CrossRef]

12. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.

13. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.

14. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]

15. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv* **2021**. [CrossRef]

16. Atek, S.; Mehidi, I.; Jabri, D.; Belkhiat, D.E. SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information. In Proceedings of the 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 8–9 May 2022.

17. Valanarasu, J.M.J.; Patel, V.M. UNeXt: MLP-based Rapid Medical Image Segmentation Network. *arXiv* **2022**, arXiv:2203.04967.

18. Chen, Y.; Ma, Y.; Zhang, J.; Li, M. Nested U-Net with attention mechanism for polyp segmentation in colonoscopy images. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *9*, 1–8.

19. Liu, X.; Tian, M.; Liang, J.Y. Image segmentation and yield prediction of densely planted cotton in Xinjiang of China using RCH-UNet. *Trans. Chin. Soc. Agric. Eng.* **2024**, *40*, 285–294.

20. Feng, X.; Zhang, X.L.; Wang, J.X. Multi crop classification extraction based on improved spatial-coordinate attention UNet. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 132–141.

21. Zhu, Y.; Huang, C. An Adaptive Histogram Equalization Algorithm on the Image Gray Level Mapping. *Physics Procedia* **2010**, *25*, 601–608. [CrossRef]

22. Ruan, J.; Xie, M.; Gao, J.; Liu, T.; Fu, Y. EGE-UNet: An Efficient Group Enhanced UNet for skin lesion segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023*; Springer: Cham, Switzerland, 2023; Volume 14223, pp. 481–490.

23. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]

24. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19.

25. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscatvey, NJ, USA; pp. 13713–13722.

26. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscatvey, NJ, USA.

27. Zhou, Z.; Siddiquee MM, R.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018. [CrossRef]

28. Ruan, J.; Xiang, S.; Xie, M.; Liu, T.; Fu, Y. MALUNet: A Multi-Attention and Light-weight UNet for Skin Lesion Segmentation. *arXiv* **2022**, arXiv:2211.01784v1.

29. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *arXiv* **2020**, arXiv:1910.03151v4.

30. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.

31. Yu, C.; Gao, C.; Wang, J.; Yu, G.; Shen, C.; Sang, N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]