

## Research Article

# Generation of Smoke Dataset for Power Equipment and Study of Image Semantic Segmentation

Rong Chang,<sup>1</sup> Zhengxiong Mao,<sup>2</sup> Jian Hu,<sup>2</sup> Haicheng Bai ,<sup>3</sup> Anning Pan ,<sup>4</sup> Yang Yang ,<sup>4</sup> and Shan Gao<sup>5</sup>

<sup>1</sup>Yuxi Power Supply Bureau, Yunnan Power Grid Corporation, Yuxi 653100, China

<sup>2</sup>Information Center, Yunnan Power Grid Co., LTD, Kunming 650032, China

<sup>3</sup>Network and Information Center, Yunnan Normal University, Kunming 650500, China

<sup>4</sup>School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China

<sup>5</sup>Guangzhou JianRuan Technology Co., Ltd., Guangzhou 650500, China

Correspondence should be addressed to Haicheng Bai; 3701@ynnu.edu.cn and Anning Pan; paninglw@163.com

Received 9 May 2023; Revised 12 November 2023; Accepted 22 November 2023; Published 11 January 2024

Academic Editor: Ramash Kumar K.

Copyright © 2024 Rong Chang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Fire in power equipment has always been one of the main hazards of power equipment. Smoke detection and recognition have always been extremely important in power equipment, as they can provide early warning before a fire breaks out. Compared to relying on smoke concentration for recognition, image-based smoke recognition has the advantage of being unaffected by indoor and outdoor environments. This paper addresses the problems of limited smoke data, difficult labeling, and insufficient research on recognition algorithms in power systems. We propose using three-dimensional virtual technology to generate smoke and image masks and using environmental backgrounds such as HDR (high dynamic range imaging) lighting to realistically combine smoke and background. In addition, to address the characteristics of smoke in power equipment, a dual UNet model named DS-UNet is proposed. The model consists of a deep and a shallow network structure, which can effectively segment the details of smoke in power equipment and handle partial occlusion. Finally, DS-UNet is compared with other smoke segmentation networks with similar structures, and it demonstrates better smoke segmentation performance.

## 1. Introduction

Electricity is one of the most important energy sources in modern human society. It is indispensable in daily life, industrial production, public facilities, economic development, and national defence. As a critical infrastructure for generating and transmitting electricity, electric power equipment provides the fundamental guarantee for the continuous supply of electricity.

With the increasing depletion of fossil fuels and the adverse effects on the environment, the demand for electricity, a clean energy source, has been increasing year by year. This poses a huge challenge for the continuous, safe, and stable operation of power equipment, which is the foundation for the generation and transmission of electrical energy. Every year, accidents caused by fires in electric

equipment occur around the world. The data show that the number of electrical fires still ranks first, accounting for as high as 50.4%. Fires in power equipment not only affect people's normal production and life but also cause casualties and economic losses [1–4]. Therefore, daily fire prevention and timely detection of fires are important means to ensure the safe operation of power equipment.

The main causes of electric equipment fires are short circuits, overloads, equipment malfunctions, damage to power lines, extreme weather, illegal operations, and electric equipment failures. Before generating an open flame, electric equipment generates a large amount of smoke. Therefore, accurate and timely identification of smoke is critical in preventing fires from becoming more severe. Currently, a large number of smoke detectors require a certain level of smoke concentration to be detected, resulting in poor

timeliness, and they also have the disadvantage of not being useable outdoors [5, 6]. With the widespread use of video surveillance equipment and drones in the power system [7, 8], these devices can provide a large amount of image data in real-time indoors and outdoors, providing the basic data for smoke identification through images. Therefore, identifying smoke based on images has a very good application prospect in the power system.

The simplest way to identify smoke in an image is to determine whether or not smoke is present, but this method cannot provide information on the location and size of the smoke, and therefore, research on this method is limited. The more common approach is to identify the location of the smoke and accurately segment the smoke, which involves object recognition and semantic segmentation. The former quickly identifies the location of the smoke in an image or video frame using an algorithm and marks the specific location with a B-Box [9]. The latter not only identifies the location of the smoke but also segments the smoke region. The former algorithm belongs to image object recognition and the latter belongs to image semantic segmentation. Obviously, accurate segmentation of the smoke can provide more useful information than just identifying the location of the smoke. This is because it not only provides the location of the smoke but also its size information and even estimates the concentration of the smoke in combination with other algorithms. However, this method is more challenging because it requires distinguishing smoke from background pixels and generating a smoke mask.

Traditional methods for smoke segmentation algorithms generally rely on the color, shape, motion, and other information of smoke. Therefore, algorithms such as motion detection, wavelet analysis, hidden Markov model (HMM), and histogram of oriented gradients (HOG) are widely used to extract smoke features [10–14]. However, traditional smoke image segmentation algorithms are affected by factors such as lighting, weather, background, and the semi-transparent nature of smoke, which leads to poor segmentation results. Neural networks can learn more abstract image feature information and have better robustness. In recent years, using neural networks for smoke segmentation has become a research hotspot, and it has achieved good results in smoke segmentation tasks in different fields [15, 16]. However, there is almost no research on smoke segmentation in the context of the power system, and there are three main problems in this regard.

First, the lack of data is a major issue for smoke segmentation in the field of electric power systems, as it is generally a common problem in the research area of smoke segmentation. Image segmentation neural networks learn image features from the input image dataset, and therefore, the dataset is an important factor that determines the final segmentation result. Fires in electric power equipment generally have small burning areas, high smoke density, and fast smoke generation, and the smoke color is mainly gray, black, and light blue. Therefore, it is different from other fields such as forest fire prevention and urban building fire prevention. In addition, smoke data in electric power

equipment are not easy to collect or obtain from the Internet as compared to other fields, and even creating a suitable environment to obtain these data is difficult due to cost-related reasons. Smoke in electric power equipment has higher density and smaller area than other fields such as forest fires. In previous smoke segmentation research, artificial synthesis methods were often used to supplement the lack of real images. This method usually involves directly overlaying smoke images with masks onto other background images. This two-dimensional method overlooks the fact that smoke can be affected by environmental lighting to a great extent when compared to real smoke. In fact, the color and brightness of smoke vary in different environments and lighting conditions. Moreover, the smoke images generated by this method are static and cannot provide continuously changing smoke images to the neural network. Therefore, there is still a significant gap between two-dimensional generated smoke and real smoke.

Second, there is an issue with the annotation of smoke images. Due to the complex and rapidly changing shape of smoke edges and its translucent nature, smoke annotation is difficult to be accurate, which affects the final segmentation effect.

Lastly, there is a lack of research on neural networks for smoke segmentation in the field of power equipment. Although there are many studies on smoke segmentation in other fields such as forest fire prevention, there are very few studies on smoke segmentation in power equipment. Unlike other fields such as forest fire prevention, the smoke generated by power equipment is more concentrated, with a higher concentration and a more diverse background environment including both indoor and outdoor settings. Therefore, the smoke segmentation results achieved in other fields cannot be easily translated to the smoke segmentation tasks in the field of power equipment.

This article proposes a solution to the problem of insufficient real smoke data for electric power equipment by using 3D virtual software to create realistic smoke data. In addition, a new neural network structure called DS-UNet (double smoke UNet) based on UNet is proposed, which is tailored to the characteristics of smoke in electric power equipment and achieves better smoke segmentation results.

The innovation of this article lies in proposing a method to address the issue of insufficient real smoke data for electric equipment smoke image segmentation by using 3D virtual software to create realistic smoke data. The software is used to simulate the realistic environment lighting with HDR, making the generated smoke consistent with the real smoke in the environment. In addition, more diverse background environments and occlusion are provided, and continuous changing smoke data (including the initial production stage of smoke) are generated to enhance the robustness of the neural network for electric equipment smoke segmentation.

We use 3D virtual software to generate precise mask images of smoke during the smoke generation process, eliminating the need for data annotation and, more importantly, improving the accuracy of subsequent segmentation.

We propose a DS-UNet neural network tailored to the characteristics of electric equipment smoke, achieving more accurate segmentation of electric equipment smoke.

## 2. Related Works

Long et al. [17] first achieved end-to-end semantic segmentation using fully convolutional networks (FCNs), marking a groundbreaking advancement in image semantic segmentation. Their key contribution was the incorporation of skip connections within the network to fuse outputs from different layers, resulting in improved segmentation results. In addition, they initiated training by directly initializing the network's parameters with a VGG16 model pretrained on ImageNet [18], followed by fine-tuning on other datasets to accelerate convergence. Since the introduction of FCN, numerous researchers have conducted extensive studies in image segmentation, including applications such as smoke segmentation. Yuan et al. [19], inspired by the concept of FCN for semantic segmentation, proposed a deep smoke segmentation network aimed at inferring high-quality segmentation masks from ambiguous smoke images. To address the significant variations in smoke appearance related to texture, color, and shape, they divided the network into coarse and fine pathways. The first pathway utilized an encoder-decoder FCN with skip connections, extracting global contextual information of the smoke to generate coarse segmentation masks. To preserve the fine spatial details of the smoke, the second pathway was also designed as an encoder-decoder FCN with skip connections, but was shallower compared to the first pathway. Ultimately, experimental results on three synthetic smoke datasets and one real smoke dataset demonstrated that the proposed method's segmentation performance significantly outperformed existing segmentation algorithms based on neural networks trained for ambiguous data.

Based on FCN, Ronneberger et al. [20] introduced the UNet deep learning image segmentation model in 2015. UNet's primary characteristic is its symmetric encoder-decoder architecture with skip connections. The encoder portion resembles a convolutional neural network (CNN), progressively reducing the input image size to extract features at different scales. The decoder portion reverses this operation, gradually upscaling the image to reconstruct finer details. Between these two parts, UNet incorporates skip connections, linking the output of the encoder with the input of the decoder to transmit both low-level and high-level feature information. Another distinctive feature of UNet is its loss function, typically a combination of cross-entropy and Dice coefficient. This combination effectively penalizes discrepancies between predicted and actual segmentation areas. Since the introduction of UNet, it has found extensive use in various classification tasks, including medical image segmentation [21–23]. In addition, UNet has been applied to smoke segmentation tasks as well [15, 24].

In addition to the aforementioned FCN and UNet, generative adversarial networks (GANs) and transformers have also been applied to image segmentation. Zhao et al. [25] employed GANs to enhance the contrast, sharpness,

and brightness of segmented images. They utilized a staged fine-tuning strategy, gradually fine-tuning layers of deep neural networks from top to bottom to achieve optimal segmentation results. Ultimately, they achieved state-of-the-art performance in the bone age assessment (BAA) task using the RSNA dataset. While transformers [26] have made significant advancements in natural language processing, they have also been applied in the field of computer vision [27, 28]. Zhou et al. [29] introduced a hybrid semantic segmentation algorithm named SCDeepLab. They combined Swin Transformer and CNN within the encoding and decoding framework of DeepLabv3+ to achieve accurate identification of tunnel lining cracks. This approach yielded excellent results in their study.

While transformers have demonstrated excellent performance in image classification and segmentation, they often require more training data. UNet, on the other hand, can achieve relatively good results with a smaller dataset, which is crucial for smoke image segmentation in cases of limited data availability. However, UNet still suffers from the issue of losing image details during the upsampling and downsampling processes. Consequently, accurately segmenting complex scenes such as occlusions and densely packed objects can be challenging in smoke segmentation. The dual-path network proposed by Yuan [19] and colleagues excels in extracting fine and coarse segmentation information from smoke data using neural networks. However, FCN's convolutional operations focus only on local information around the current pixel, lacking broader contextual information. Therefore, this paper introduces a dual-path network based on UNet to address the aforementioned challenges in smoke segmentation. Unlike Feiniu Yuan's [19] approach, we utilize a dual UNet structure rather than FCN. In addition, both networks adopt the UNet architecture instead of the VGG16 structure. To retain more smoke information, there are more skip connections between downsampling and upsampling layers. While this structure increases computational complexity compared to UNet, it yields superior segmentation results, making the added computational load worthwhile.

*2.1. Generating a 3D Virtual Smoke Dataset.* As mentioned earlier, there is a lack of publicly available smoke data for power equipment. In other smoke segmentation tasks, a common method is to directly overlay the smoke image with an alpha channel on the background image, as shown in Figure 1, to address this issue. However, this direct overlay method results in a very rough fusion of smoke and background. On the other hand, the smoke has a semi-transparent nature, and the alpha channel of the smoke image overlaid in this 2D way usually lacks information that can reflect this characteristic of smoke, making it difficult to blend with the background. Moreover, this method can only generate static smoke images and cannot provide continuously changing smoke. These issues are not conducive to the training of neural networks because they cannot realistically reflect the relationship between smoke and background in real images or the characteristics of smoke.

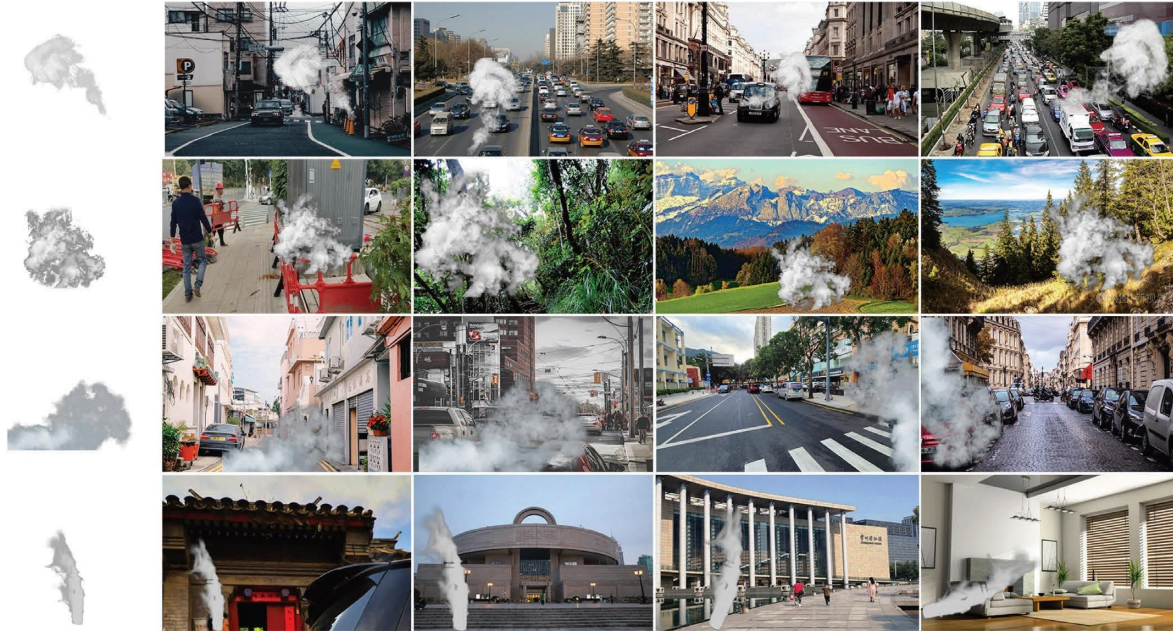


FIGURE 1: Two-dimensional composite images.

In order to address the aforementioned issues, we propose using three-dimensional virtual technology to generate realistic smoke as a method to overcome the lack of smoke data. Currently, the use of three-dimensional virtual technology to generate smoke is very mature and has been widely used in the film and television industry to create realistic smoke and fire in various productions. The smoke generated by three-dimensional virtual technology not only solves the shortcomings of two-dimensional smoke mentioned above but can also generate continuously animated smoke, which is extremely helpful for neural networks to learn the continuous changes in smoke.

There are many 3D software programs that can generate smoke. In this study, we use the open-source Blender [30] software to generate virtual and realistic smoke. Blender is a free and open-source 3D creation suite. It supports the entirety of the 3D pipeline—modelling, rigging, animation, simulation, rendering, compositing, and motion tracking, and even video editing and game creation. Blender can generate very realistic smoke effects, and various features such as the density and color of the smoke can be adjusted freely. To better integrate with the background environment, we chose to use the HDR images as background images and used them as light sources to illuminate the generated smoke. This allows the smoke to reflect the lighting characteristics of the background realistically, producing real light and shadow effects. At the same time, the smoke images also come with their own mask data, avoiding the problem of manual labeling. The detailed steps of generating smoke images in Blender are as follows and the process is shown in Figure 2:

- (1) Smoke generation: Blender can directly use the fluid module to generate realistic smoke. First, we create a smoke domain and then create a smoke generator inside the domain. We create an object with a similar

shape to the object that needs to produce smoke in the scene and use it as a smoke generator. Smoke will be produced from the surface of the object. At the same time, we set the parameters of the domain and smoke generator (see Figure 3). As shown in Figure 4, we created an object similar to an insulator as a smoke generator, which will be composited into the image of the power equipment.

- (2) To integrate the smoke into the power equipment image and achieve a realistic effect, we use different HDR images to illuminate the smoke and serve as the background for the rendering process. The smoke not only reflects the background lighting but also casts shadows onto the background, making it more realistic.
- (3) Rendering the image sequence: During the process, we set the rendering size to  $1000 \times 600$ . The image background and smoke are rendered and output together. The color mode is set to RGB.
- (4) Render the smoke mask series: After rendering the previous set of images, we separately render the smoke mask image as ground truth for segmentation training using DS-UNet. This way, we do not have to manually label this part of the image. As shown in Figure 4, some generated smoke images and their corresponding smoke masks are displayed. This method can generate static smoke images and produce continuous images. Figure 5 demonstrates the sequence of generated smoke images.

Obviously, compared with manual labeling of smoke data, using Blender to generate real smoke data avoids high data collection and labeling costs. Therefore, it greatly reduces time and manpower costs. Table 1 simply compares the time and manpower costs required for one person to

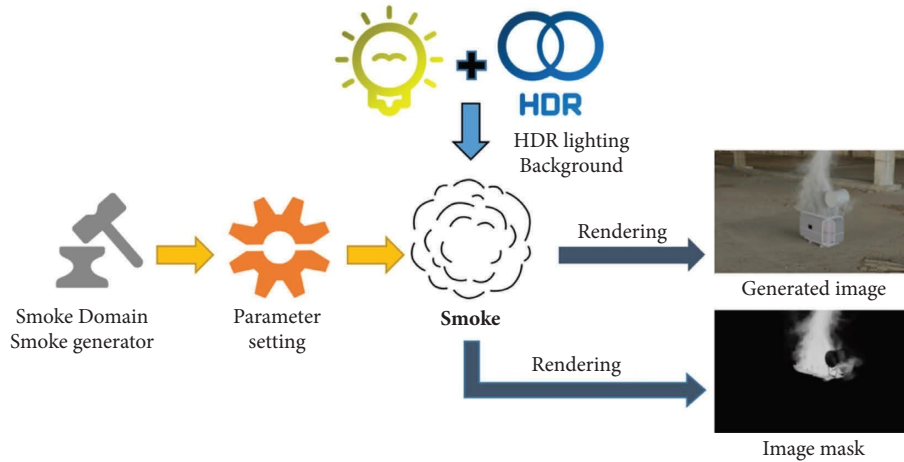


FIGURE 2: The process of generating smoke images in Blender.

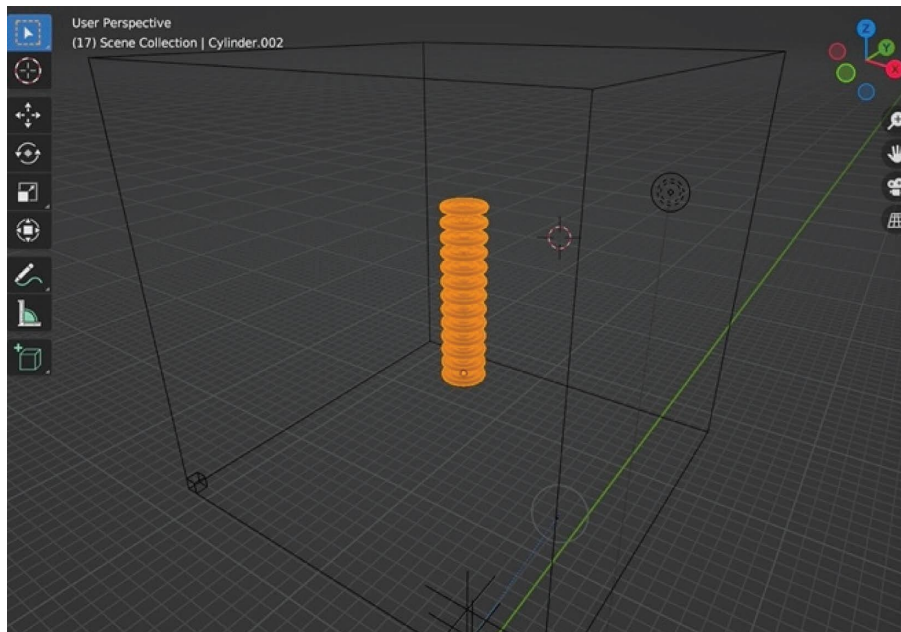


FIGURE 3: Creating an object similar to an insulator as a smoke generator.

complete 10 scenes and 5000 pieces of smoke data. It should be noted that the image acquisition process in manual data labeling is affected by various spatial factors such as scenes and regions, so the estimated time will be longer, which does not include transportation costs. In addition, due to the irregularity of the edges of smoke images, labeling will take longer than other types of data. According to experience, it takes about 1 minute per image, so 5000 images require about 83.3 hours, which is about 3.47 days. This is the ideal case, and normally, no one can work continuously for so long. From the table, we can see that the manpower cost of the manual data labeling method exists in every process, while the virtual data generation method requires manpower mainly in the process of creating virtual scenes, and rendering basically does not require much manpower, which is completely performed automatically by the computer (the computer hardware configuration required for rendering

time in the table: NVIDIA 3080, 32G memory, and Intel i7. If the computer has a better performance, rendering time will be shorter).

It should be noted that the smoke mask is a grayscale image, not a binary image. Therefore, before training, we need to convert the image to a binary image. We found that using a threshold of 100 can better preserve the edges of the smoke. In order to have better robustness for the neural network used later, in addition to the generated smoke images, we also have real images that are accumulated from daily work or obtained from the Internet. In order to obtain more accurate segmentation masks, we use Photoshop's smart selection tool to select and segment the real images. In the end, our dataset consists of a total of 6100 images, including 5100 generated images and 1000 real images. We will divide 3/4 of the generated images for the training set and 1/4 for the validation set and use the 1000 real images as the



FIGURE 4: The smoke synthesis images and image masks generated by 3D virtual technology.

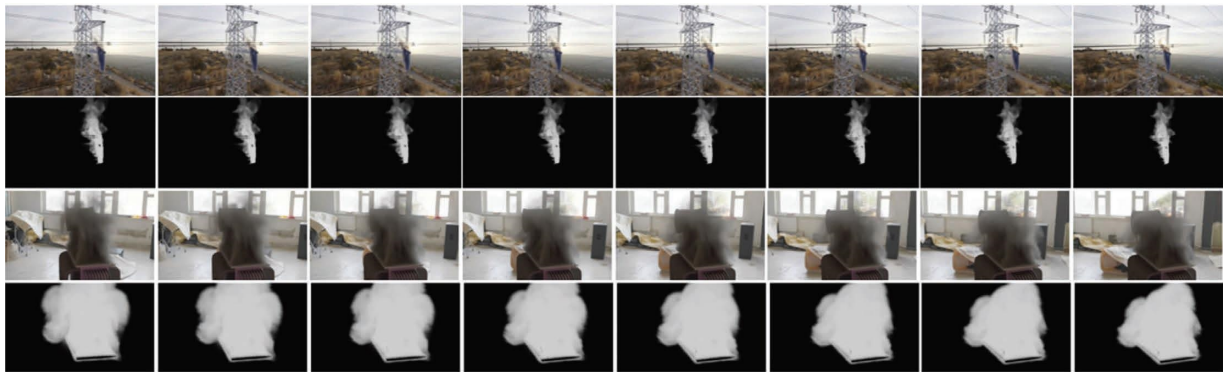


FIGURE 5: Generated sequence of images.

TABLE 1: Cost comparison.

Training	Step	Time	Manpower	Spatial influence
Manual data labeling	Image collection	10–20 days	Yes	Yes
	Data labeling	3.47 days	Yes	No
Virtual data generation	Creating virtual scenes	3–4 days	Yes	No
	Rendering	2–3 days	No	No

test set. In addition, in the training process, we use a series of image enhancement techniques to augment the dataset, such as rotation, flipping, cropping, and contrast changes. The detailed composition of the datasets is shown in Table 2. The data used to support this study are available at the following website: [https://github.com/baihch7982/power\\_smoke\\_datasets.git](https://github.com/baihch7982/power_smoke_datasets.git).

**2.2. Double Smoke UNet (DS-UNet).** Compared to other semantic segmentation networks, UNet has a good image segmentation effect with only a small amount of data and is widely used in various semantic segmentation tasks, especially in medical image segmentation. However, in the task of smoke segmentation, due to the semitransparent and unclear edge characteristics of smoke, using UNet alone cannot achieve good segmentation of smoke edges, and the

TABLE 2: Composition of the dataset.

Training	Validation	Testing	Total
3825	1275	1000	6100

model loses some details of smoke during the training process. This is because the UNet network extracts high-level abstract features of images through continuous downsampling and restores them through continuous upsampling. Although information is passed through skip connections similar to ResNet, a lot of information is still lost because the downsampling and upsampling processes are not reversible.

Inspired by Feiniu Yuan’s [19] dual-path network, we used a dual UNet to improve the UNet model for smoke segmentation and improve the segmentation performance of smoke. It should be noted that the only similarity between

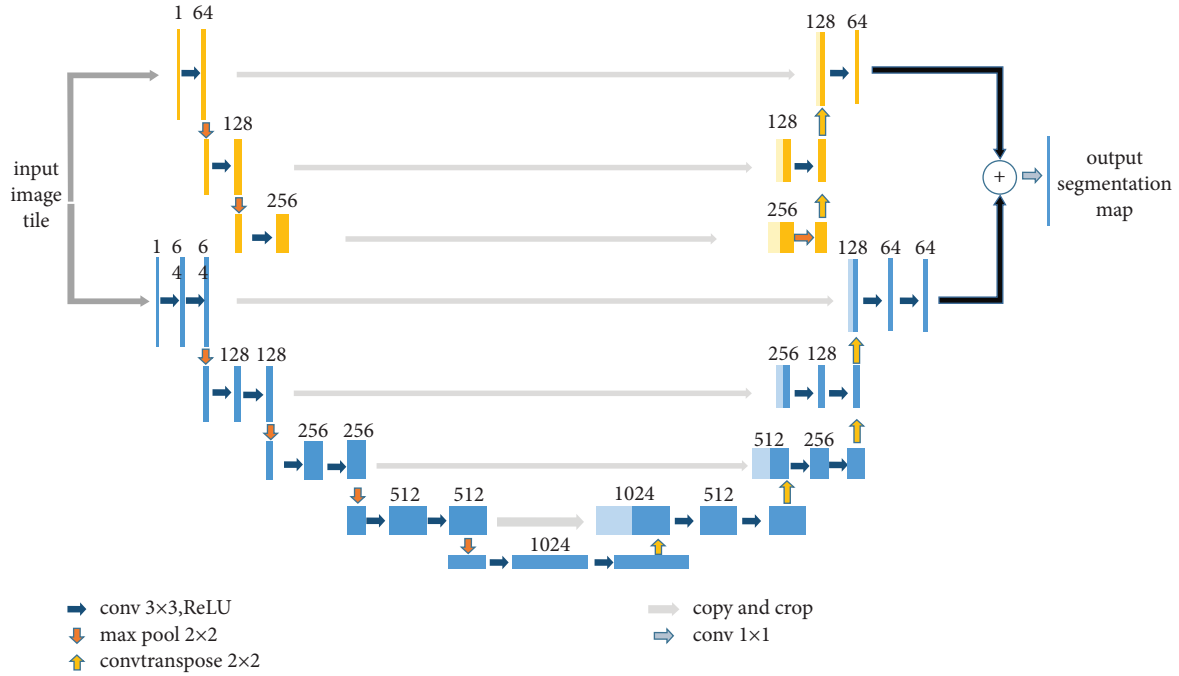


FIGURE 6: DS-UNet model network architecture.

our work and the work proposed by the authors in reference [19] is that both use a deep network to allow the model to learn high-level features of the image and a shallow network to allow the model to learn shallow information about the image. Apart from this, our network structure is completely different.

The DS-UNet model uses a fully convolutional structure, as shown in Figure 6. Like UNet, the deep network uses the UNet network structure. Therefore, after five identical downsampling operations, the input image is gradually reduced in scale, while the channel dimension changes in the order of 64, 128, 256, 512, and 1024. Then, the image scale is restored to the input size by upsampling. In this process, except for the bottom layer, the same layer is connected by skip connections through concatenation. The downsampling process is implemented using max pooling, while the upsampling process uses transpose convolution, rather than interpolation used in previous upsampling methods. The difference between the two is that upsampling completes the upsampling through interpolation without training parameters, while transpose convolution has parameters that can be trained. Although upsampling is faster without parameters, the trainable parameters in transpose convolution can better control the restoration process from a high abstract feature space.

The shallow network also uses the UNet network structure, but to preserve more information, the number of layers in the network should be shallow. At the same time, the number of convolutions in each layer is reduced from 2 to 1 (please see Figure 7 for the detailed network structure). The orange upper part is the shallow network for detailed feature extraction and the blue lower part is the deep network for high abstract feature extraction. After different downsampling and transpose convolution, the two networks

obtain the same size and dimension, and we concatenate them directly. The channel dimension is the sum of the output dimensions of the two networks. Finally, the final segmentation result is obtained by adjusting to the same dimension and size as the input image using a 1×1 convolution and taking the maximum probability value of each pixel by using sigmoid.

During the training of the whole network, we used PyTorch as the deep learning framework. To prevent the repetition of applying sigmoid to the output and increase prediction errors, we used binary cross-entropy loss as the network's loss function instead of binary cross-entropy loss with logits. The loss function is shown in the following equation:

$$\text{BCE loss} = -\frac{1}{n} \sum_{i=1}^n [y_i \cdot \log p(y_i) + (1 - y_i) \cdot \log(1 - p(y_i))]. \quad (1)$$

**2.3. Experiments.** PyTorch was used as the deep learning framework during training, and we trained the models using an NVIDIA RTX 3080 GPU with 8 GB of memory. The training was conducted for 300 epochs. The training parameters of the specific settings include epoch: 300, batch size: 16, IoU threshold: 0.2, initial learning rate: 0.01, image size: 500×300, and momentum: 0.936. Figure 7 shows the loss-epoch curve for DS-UNet.

To evaluate the performance of our proposed DS-UNet, we compared it with UNet and Feiniu Yuan's [19] dual-path FCN segmentation algorithm. We trained these comparative methods on the same power equipment smoke training data using the code reproduced according to the respective papers.

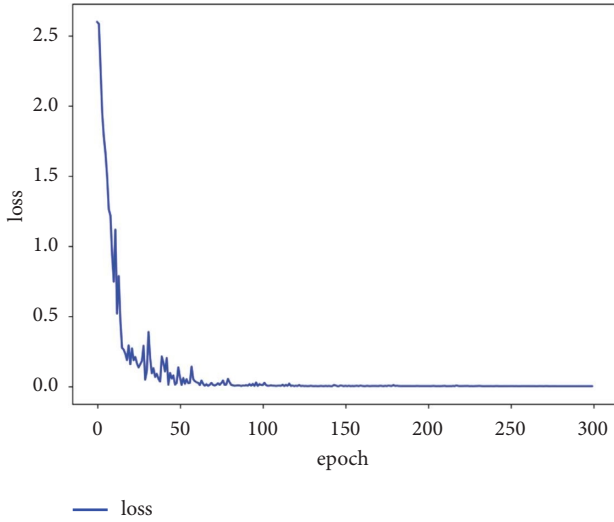


FIGURE 7: The loss-epoch curve for DS-UNet.

After training, we tested the mean pixel accuracy (mPA) of the three models using both the validation dataset and the test dataset. The results are shown in Table 3. We observed that DS-UNet consistently outperformed the other three models.

The evaluation metrics for image segmentation generally include mean intersection over union (mIoU) and Dice coefficient. Therefore, in order to further compare our algorithm with other algorithms, we calculated the widely used metrics: mIoU and Dice coefficient. For mIoU, the closer the value is to 1, the better the segmentation effect. The calculation formula for mIoU and Dice are shown in the following equations:

$$mIoU = \frac{1}{n} \sum_{i=1}^n \frac{PR_i \cap GT_i}{PR_i \cup GT_i}, \quad (2)$$

$$Dice = \frac{2|PR \cap GT|}{|PR| + |GT|}, \quad (3)$$

where  $PR_i$  denotes the predicted segmentation result of the  $i$ th image,  $GT_i$  denotes the corresponding ground truth, and  $n$  denotes the total number of images in the dataset.

As shown in Tables 4 and 5, as we can see, our method performs significantly better than the other methods on both the validation dataset and test dataset. Our method achieves the highest mIoU among all the compared methods, indicating that our predicted segmentation is closest to the ground truths. It should be noted that if the shallow network shown in the orange color in the upper part of the figure is removed, our model will degrade to the UNet model. As can be seen from Tables 4 and 5, adding the shallow network has a significant effect on the segmentation performance.

Typically, the complexity of a neural network model is described using two metrics: the number of parameters and the computational cost (time complexity). Therefore, in order to further illustrate the complexity of DS-UNet, we have counted the number of parameters as well as the amount of data used for inference (Table 6). As shown in Table 6, the number of parameters in the model reaches

TABLE 3: Comparison of image segmentation accuracy of different models on the validation set.

Models	mPA (validation)	mPA (test)
UNet	93.27	86.32
Feiniu Yuan	94.45	90.17
DS-UNet	95.31	92.33

TABLE 4: Comparison of mIoU among different models.

Models	mIoU (validation)	mIoU (test)
UNet	0.6713	0.6754
Feiniu Yuan	0.7046	0.6662
DS-UNet	0.7221	0.6934

TABLE 5: Comparison of Dice among different models.

Models	Dice (validation)	Dice (test)
UNet	0.8033	0.8063
Feiniu Yuan	0.8267	0.7997
DS-UNet	0.8386	0.8189

TABLE 6: The number of parameters and computational cost (time complexity) of DS-UNet.

Categories	Values
Total params	31,037,697
Trainable params	31,037,697
Nontrainable	0
Input size (MB)	3.00
Forward/backward pass size (MB)	3718.00
Params size (MB)	118.40
Estimated total size (MB)	3839.40
Parameter count	31,037,697

31037697, and the amount of data used for forward inference and backward inference reaches 3718.00. Due to the addition of branches, both the number of parameters and the amount of data used for inference increase. However, we believe that this is worthwhile in order to achieve better smoke segmentation results.

At the same time, we selected four data images and created heat maps. As shown in Figure 8, we can see that the model has accurately identified the smoke regions in the images. This further demonstrates DS-UNet's ability to accurately recognize smoke images in an electric equipment environment.

Smoke is different from other objects, such as cars and pedestrians (their edges are clear). It has the characteristics of constantly changing shape, transparency, and blurred edges. Therefore, in daily use, it is very important to correctly segment continuous smoke images (such as videos). To test the segmentation ability of DS-UNet for continuous smoke data, we selected three different scenes with a duration of 1 minute (30 seconds/frame, a total of 1800 frames) for video data testing. The test results show that DS-UNet has a good segmentation effect for continuous smoke, as shown in Figure 9, which displays the segmentation results of 1 second and 30 frames from these three scenes.





FIGURE 8: Heatmap examples of DS-UNet.



FIGURE 9: Segmentation effect of DS-UNet on continuous smoke images (two lines per scene).

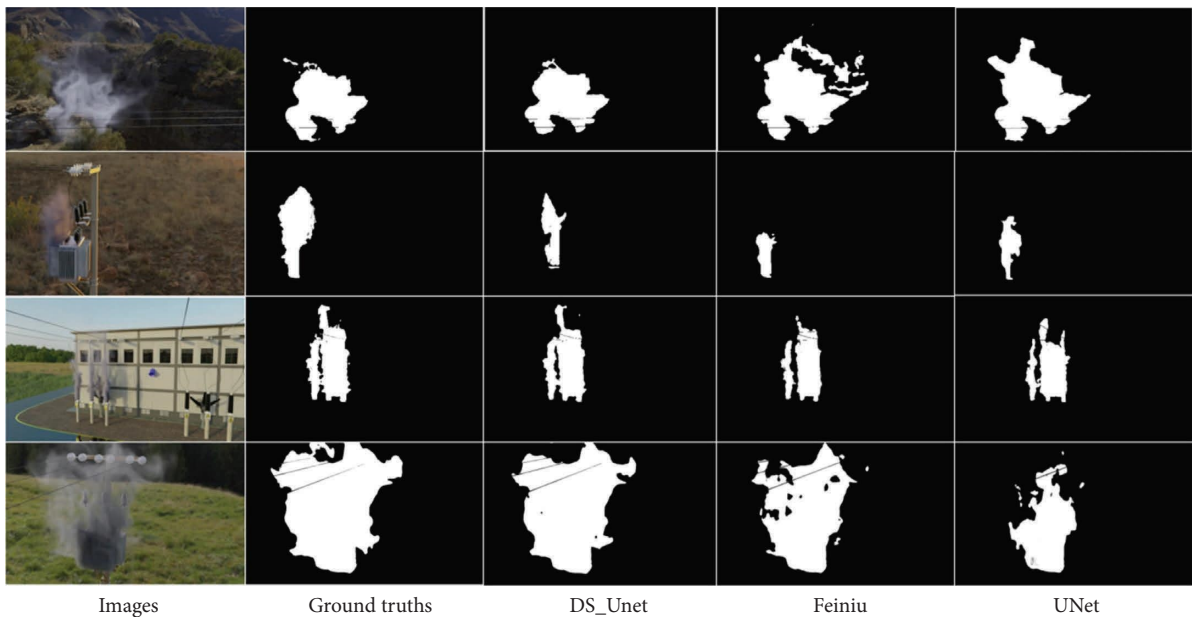


FIGURE 10: Comparison of segmentation results on generated smoke images and real images (the first two rows are generated images and the last two rows are real images).

The qualitative comparison between our method and other deep learning methods on the smoke test dataset of power equipment is shown in Figure 10. The first column shows the test images, the second column shows the

corresponding ground truths, and the other columns show the segmentation results of different methods. These images include both the generated images by our method (first two rows) and real images (last two rows). The experimental

results demonstrate that our method has good segmentation performance on both synthetic and real smoke images. Moreover, our method separates the smoke with clearer edges and more accurate positions compared with other methods.

### 3. Conclusion

Electric equipment fires have always been one of the main hazards of electric equipment. Smoke detection and recognition have always been extremely important in electric equipment, as they can provide early warning before an open flame occurs. Compared to relying on smoke concentration for detection, image-based smoke recognition has the advantage of being unaffected by indoor and outdoor environmental conditions. This paper addresses the problems of limited smoke data in the electrical system, difficulty in labeling data, and inadequate research on recognition algorithms. We propose using a 3D virtual technology to generate smoke and image masks and using the environment background such as HDR lighting to enable smoke to be realistically combined with the background. Inspired by the dual-path networks [19], we also propose the DS-UNet model. The model uses two UNet structures, one deep structure network for extracting abstract features of the data and one shallow structure network for extracting detailed features of the data. We trained the model on generated smoke data and conducted experiments on the generated and real images. Comparative experiments showed that the DS-UNet model has a significantly better smoke segmentation performance in electric equipment than other similar models.

### Data Availability

The data used to support the findings of this study are available at [https://github.com/baihch7982/power\\_smoke\\_datasets.git](https://github.com/baihch7982/power_smoke_datasets.git).

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- [1] A. Kelley, "California power company says a faulty wire may have sparked Silverado Fire, Changing America," 2020, <https://thehill.com/changing-america/resilience/natural-disasters/523141-california-power-company-says-a-faulty-wire-may/>.
- [2] National Fire Protection Association, "National fire protection association. (n.d.)," 2023, <https://www.nfpa.org>.
- [3] Electrical Safety Foundation International, "The electrical safety foundation international. (n.d.)," 2023, <https://www.esfi.org/>.
- [4] United States Consumer Product Safety Commission, "The United States consumer product safety commission. (n.d.)," 2023, <https://www.usa.gov/>.
- [5] R. M. Nesse, "The smoke detector principle: natural selection and the regulation of defensive responses," *Annals of the New York Academy of Sciences*, vol. 935, no. 1, pp. 75–85, 2001.
- [6] R. M. Nesse, "Natural selection and the regulation of defenses: a signal detection analysis of the smoke detector principle," *Evolution and Human Behavior*, vol. 26, no. 1, pp. 88–105, 2005.
- [7] M. Alsafasfeh, I. Abdel-Qader, B. Bazuin, Q. Alsafasfeh, and W. Su, "Unsupervised fault detection and analysis for large photovoltaic systems using drones and machine vision," *Energies*, vol. 11, no. 9, p. 2252, 2018.
- [8] M. N. Boukoberine, Z. Zhou, and M. Benbouzid, "Power supply architectures for drones-a review," in *Proceedings of the IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, pp. 5826–5831, Lisbon, Portugal, October 2019.
- [9] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, <https://arxiv.org/abs/2207.02696>.
- [10] D. Kim and Y.-F. Wang, "Smoke detection in video," in *Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering*, pp. 759–763, Washington, DC, USA, April 2009.
- [11] C. Yuan, Z. Liu, and Y. Zhang, "Learning-based smoke detection for unmanned aerial vehicles applied to forest fire surveillance," *Journal of Intelligent and Robotic Systems*, vol. 93, no. 1-2, pp. 337–349, 2019.
- [12] Z. Na, W. Huiqin, and H. Yan, "Smoke image segmentation algorithm based on rough set and region growing," *Journal of Frontiers of Computer Science and Technology*, vol. 11, p. 1296, 2017.
- [13] Y. Zhao, "Candidate smoke region segmentation of fire video based on rough set theory," *Journal of Electrical and Computer Engineering*, vol. 2015, Article ID 280415, 8 pages, 2015.
- [14] A. Filonenko, D. C. Hernández, and K.-H. Jo, "Fast smoke detection for video surveillance using CUDA," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 725–733, 2018.
- [15] W. S. Mseddi, R. Ghali, M. Jmal, and R. Attia, "Fire detection and segmentation using YOLOv5 and U-net," in *Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 741–745, Dublin, Ireland, August 2021.
- [16] S. Frizzi, M. Bouchouicha, J. Ginoux, E. Moreau, and M. Sayadi, "Convolutional neural network for smoke and fire semantic segmentation," *IET Image Processing*, vol. 15, no. 3, pp. 634–647, 2021.
- [17] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Boston, MA, USA, June 2015.
- [18] O. Russakovsky, J. Deng, H. Su et al., "ImageNet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [19] F. Yuan, L. Zhang, X. Xia, B. Wan, Q. Huang, and X. Li, "Deep smoke segmentation," *Neurocomputing*, vol. 357, pp. 248–260, 2019.
- [20] O. Ronneberger, P. Fischer, T. Brox, and U-net, "Convolutional networks for biomedical image segmentation," in *Proceedings of the 2015 Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference*, pp. 234–241, Munich, Germany, October 2015.
- [21] L. Yu, P. Wang, X. Yu, Y. Yan, and Y. Xia, "A holistically-nested U-Net: surgical instrument segmentation based on convolutional neural network," *Journal of Digital Imaging*, vol. 33, no. 2, pp. 341–347, 2020.

- [22] T. Falk, D. Mai, R. Bensch et al., “U-Net: deep learning for cell counting, detection, and morphometry,” *Nature Methods*, vol. 16, no. 1, pp. 67–70, 2019.
- [23] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, “U-net and its variants for medical image segmentation: a review of theory and applications,” *IEEE Access*, vol. 9, pp. 82031–82057, 2021.
- [24] Y. Zheng, Z. Wang, B. Xu, and Y. Niu, “Multi-scale semantic segmentation for fire smoke image based on global information and U-net,” *Electronics*, vol. 11, no. 17, p. 2718, 2022.
- [25] C. Zhao, J. Han, Y. Jia, L. Fan, and F. Gou, “Versatile framework for medical image processing and analysis with application to automatic bone age assessment,” *Journal of Electrical and Computer Engineering*, vol. 2018, 13 pages, 2018.
- [26] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [27] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” 2020, <https://arxiv.org/abs/2010.11929>.
- [28] Z. Liu, Y. Lin, Y. Cao et al., “Swin transformer: hierarchical vision transformer using shifted windows,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, Montreal, Canada, October 2021.
- [29] Z. Zhou, J. Zhang, and C. Gong, “Hybrid semantic segmentation for tunnel lining cracks based on Swin Transformer and convolutional neural network,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 17, pp. 2491–2510, 2023.
- [30] Blender Software, “The Blender Software (n.d.),” 2023, <https://www.blender.org/about/>.