

## Article

# Deep Learning-Based Infrared Image Segmentation for Aircraft Honeycomb Water Ingress Detection

Hang Fei <sup>1,\*</sup> , Hongfu Zuo <sup>1,\*</sup>, Han Wang <sup>2</sup>, Yan Liu <sup>1</sup>, Zhenzhen Liu <sup>3</sup>  and Xin Li <sup>1,4</sup>

<sup>1</sup> Civil Aviation Key Laboratory of Aircraft Health Monitoring and Intelligent Maintenance, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; jiainyin@nuaa.edu.cn (Y.L.); lixin1990@nuaa.edu.cn (X.L.)

<sup>2</sup> Shanghai Aeronautical Measurement and Control Research Institute, Shanghai 201601, China; wanghan07@nuaa.edu.cn

<sup>3</sup> School of Mechanical Engineering, Anhui University of Technology, Ma'anshan 243032, China; lzz1201@nuaa.edu.cn

<sup>4</sup> School of Automotive and Rail Transit, Nanjing Institute of Technology, Nanjing 211167, China

\* Correspondence: feihang@nuaa.edu.cn (H.F.); rms@nuaa.edu.cn (H.Z.)

**Abstract:** The presence of water accumulation on aircraft surfaces constitutes a considerable hazard to both performance and safety, necessitating vigilant inspection and maintenance protocols. In this study, we introduce an innovative semantic segmentation model, grounded in deep learning principles, for the precise identification and delineation of water accumulation areas within infrared images of aircraft exteriors. Our proposed model harnesses the robust features of ResNet, serving as the foundational architecture for U-Net, thereby augmenting the model's capacity for comprehensive feature characterization. The incorporation of channel attention mechanisms, spatial attention mechanisms, and depthwise separable convolution further refines the network structure, contributing to enhanced segmentation performance. Through rigorous experimentation, our model surpasses existing benchmarks, yielding a commendable 22.44% reduction in computational effort and a substantial 38.89% reduction in parameter count. The model's outstanding performance is particularly noteworthy, registering a 92.67% mean intersection over union and a 97.97% mean pixel accuracy. The hallmark of our innovation lies in the model's efficacy in the precise detection and segmentation of water accumulation areas on aircraft skin. Beyond this, our approach holds promise for addressing analogous challenges in aviation and related domains. The enumeration of specific quantitative outcomes underscores the superior efficacy of our model, rendering it a compelling solution for precise detection and segmentation tasks. The demonstrated reductions in computational effort and parameter count underscore the model's efficiency, fortifying its relevance in broader contexts.

**Keywords:** aircraft skin; water accumulation detection; infrared image segmentation; deep learning



**Citation:** Fei, H.; Zuo, H.; Wang, H.; Liu, Y.; Liu, Z.; Li, X. Deep Learning-Based Infrared Image Segmentation for Aircraft Honeycomb Water Ingress Detection. *Aerospace* **2024**, *11*, 961. <https://doi.org/10.3390/aerospace11120961>

Academic Editor: Wim J. C. Verhagen

Received: 30 September 2024

Revised: 14 November 2024

Accepted: 19 November 2024

Published: 22 November 2024



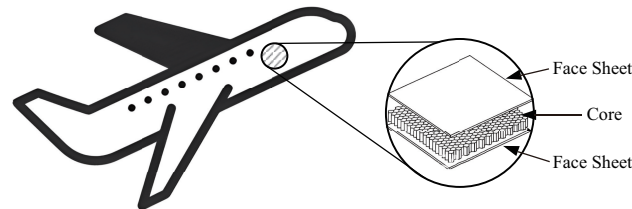
**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The aircraft skin is a critical component of an aircraft, with its structure illustrated in Figure 1. During flight operations, it is subjected to various complex loads and environmental influences, rendering it susceptible to defects such as dents, cracks, corrosion, and others [1]. Among these defects, water ingress in aircraft honeycomb structures is particularly hazardous, as it can lead to a deterioration in structural performance and potentially catastrophic accidents [2]. Swiftly identifying areas of water accumulation within the honeycomb panels and accurately assessing their impact on structural strength and aerodynamic performance can help promptly identify and eliminate potential risks. This, in turn, extends the aircraft's lifespan and reduces maintenance costs.

Currently, there are several methods commonly used for detecting water accumulation on aircraft skin, including the ultrasonic method, resistance method, capacitance method, X-ray method, and others [3]. However, these methods have various limitations, such as

low efficiency, complicated operation, and expensive equipment. Recently, infrared thermal imaging technology has emerged as a promising non-destructive testing technique for inspecting water accumulation on aircraft skin. This technology uses infrared radiation to measure surface temperatures and employs image processing techniques to analyze damage distribution. In comparison to other methods, infrared thermal imaging offers several advantages, including non-contact and non-destructive testing, high sensitivity, efficiency, and rapid sweeping capabilities [4].



**Figure 1.** Structure of aircraft skin.

Although infrared inspection methods are used for off-airport inspections and are deployed in many domestic and international airports, there are still numerous technical challenges, especially concerning image processing and accurate segmentation of water accumulation areas from infrared images of aircraft skin. Accurate segmentation is crucial as it forms the basis for identifying and repairing skin damage. Infrared images possess inherent characteristics of low quality, high noise, and poor contrast, making it difficult to accurately segment water accumulation areas from the aircraft skin surface using traditional image segmentation methods. These methods may also misidentify background and cold-end parts as water accumulation areas due to radiation similarity.

Our proposal, therefore, conducts a comprehensive study on the segmentation of infrared images of aircraft skin water accumulation, along with the development of effective infrared image segmentation methods, to address these challenges.

The remainder of this paper is organized as follows: Section 2 provides a brief overview of related work. Section 3 describes the proposed method in detail. Section 4 shows the experimental setup and evaluation metrics. Section 5 presents the experimental results and analysis. Section 6 discusses the advantages and disadvantages of the proposed model. Finally, Section 7 concludes the paper.

## 2. Related Work

Image segmentation is a fundamental task in the field of computer vision. Its primary objective is to subdivide an image into multiple sub-regions with similar or identical features [5]. This technique finds extensive use across several domains, including target detection [6,7], medical image analysis [8–10], autonomous driving [11], robot navigation [12], and more [13,14]. Infrared image segmentation delineates targets from backgrounds in infrared images via algorithms. However, infrared images possess inferior resolution and contrast compared to visible images [15]. They also lack rich texture information, which poses many challenges for infrared image segmentation.

Traditional image segmentation methods mainly include thresholding [16], edge detection [17], region growing [18], watershed [19], etc. Neves et al. [20] used a combination of wavelet transform and watershed transform to segment infrared images of ship surveillance, which can be robust to the noise commonly found in infrared images. Zhou et al. [21] proposed an infrared image segmentation method based on Otsu and genetic algorithm, which significantly improved the quality of image segmentation. Cui et al. [22] suggested using edge detection for infrared image segmentation to facilitate state detection and fault diagnosis of petrochemical thermal equipment. The results demonstrated that the optimal threshold edge detection operator was ideal for infrared image target extraction. Zhang et al. [23] proposed a generalized gradient vector flow snake model for infrared image segmentation, which outperformed other snake models with respect to infrared image

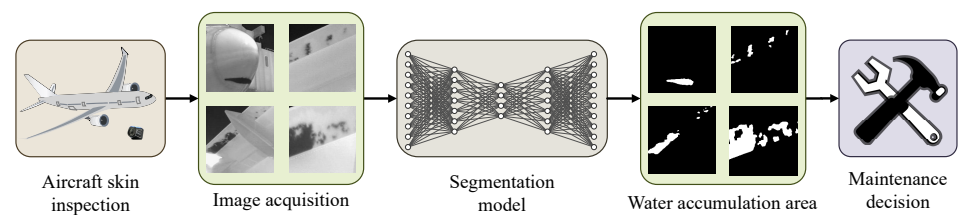
segmentation by combining generic gradient vector flow (GGVF) and normally biased gradient vector flow (NBGVF) approaches. Liu et al. [24] applied the two-dimensional maximum inter-class variance method for infrared image segmentation and introduced an improved particle swarm algorithm to accelerate the computational time while ensuring effective segmentation. Although traditional methods can yield promising results, they are susceptible to noise and background interference and often depend on manual feature selection and thresholding. Furthermore, they are not capable of handling large-scale images or complex object shapes.

For the past few years, the wide application of deep learning techniques in image processing has provided a new paradigm for image segmentation [25], especially convolutional neural network [26]. Deep learning methods automatically learn the features and representations of images by using convolutional neural network models to achieve more accurate, robust, and efficient image segmentation results. In 2015, Long et al. [27] introduced an end-to-end image segmentation model called a fully convolutional network (FCN). This model replaced the fully connected layer with a convolutional layer, enabling it to directly produce pixel-level segmentation results. FCN laid the groundwork for subsequent image segmentation models. Also in 2015, Ronneberger et al. [28] proposed U-Net, a semantic segmentation convolutional neural network. U-Net utilizes an encoder–decoder structure and skip connections to better preserve high-resolution information and contextual features. It has found widespread application in medical image segmentation and other fields. In 2016, Paszke et al. [29] presented ENet, a lightweight real-time semantic segmentation network. ENet achieved efficient image segmentation by reducing network parameters and adopting a fast structure, making it suitable for real-time segmentation on mobile and embedded devices. In 2017, the Google team introduced a series of image semantic segmentation models under the name DeepLab [30]. These models utilized dilated convolution to extend the receptive field and incorporated conditional random fields (CRFs) to enhance segmentation accuracy. DeepLab achieved significant improvements in both accuracy and speed. In the same year, Mask R-CNN proposed by He et al. [31] and PSPNet proposed by Zhao et al. [32] also contributed to the performance improvements in image segmentation models. In 2020, Sun et al. [33] proposed a network structure called HRNet, which focused on retaining high-resolution information. This structure significantly improved the accuracy of image segmentation by constructing multiple branches at different resolutions.

To date, many scholars at home and abroad have conducted research on infrared image segmentation based on deep learning. Zhang et al. [34] developed Res-UNet, an infrared image segmentation model that utilizes residual learning, transfer learning, and U-Net to detect photovoltaic panel errors. Nan et al. [35] proposed a lightweight real-time infrared image instance segmentation algorithm model based on Tiny YOLO, which achieved high segmentation accuracy for small targets. Chen et al. [36] introduced R-Net, a new deep learning-based infrared image segmentation method that effectively handles low-resolution, weakly illuminated, and noise-interrupted infrared images for better segmentation results. Yu et al. [37] integrated a hierarchical-split depthwise separable convolutional block into a U-Net network for wind turbine blade infrared image segmentation, which achieved higher accuracy. Zhao et al. [38] improved the DeepLabv3+ semantic segmentation algorithm by utilizing the convolutional block attention module (CBAM) to accurately segment armor targets from complex battlefield environments.

Although the methods mentioned above can achieve satisfactory segmentation results, there is limited research on deep learning-based infrared image segmentation for detecting water accumulation in aircraft skin. To address this gap, this study proposes an infrared image segmentation model for aircraft skin water detection. The detailed operation process is depicted in Figure 2. This model utilizes ResNet [39] as the feature extractor of the U-Net backbone and introduces CBAM to enhance its performance. Additionally, the structure of the network model is optimized using depthwise separable convolutional blocks. In comparison to the baseline Res-UNet, our proposed model exhibits a notable reduction in computational complexity by 22.44% and a decrease in parameter count by 38.89%.

Moreover, there is an improvement in the mean intersection over union by 1.41% and an increase in the mean pixel accuracy by 0.32%.



**Figure 2.** Schematic of aircraft skin water accumulation detection via deep learning.

The main contributions of this paper are as follows:

1. We propose an improved U-Net-based method for infrared image segmentation of aircraft skin water accumulation areas, which has not been previously reported.
2. We enhance the model's robustness and accuracy by introducing a channel and spatial attention mechanism to the network.
3. We optimize the complexity of the model by replacing traditional convolutions with depthwise separable convolutions.
4. We construct a dataset of infrared images for aircraft skin water accumulation areas and evaluate the performance of the proposed method through comparative experiments.

### 3. Methodology

#### 3.1. Network Architecture

The proposed network architecture constitutes the core algorithm of this paper illustrated in Figure 3, which comprises two parts: the encoding process on the left contraction path and the decoding process on the right expansion path. The network's encoder adopts ResNet as the backbone to enhance its feature extraction ability. To further improve performance, the channel attention mechanism and spatial attention mechanism are introduced after the residual block, which enables the model to focus on key information related to the task at hand. Additionally, depthwise separable convolution is implemented to replace normal convolution in residual blocks, reducing the network's parameter count and computational costs while maintaining model accuracy. Both paths of the network consist of four analogous submodules, with detailed descriptions provided below.

The model takes the infrared image of the aircraft skin and transforms it into a 3D matrix as input. The contraction path initiates with two  $3 \times 3$  convolutions that extract features from the input image. Subsequently, a batch normalization layer is embedded after the convolution layer to prevent overfitting, followed by a rectified linear unit (ReLU) [40] as the activation function. The feature maps are then inputted into the residual convolution block of the first submodule. In the initial stage, a  $2 \times 2$  max pooling downsampling operation halves the feature size and doubles the number of channels, thereby reducing the influence of rotation and translation while also averting overfitting. Following this, two residual blocks with depthwise separable convolution double the number of channels, leaving the feature map size unchanged. A channel attention mechanism and a spatial attention mechanism are subsequently applied, allowing the extraction of finer and more effective features. This process repeats as the obtained feature maps are fed into the next submodule.

Consistent with the contraction path, the expansion path also contains four submodules with similar structures. Each submodule has a bilinearly interpolated upsampling layer that restores the feature map's resolution. Subsequently, two normal convolution blocks are incorporated after the upsampling process to extract detailed features. Unlike the contraction path, the expansion path includes a skip connection [41], linking low-level feature mapping with high-level feature mapping. As the network goes deeper, the perceptual field broadens, and rich detailed information remains within the high-level feature

mappings. Connecting rich spatial information from the low-level feature mapping in series with the high-level feature mapping mitigates data loss and benefits segmentation by providing multi-scale and multi-level information. Finally, a  $1 \times 1$  convolution layer applies a sigmoid activation function to project multichannel feature maps onto the desired segmentation results.

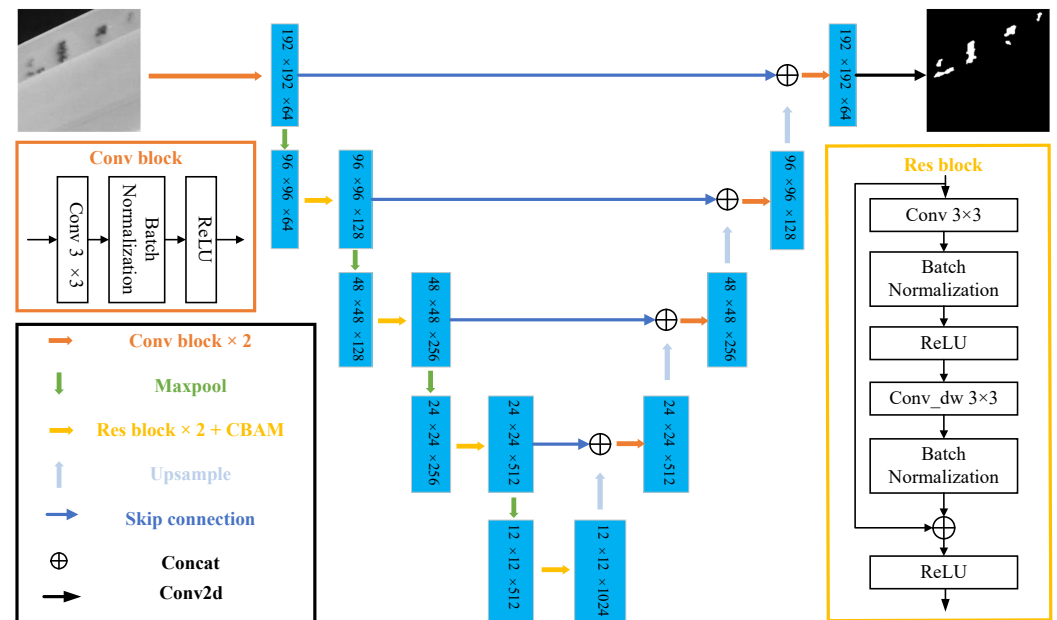


Figure 3. Structure of the proposed network.

### 3.2. Convolutional Block Attention Module

Attention mechanisms have been widely used in various fields, such as computer vision, natural language processing, and others [42,43]. Convolutional neural networks (CNNs) tend to disregard essential information while completing tasks, such as image classification and target detection, due to various feature scales and channel numbers. To address this issue, Woo et al. [44] introduced CBAM, designed to help CNNs focus on relevant yet meaningful features in two dimensions: channel and space.

Figure 4 illustrates the structure of CBAM, comprising two submodules: the channel attention module and the spatial attention module. The convolution layer's output sequentially passes through the channel attention module and the spatial attention module. By element-wise multiplication with the input features, the final output is obtained.

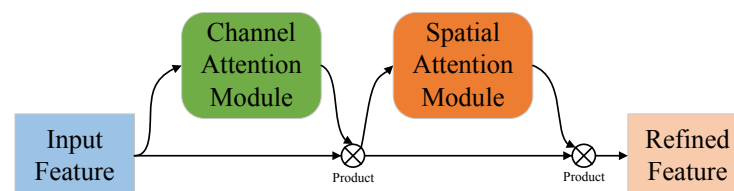
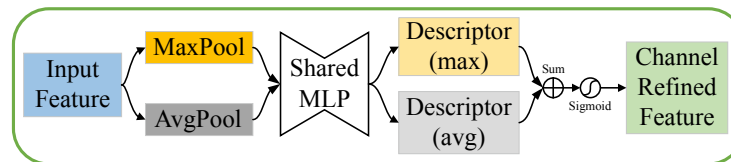


Figure 4. Structure of CBAM.

The channel attention module refines feature maps by assessing the significance of individual channels. Figure 5 demonstrates how spatial information is garnered from the feature map via maximum and average pooling, creating two scalars ( $F_{max}^c \in \mathbb{R}^{C \times 1 \times 1}$  and  $F_{avg}^c \in \mathbb{R}^{C \times 1 \times 1}$ ) that suggest the importance of each channel. These descriptors then enter a multilayer perceptron (MLP) with a hidden layer that determines the relevance of each channel, enhancing useful features while suppressing unimportant ones. Lastly, the two descriptors combine element-wise and are normalized using a sigmoid function to produce a channel attention mapping  $M_c \in \mathbb{R}^{C \times 1 \times 1}$ .

$$\begin{aligned}
 M_c(F) &= \sigma(MLP(MaxPool(F)) \\
 &\quad + MLP(AvgPool(F))) \\
 &= \sigma(W_1(W_0(F_{max}^c)) + W_1(W_0(F_{avg}^c))).
 \end{aligned} \tag{1}$$

where  $\sigma$  denotes the sigmoid function,  $W_0 \in \mathbb{R}^{C/r \times C}$  and  $W_1 \in \mathbb{R}^{C \times C/r}$  are the weights of the *MLP*, and  $r$  is the channel scaling ratio.

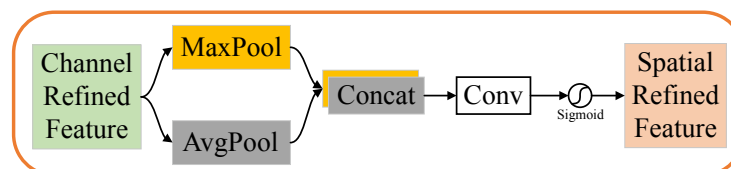


**Figure 5.** Processes in the channel attention module.

In contrast, the spatial attention module determines feature map pixel importance by resizing the maps. Figure 6 shows how the channel information undergoes maximum and average pooling to create two 2D mappings:  $F_{max}^s \in \mathbb{R}^{1 \times H \times W}$  and  $F_{avg}^s \in \mathbb{R}^{1 \times H \times W}$ , respectively. These two feature mappings are then stitched together along the channel direction and fed into a standard convolutional layer for convolutional operation. Lastly, the spatial attention mapping is normalized by the sigmoid function. The specific formula is as follows:

$$\begin{aligned}
 M_s(F) &= \sigma(f^{7 \times 7}([MaxPool(F); AvgPool(F)])) \\
 &= \sigma(f^{7 \times 7}([F_{max}^s; F_{avg}^s])).
 \end{aligned} \tag{2}$$

where  $\sigma$  denotes the sigmoid function, and  $f^{7 \times 7}$  is the convolution operation of  $7 \times 7$ .



**Figure 6.** Processes in the spatial attention module.

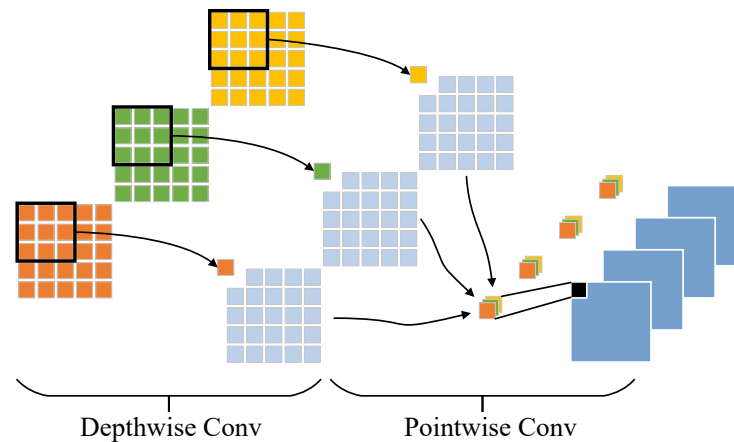
The physical significance of introducing CBAM lies in its ability to mimic human visual attention, which naturally focuses on the most important aspects of a scene. In the context of infrared thermal imaging, the attention mechanism allows the model to highlight key temperature variations associated with water accumulation, which is crucial for precise segmentation. By focusing on areas with significant thermal contrast, CBAM enhances the model's sensitivity to subtle changes in temperature, ultimately improving the accuracy of moisture detection on the aircraft skin. It is proved that incorporating the CBAM attention mechanism into the model increases segmentation accuracy, with the best results achieved when the CBAM module is present in all four residual blocks.

### 3.3. Depthwise Separable Convolution

Depthwise separable convolution is a novel architecture for convolutional neural networks that was initially proposed by François Chollet [45] in 2017. It has been demonstrated that using this method, rather than traditional convolutional operations, can substantially reduce the number of network parameters and computational costs while maintaining model accuracy. This approach has been successful in numerous applications, particularly in resource-constrained mobile device environments, such as MobileNet [46].

Depthwise separable convolution is a two-step architecture, as shown in Figure 7, consisting of depthwise convolution followed by pointwise convolution. In depthwise

convolution, a separate convolution kernel is applied to each channel of the input tensor to generate outputs in distinct channels. This step effectively captures correlations between different channels of the input tensor. In contrast, pointwise convolution applies a  $1 \times 1$  convolution kernel to each pixel of the output tensor after depthwise convolution, combining information from all channels to produce the final output tensor. The operation flow diagram is depicted in Figure 7.

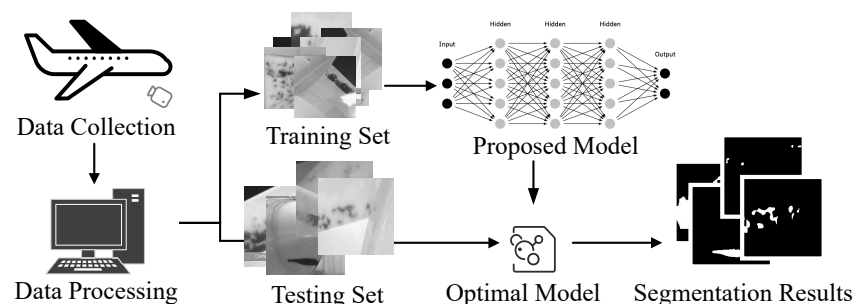


**Figure 7.** Processes in depthwise separable convolution.

Compared to traditional convolutional operations, depthwise separable convolution can significantly reduce the number of parameters and computational operations while maintaining high accuracy. However, it is not recommended to replace normal convolution in a model without careful consideration. Doing so may make it difficult for the model to learn contextual relationships. If the number of parameters is too small, it can reduce the model's ability to fit complex backgrounds and targets, resulting in a sharp decrease in accuracy. In this paper, we discovered that using depthwise separable convolution only in residual blocks can decrease the number of parameters without sacrificing accuracy.

#### 4. Experiment

This paper employs a deep learning approach to segment infrared images of the aircraft skin surface, identifying areas of water accumulation. Figure 8 presents a step-by-step demonstration of our research. We present a detailed description of the experimental process.



**Figure 8.** Diagram of research method.

##### 4.1. Dataset Collection

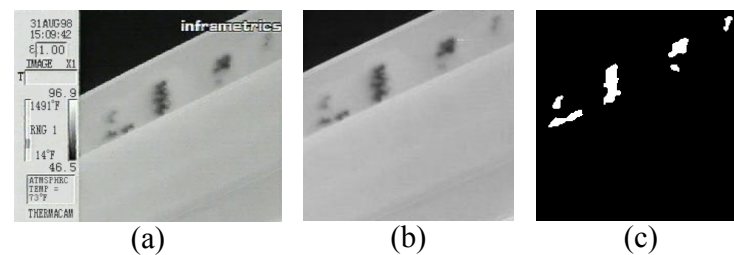
In order to obtain infrared images of the aircraft skin, detection is conducted on an aircraft that has just landed. Multiple sections of the aircraft, including the wings, fuselage, and other areas prone to moisture accumulation, are imaged to ensure the model's applicability across various structural regions. It is worth mentioning that the FLIR E8 infrared thermal camera, manufactured in the United States and shown in Figure 9, is used in the study, and its parameters are given as follows: The infrared resolution is  $320 \times 240$ ,

the thermal sensitivity is 0.06 degrees Celsius, and the field of view is 33 degrees. The camera provides an accuracy of  $\pm 2^\circ\text{C}$  or  $\pm 2\%$  of the reading, with an operating ambient temperature range of  $10^\circ\text{C}$  to  $35^\circ\text{C}$  and a minimum measurable object temperature of  $0^\circ\text{C}$ . For accuracy in temperature measurements, the emissivity was set to 1, and the reflected apparent temperature (RAT) was set to  $22.78^\circ\text{C}$ .



**Figure 9.** The FLIR E8 infrared thermal camera.

The experimental data were derived from an infrared thermal imaging video. The video frames were extracted with an interval of 3 frames, resulting in a total of 1180 infrared images with a resolution of  $320 \times 240$ , as depicted in Figure 10a.



**Figure 10.** Example of the dataset. (a) Original image. (b) Processed image. (c) Ground truth.

To eliminate irrelevant information and focus only on the region of interest, the original images are cropped to a size of  $206 \times 206$ , removing the parameter label on the left side. This is shown in Figure 10b. The 1180 infrared images are randomly split into 944 images for the training set, 118 for the validation set, and 118 for the test set, following an 8:1:1 ratio.

Finally, every image is manually annotated at the pixel level using an annotation tool, as displayed in Figure 10c. The true values were converted to grayscale images with a 0–1 distribution, where 0 represents the background and 1 represents standing water.

In this study, the proposed model segments the infrared images into two classes: water accumulation and non-water areas. This binary classification approach was chosen due to the specific nature of the task at hand. Water accumulation on aircraft surfaces, especially in the context of infrared thermal imaging, is typically the main feature of interest for maintenance purposes. The model's primary objective is to identify and differentiate regions with water accumulation from the rest of the image, which are typically characterized by standard surface temperatures. Therefore, the binary classification of water versus non-water regions simplifies the problem and focuses on the most relevant information for maintenance decision-making.

#### 4.2. Training Strategy

The model presented in this paper is established using the PyTorch deep learning framework. It is an improved U-shaped network that uses ResNet as the backbone. The model is trained and tested on a desktop computer with an Intel(R) Core(TM) i7-12700K 3.60 GHz CPU and an NVIDIA GeForce RTX 3070 GPU.

To speed up the convergence of the model and avoid falling into the local optimum, the optimizer employs the SGD optimization algorithm with momentum. The learning

rate is set to 0.05, the momentum is set to 0.5, and the batch size is set to 8. The model is trained for 50 epochs in case of overfitting. In this paper, binary cross entropy [47] is implemented to calculate the loss between the prediction and the ground truth, which is defined as follows:

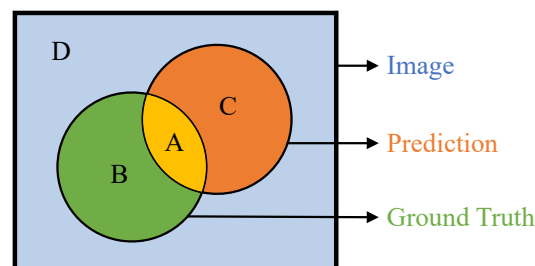
$$Loss = -\frac{1}{N} \sum_{i=1}^M y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i). \quad (3)$$

where  $N$  is the batch size,  $M$  is the number of classes,  $y_i$  is the true value of the class  $i$ , and  $\hat{y}_i$  is the predicted value of the class  $i$ .

## 5. Evaluation Metrics

To comprehensively assess the performance of our segmentation model, commonly used metrics [48] are employed in our experiment, as follows:

Figure 11 illustrates the assessment of the model's prediction. Here, region A consists of the portion of the predicted target that intersects with the true target, while region B represents the portion of the target that the model failed to predict. Region C corresponds to the portion of the target that the model incorrectly predicted, and region D contains the background that was accurately predicted.



**Figure 11.** Diagram of the evaluation index.

The confusion matrix [49] is a fundamental tool for assessing the performance of a classification model. It presents the relationships between true and predicted categories in a matrix form. Specifically, Table 1 displays the confusion matrix for the binary classification problem. The performance of a classification model is evaluated based on the number of true positive (TP), false negative (FN), false positive (FP), and true negative (TN) samples. These categories correspond to regions A, B, C, and D in Figure 11, respectively.

**Table 1.** Confusion matrix.

| Confusion Matrix |          | Predicted |          |
|------------------|----------|-----------|----------|
|                  |          | Positive  | Negative |
| Ground Truth     | Positive | TP        | FN       |
|                  | Negative | FP        | TN       |

1. Mean pixel accuracy (MPA): Calculates the average accuracy of pixel-wise classifications across all pixels in an image. The formula is defined as follows:

$$MPA = \frac{1}{2} \left( \frac{TP}{TP + FN} + \frac{TN}{FP + TN} \right). \quad (4)$$

2. Mean intersection over union (MIoU): Computes the average intersection-over-union values for each class, providing a measure of segmentation accuracy. The formula is defined as follows:

$$MIoU = \frac{1}{2} \left( \frac{TP}{TP + FN + FP} + \frac{TN}{FP + TN + FN} \right). \quad (5)$$

3. Precision: Quantifies the ratio of correctly predicted positive instances to the total predicted positive instances, indicating the accuracy of positive predictions made by the model. The formula is defined as follows:

$$Precision = \frac{TP}{TP + FP}. \quad (6)$$

4. Recall: Measures the ratio of correctly predicted positive instances to the total actual positive instances, reflecting the model's ability to capture all relevant instances in the dataset. The formula is defined as follows:

$$Recall = \frac{TP}{TP + FN}. \quad (7)$$

5. Dice coefficient: Quantifies the similarity between the predicted and ground truth segmentations by measuring the overlap between the two. The formula is defined as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP}. \quad (8)$$

6. Hausdorff Distance (HD): Measures the maximum distance between the predicted and ground truth boundaries, defined as follows:

$$HD(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\}. \quad (9)$$

where  $A$  and  $B$  represent the boundary points of the predicted and ground truth segmentation regions, and  $d(a, b)$  is the Euclidean distance between two boundary points.

7. Cohen's Kappa: Quantifies the overall agreement between the predicted labels and the ground truth, and is calculated as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e}. \quad (10)$$

where  $p_o$  is the observed agreement, and  $p_e$  is the expected agreement by chance, both calculated from the confusion matrix.

8. Floating point operations per second (FLOPs): Measure of the computational complexity of a model, representing the total number of floating-point arithmetic operations.
9. Parameters (Params): The number of model parameters, which measures the total quantity of trainable parameters in a neural network model.
10. Inference time: Refers to the amount of time required to perform inference using a deep learning model, which can directly reflect the efficiency of the model operation.

## 6. Results Analysis

This paper demonstrates the robustness and superiority of the proposed model by comparing it with several state-of-the-art models, including FCN, U-Net, DeepLabv3, PSPNet, and HRNet. Figure 12 displays typical aircraft skin infrared images and their corresponding segmentation results.

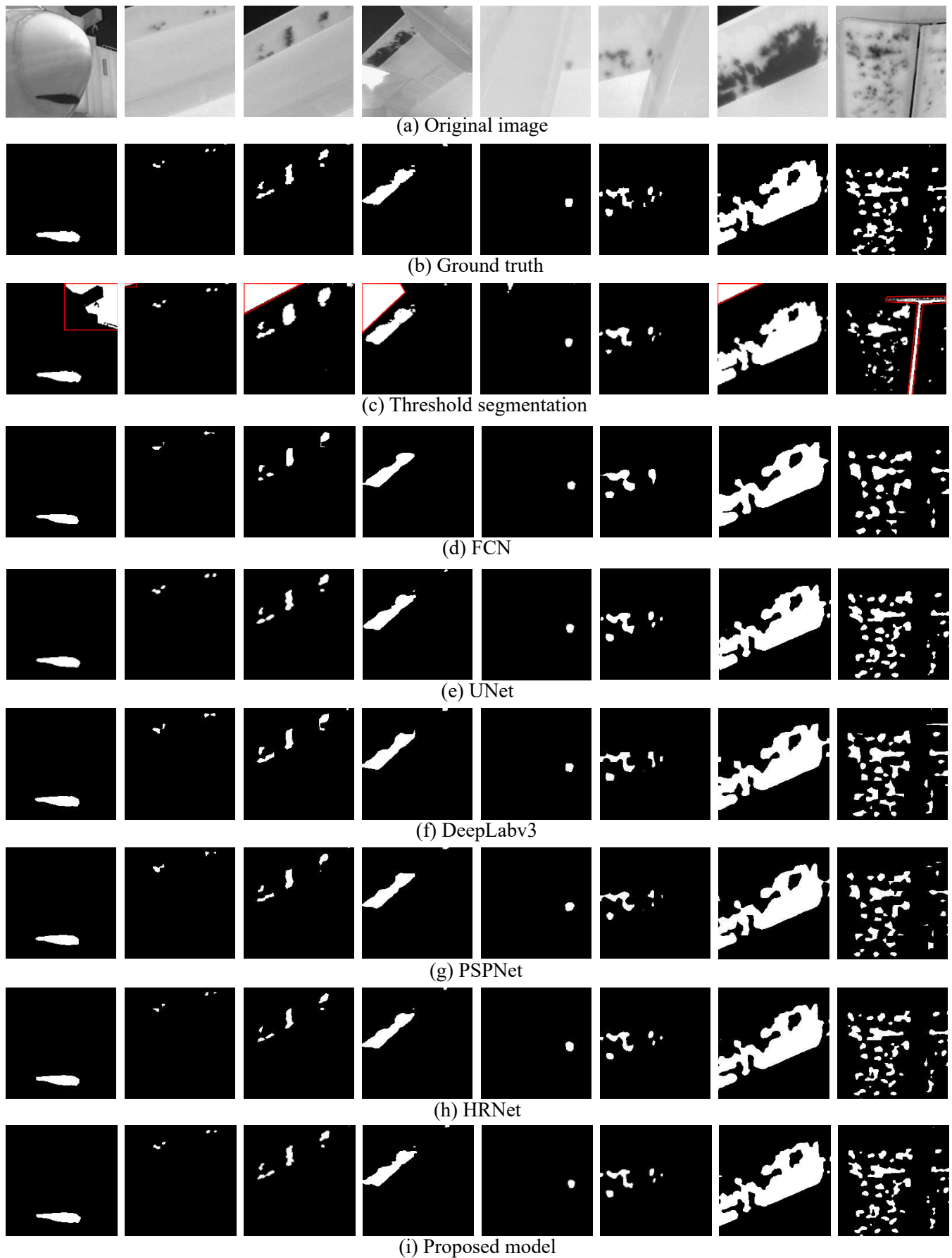


Figure 12. Examples of segmentation results.

Traditional image segmentation methods struggle to accurately identify water accumulation areas in infrared images obtained under actual external field conditions, often misidentifying cold-end parts or the background as the target. Figure 12c depicts such a scenario, where the incorrect identification, highlighted by the red boxes, fails to provide accurate information for subsequent maintenance work. In contrast, deep learning-based image segmentation methods can mitigate this issue effectively, as shown in Figure 12d–i. However, it is evident that FCN and UNet exhibit low segmentation accuracy, particularly when dealing with small targets, resulting in misdetection and omission. Despite DeepLabv3, PSPNet, and HRNet performing well in certain cases, they still have deficiencies in processing edge details. The method proposed in this paper not only achieves precise segmentation of water accumulation areas but also effectively addresses edge details.

When handling small targets, our method deliberately enhances the model's ability to capture target boundaries by incorporating an attention mechanism, thereby improving the precise processing of details and significantly reducing false positives and missed detections. Moreover, we tackle the shortcomings of PSPNet and HRNet in handling edge details. Additionally, we have optimized the training process by implementing a more efficient depth-separable convolutional design, which accelerates model convergence and enhances generalizability.

After conducting extensive experiments, our approach surpasses the performance of established methods such as FCN, UNet, DeepLabv3, PSPNet, and HRNet. This superiority is evident not only in terms of overall performance but also in the attainment of satisfactory results for small targets and edge details.

Subsequently, the performance of the proposed model is quantitatively evaluated based on two criteria: segmentation accuracy and inference speed.

#### 6.1. Segmentation Accuracy

We first evaluate the segmentation performance of our proposed model using MPA and MIoU, which are widely recognized metrics in segmentation tasks. However, both have certain limitations. Specifically, MPA does not account for the relative importance of each class, which may be problematic in imbalanced datasets where certain classes, such as water accumulation, are underrepresented. Similarly, MIoU may be influenced by class imbalance, as it averages the intersection over all classes, which could lead to a biased evaluation when small classes dominate the results.

To address these limitations and provide a more comprehensive evaluation of the model's performance, we also include additional metrics, such as precision, recall, the Dice coefficient, the Hausdorff distance, and Cohen's Kappa. These metrics offer a more balanced and robust evaluation, especially in the context of imbalanced datasets where small classes are crucial for accurate performance assessment.

Table 2 presents the experimental results of the test set, where bold numbers indicate the best performance in each category. It illustrates that the proposed method achieves outstanding performance with a 92.68% MIoU and a 97.58% MPA. It can be observed that the two lightweight models, FCN and UNet, have achieved good segmentation results, with MIoU values of 84.38% and 89.42%, and MPA values of 91.52% and 97.56%, respectively.

**Table 2.** Segmentation accuracy.

| Model          | MPA           | MIoU          | Precision     | Recall        | Dice          | HD          | $\kappa$    |
|----------------|---------------|---------------|---------------|---------------|---------------|-------------|-------------|
| FCN            | 91.52%        | 84.38%        | 99.19%        | 98.57%        | 98.88%        | 23.44       | 0.81        |
| U-Net          | 97.56%        | 89.42%        | <b>99.86%</b> | 98.64%        | 99.24%        | 17.99       | 0.91        |
| DeepLabv3      | 96.84%        | 85.43%        | 99.75%        | 97.90%        | 98.81%        | <b>9.66</b> | 0.87        |
| PSPNet         | 92.13%        | 86.08%        | 99.30%        | 98.88%        | 99.09%        | 13.52       | 0.83        |
| HRNet          | 96.79%        | 89.90%        | 99.78%        | 98.93%        | 99.35%        | 14.72       | 0.89        |
| Proposed model | <b>97.97%</b> | <b>92.68%</b> | 99.81%        | <b>99.27%</b> | <b>99.54%</b> | 10.81       | <b>0.93</b> |

The bold indicates the best results in the table.

The more recent segmentation models, namely DeepLabv3, PSPNet, and HRNet, did not perform as well, with MIoU values of 85.43%, 86.08%, and 89.90%, and MPA values of 96.84%, 92.13%, and 96.79%, respectively, even falling short of UNet. This indicates that larger and deeper networks do not necessarily lead to better segmentation results.

Our model outperforms others in terms of MPA, MIoU, recall, Dice, and Cohen's Kappa. The proposed method achieves nearly 3% higher MIoU than the best-performing model, showcasing its superiority in segmentation accuracy.

### 6.2. Inference Speed

Apart from segmentation accuracy, a model's inference speed is also a critical indicator of its quality. Referring to Table 3 (bold indicates the best performance), we can observe that the FCN and UNet models are relatively small, with computational volumes of 11.37 G and 17.51 G, and parameter counts of 15.25 M and 13.40 M, respectively. FCN processes an image in 3.46 ms, while UNet takes 5.88 ms to do the same.

**Table 3.** Inference speed.

| Model          | FLOPs          | Params         | Inference Time |
|----------------|----------------|----------------|----------------|
| FCN            | <b>11.37 G</b> | 15.25 M        | <b>3.46 ms</b> |
| U-Net          | 17.51 G        | <b>13.40 M</b> | 5.88 ms        |
| DeepLabv3      | 23.08 G        | 39.63 M        | 7.62 ms        |
| PSPNet         | 28.42 G        | 52.49 M        | 10.27 ms       |
| HRNet          | 13.20 G        | 65.85 M        | 42.98 ms       |
| Proposed model | 33.06 G        | 35.07 M        | 15.02 ms       |

The bold indicates the best results in the table.

DeepLabv3 doubles the computational volume and parameter count, reaching 23.08 G of computation and 39.63 M parameters. Meanwhile, PSPNet and HRNet quadruple the number of parameters, resulting in 52.49 M and 65.85 M, respectively, and they take longer to process an image.

The model proposed in this paper reduces the number of parameters by half and decreases the image processing time by two-thirds compared to the best-performing model HRNet. This demonstrates the efficiency of our approach, achieving a balance between segmentation accuracy and faster inference speed.

### 6.3. Ablation Study

To demonstrate the effectiveness of the convolutional block attention module (CBAM) and depthwise separable convolution (DW) used in this research, we conducted a series of ablation experiments. The experiments were carried out using the aforementioned dataset, and we utilized Res-UNet as the base architecture. The dataset consisted of 944 images for training, 118 images for validation, and 118 images for testing.

Following the same experimental scheme as mentioned earlier, we trained and tested four models: Res-UNet without any additional modules, Res-UNet with CBAM, Res-UNet with dw, and the model proposed in this paper, which combines both CBAM and dw. The experimental results are summarized in Table 4, where bold values indicate the best performance across all methods.

**Table 4.** Ablation study results.

| Model               | CBAM | DW | MIoU          | MPA           | Precision     | Recall        | HD          | $\kappa$    | FLOPs          | Params         | Inference Time |
|---------------------|------|----|---------------|---------------|---------------|---------------|-------------|-------------|----------------|----------------|----------------|
| Res-UNet (Baseline) |      |    | 91.27%        | 97.65%        | 96.81%        | 99.82%        | <b>9.05</b> | 0.83        | 42.69 G        | 57.16 M        | 15.68 ms       |
| Res-UNet with CBAM  | ✓    |    | 92.19%        | 97.90%        | 96.53%        | <b>99.92%</b> | 10.95       | <b>0.93</b> | 42.70 G        | 57.34 M        | 18.16 ms       |
| Res-UNet with dw    |      | ✓  | 91.43%        | 97.20%        | 96.61%        | <b>99.92%</b> | 10.81       | 0.92        | 33.11 G        | <b>34.93 M</b> | <b>9.54 ms</b> |
| Proposed model      | ✓    | ✓  | <b>92.68%</b> | <b>97.97%</b> | <b>99.81%</b> | 99.27%        | 10.81       | <b>0.93</b> | <b>33.06 G</b> | 35.11 M        | 15.02 ms       |

✓ represents the use of the module and the bold indicates the best results in the table.

As shown in Table 4, the benchmark model Res-UNet achieves an MIoU of 91.27%, MPA of 97.65%, FLOPs of 42.69 G, Params of 57.16 M, and takes 15.68 ms to process an image.

When CBAM is added, the MIoU improves by almost 1% compared to Res-UNet, while the computational and parameter counts do not change significantly.

Replacing the normal convolutional layer with depthwise separable convolution does not cause a drastic change in the model's accuracy, but it results in a 22.44% reduction in the computational workload and a 38.89% reduction in the number of parameters compared to Res-UNet.

Finally, it is evident that the model proposed in this paper, which utilizes depthwise separable convolution along with CBAM, not only improves the segmentation accuracy by 1.41% compared to the benchmark model but also reduces the computation by 22.41% and the number of parameters by 38.58%. This demonstrates that CBAM indeed enhances the segmentation accuracy of the model, and the usage of depthwise separable convolution effectively reduces the model's complexity, validating the soundness of the model design in this research.

## 7. Discussion

Our experimental results demonstrated that the proposed method achieved remarkable performance in terms of segmentation accuracy compared with established methods. It highlights the effectiveness of our approach in addressing the segmentation of infrared images, particularly in the context of detecting water accumulation on aircraft surfaces, which will provide significant assistance for aircraft maintenance decision-making. The incorporation of attention mechanisms allows the model to better perceive and highlight crucial regions in the infrared images. It is particularly advantageous for detecting areas of water accumulation on aircraft surfaces, where specific patterns and details are essential. Moreover, they empower the model with increased adaptability, enabling it to dynamically shift focus in response to distinct characteristics in various infrared images. This adaptability significantly enhances the model's generalization and robustness across diverse scenarios.

Notably, it is important to acknowledge the limitations of our approach. As infrared thermal imaging is primarily effective for detecting surface-level temperature variations, moisture trapped within deeper layers may not exhibit significant contrast in the thermal images, limiting the model's ability to detect subsurface moisture accurately.

In practical terms, our proposed model has the potential to significantly enhance aircraft maintenance processes by enabling the detection of water accumulation on aircraft surfaces, which can be critical for identifying potential safety risks. However, integrating this technology into existing maintenance workflows presents several challenges. One such challenge is ensuring that infrared thermal imaging systems are deployed effectively across a fleet of aircraft with varying operational conditions, which may require standardizing data acquisition procedures and maintaining consistent imaging quality. Furthermore, the model would need to be adapted to work across different aircraft models and environmental conditions, which could involve collecting and processing large amounts of diverse data.

Another key challenge is the model's interpretability. Although attention mechanisms improve segmentation accuracy by focusing on relevant regions of the image, they also add complexity to the model, making it more difficult to explain the rationale behind its decisions. In safety-critical applications like aircraft maintenance, understanding why the model makes a particular decision is crucial for gaining trust and ensuring correct actions are taken based on the model's outputs.

## 8. Conclusions

In this study, a novel infrared image segmentation network is designed for the detection of water accumulation on aircraft surfaces. By integrating attention mechanisms and depthwise separable convolution, our model exhibits superior accuracy and efficiency com-

pared with the existing networks. Specifically, our approach achieved an MPA of 97.97%, an MIoU of 92.68%, a recall of 99.27%, a Dice coefficient of 99.54%, and a Cohen's Kappa of 0.93, demonstrating significant improvement over traditional methods. The positive outcomes indicate the viability of our approach in aircraft structural damage detection while recognizing the need for ongoing refinement and exploration.

Looking forward, our future research endeavors will focus on exploring ways to mitigate computational costs, optimizing hyperparameters, and investigating methods to enhance the model's interpretability. To address the challenge of detecting moisture within deeper structures, we plan to explore multi-modal sensing approaches, such as integrating infrared with ultrasound or terahertz imaging, to improve sensitivity to subsurface moisture. Additionally, we will collect more extensive datasets and conduct experiments across various aircraft models to further validate the robustness and generalization capabilities of our proposed approach. We believe this work lays a solid foundation for future advancements in aviation safety, emphasizing the continuous pursuit of innovation and improvement.

**Author Contributions:** Conceptualization, H.F. and H.Z.; methodology, H.F. and X.L.; software, H.F.; validation, H.F., Y.L. and Z.L.; formal analysis, H.W.; funding acquisition, H.Z.; investigation, H.F.; resources, H.Z.; data curation, H.Z.; writing—original draft preparation, H.F.; writing—review and editing, H.F., Y.L. and Z.L.; visualization, H.W.; supervision, X.L.; project administration, H.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (U2133202).

**Data Availability Statement:** The original contributions presented in this study are included in the article, and further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Meng, D.; Boer, W.; Juan, X.; Kasule, A.N.; Hongfu, Z. Visual inspection of aircraft skin: Automated pixel-level defect detection by instance segmentation. *Chin. J. Aeronaut.* **2022**, *35*, 254–264.
2. Alemour, B.; Badran, O.; Hassan, M.R. A review of using conductive composite materials in solving lightening strike and ice accumulation problems in aviation. *J. Aerosp. Technol. Manag.* **2019**, *11*, e1919. [[CrossRef](#)]
3. Wei, K.; Yang, Y.; Zuo, H.; Zhong, D. A review on ice detection technology and ice elimination technology for wind turbine. *Wind Energy* **2020**, *23*, 433–457. [[CrossRef](#)]
4. Chulkov, A.; Vavilov, V.; Moskovchenko, A.; Pan, Y.Y. Quantitative evaluation of water content in composite honeycomb structures by using one-sided IR thermography: Is there any promise? In Proceedings of the Thermosense: Thermal Infrared Applications XXXIX, Anaheim, CA, USA, 9–13 April 2017; SPIE: St Bellingham, WA, USA, 2017; Volume 10214, pp. 213–220.
5. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [[CrossRef](#)] [[PubMed](#)]
6. Zaidi, S.S.A.; Ansari, M.S.; Aslam, A.; Kanwal, N.; Asghar, M.; Lee, B. A survey of modern deep learning based object detection models. *Digit. Signal Process.* **2022**, *126*, 103514. [[CrossRef](#)]
7. Roy, A.M.; Bhaduri, J. DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism. *Adv. Eng. Inform.* **2023**, *56*, 102007. [[CrossRef](#)]
8. Wang, R.; Lei, T.; Cui, R.; Zhang, B.; Meng, H.; Nandi, A.K. Medical image segmentation using deep learning: A survey. *IET Image Process.* **2022**, *16*, 1243–1267. [[CrossRef](#)]
9. Lin, A.; Chen, B.; Xu, J.; Zhang, Z.; Lu, G.; Zhang, D. Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–15. [[CrossRef](#)]
10. Aboelenein, N.M.; Piao, S.; Noor, A.; Ahmed, P.N. MIRAUNet: An improved neural network based on U-Net for gliomas segmentation. *Signal Process. Image Commun.* **2022**, *101*, 116553. [[CrossRef](#)]
11. Abadi, A.D.; Gu, Y.; Goncharenko, I.; Kamijo, S. Detection of Cyclist's Crossing Intention based on Posture Estimation for Autonomous Driving. *IEEE Sens. J.* **2023**, *23*, 11274–11284. [[CrossRef](#)]
12. Nahavandi, S.; Alizadehsani, R.; Nahavandi, D.; Mohamed, S.; Mohajer, N.; Rokouzzaman, M.; Hossain, I. A Comprehensive Review on Autonomous Navigation. *arXiv* **2022**, arXiv:2212.12808.
13. Liu, H. Image Segmentation Techniques for Intelligent Monitoring of Putonghua Examinations. *Adv. Math. Phys.* **2022**, *2022*, 4302666. [[CrossRef](#)]

14. Kazakeviciute-Januskeviciene, G.; Janusonis, E.; Bausys, R. Evaluation of the segmentation of remote sensing images. In Proceedings of the 2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania, 22 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–7.
15. Liang, C.; Zhang, Y. Multi Object Infrared Image Segmentation Based on Multi-level Feature Fusion. In Proceedings of the International Conference on Computer Science, Engineering and Education Applications, Warsaw, Poland, 17–19 March 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 26–36.
16. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
17. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Nock, R.; Nielsen, F. Statistical region merging. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 1452–1458. [\[CrossRef\]](#)
19. Najman, L.; Schmitt, M. Watershed of a continuous function. *Signal Process.* **1994**, *38*, 99–112. [\[CrossRef\]](#)
20. Neves, S.; Da Silva, E.; Mendonca, G. Wavelet-watershed automatic infrared image segmentation method. *Electron. Lett.* **2003**, *39*, 903–904. [\[CrossRef\]](#)
21. Zhou, S.; Yang, P.; Xie, W. Infrared image segmentation based on Otsu and genetic algorithm. In Proceedings of the 2011 International Conference on Multimedia Technology, Hangzhou, China, 26–28 July 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 5421–5424.
22. Cui, L.L.; Lai, H.; Tang, Y.W.; Qi, M.J. Application of petrochemical heat equipments' state inspection based on the edge detection of infrared image segmentation. *Appl. Mech. Mater.* **2013**, *303*, 970–974. [\[CrossRef\]](#)
23. Zhang, R.; Zhu, S.; Zhou, Q. A novel gradient vector flow snake model based on convex function for infrared image segmentation. *Sensors* **2016**, *16*, 1756. [\[CrossRef\]](#)
24. Liu, S.T.; Wang, Z.; Wang, Z. Fast infrared image segmentation method based on 2D OTSU and particle swarm optimization. In Proceedings of the Third International Workshop on Pattern Recognition, Jinan, China, 19–21 May 2018; SPIE: St Bellingham, WA, USA, 2018; Volume 10828, pp. 94–100.
25. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A survey of deep learning and its applications: A new paradigm to machine learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [\[CrossRef\]](#)
26. Rafique, A.A.; Ghadi, Y.Y.; Alsuhbany, S.A.; Chelloug, S.A.; Jalal, A.; Park, J. CNN Based Multi-Object Segmentation and Feature Fusion for Scene Recognition. *Comput. Mater. Contin.* **2022**, *73*, 4657–4675.
27. 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.
28. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Proceedings of the 18th International Conference, Munich, Germany, 5–9 October 2015, Proceedings, Part III 18*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
29. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* **2016**, arXiv:1606.02147.
30. 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. [\[CrossRef\]](#)
31. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
32. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
33. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [\[CrossRef\]](#)
34. Zhang, H.; Hong, X.; Zhou, S.; Wang, Q. Infrared image segmentation for photovoltaic panels based on Res-UNet. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Xi'an, China, 8–11 November 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 611–622.
35. Nan, J.; Bo, L. A lightweight real-time infrared image instance segmentation algorithm based on tiny YOLO. In Proceedings of the Seventh Symposium on Novel Photoelectronic Detection Technology and Applications, Kunming, China, 5–7 November 2020; SPIE: St Bellingham, WA, USA, 2021; Volume 11763, pp. 1098–1107.
36. Chen, S.; Xu, X.; Yang, N.; Chen, X.; Du, F.; Ding, S.; Gao, W. R-Net: A novel fully convolutional network-based infrared image segmentation method for intelligent human behavior analysis. *Infrared Phys. Technol.* **2022**, *123*, 104164. [\[CrossRef\]](#)
37. Yu, J.; He, Y.; Liu, H.; Zhang, F.; Li, J.; Sun, G.; Zhang, X.; Yang, R.; Wang, P.; Wang, H. An improved U-Net model for infrared image segmentation of wind turbine blade. *IEEE Sens. J.* **2022**, *23*, 1318–1327. [\[CrossRef\]](#)
38. Zhao, X.; Dai, J.; Li, L.p.; Su, Z.; Ma, X.f. Deep learning-based laser and infrared composite imaging for armor target identification and segmentation in complex battlefield environments. *Infrared Phys. Technol.* **2023**, *132*, 104725. [\[CrossRef\]](#)
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
40. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.

41. Liu, F.; Ren, X.; Zhang, Z.; Sun, X.; Zou, Y. Rethinking skip connection with layer normalization. In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain, 8–13 December 2020; pp. 3586–3598.
42. Choromanski, K.; Likhoshesterov, V.; Dohan, D.; Song, X.; Kane, A.; Sarlos, T.; Hawkins, P.; Davis, J.; Mohiuddin, A.; Kaiser, L.; et al. Rethinking attention with performers. *arXiv* **2020**, arXiv:2009.14794.
43. Guo, M.H.; Xu, T.X.; Liu, J.J.; Liu, Z.N.; Jiang, P.T.; Mu, T.J.; Zhang, S.H.; Martin, R.R.; Cheng, M.M.; Hu, S.M. Attention mechanisms in computer vision: A survey. *Comput. Vis. Media* **2022**, *8*, 331–368. [[CrossRef](#)]
44. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
45. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
46. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
47. Yi-de, M.; Qing, L.; Zhi-Bai, Q. Automated image segmentation using improved PCNN model based on cross-entropy. In Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004; IEEE: Piscataway, NJ, USA, 2004; pp. 743–746.
48. Wang, Z.; Wang, E.; Zhu, Y. Image segmentation evaluation: A survey of methods. *Artif. Intell. Rev.* **2020**, *53*, 5637–5674. [[CrossRef](#)]
49. Powers, D.M. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv* **2020**, arXiv:2010.16061.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.