



Hongbin Cao¹, Yuxi Gao¹, Weiwei Cai², Zhuonong Xu^{1,*} and Liujun Li³

- ¹ College of Computer & Information Engineering, Central South University of Forestry and Technology, Changsha 410004, China
- ² School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China
- ³ Department of Soil and Water Systems, University of Idaho, Moscow, ID 83844, USA
- * Correspondence: t19980580@csuft.edu.cn

Abstract: Road cracks are one of the external manifestations of safety hazards in transportation. At present, the detection and segmentation of road cracks is still an intensively researched issue. With the development of image segmentation technology of the convolutional neural network, the identification of road cracks has also ushered in new opportunities. However, the traditional road crack segmentation method has these three problems: 1. It is susceptible to the influence of complex background noise information. 2. Road cracks usually appear in irregular shapes, which increases the difficulty of model segmentation. 3. The cracks appear discontinuous in the segmentation results. Aiming at these problems, a network segmentation model of HC-Unet++ road crack detection is proposed in this paper. In this network model, a deep parallel feature fusion module is first proposed, one which can effectively detect various irregular shape cracks. Secondly, the SEnet attention mechanism is used to eliminate complex backgrounds to correctly extract crack information. Finally, the Blurpool pooling operation is used to replace the original maximum pooling in order to solve the crack discontinuity of the segmentation results. Through the comparison with some advanced network models, it is found that the HC-Unet++ network model is more precise for the segmentation of road cracks. The experimental results show that the method proposed in this paper has achieved 76.32% mIOU, 82.39% mPA, 85.51% mPrecision, 70.26% dice and Hd95 of 5.05 on the self-made 1040 road crack dataset. Compared with the advanced network model, the HC-Unet++ network model has stronger generalization ability and higher segmentation accuracy, which is more suitable for the segmentation detection of road cracks. Therefore, the HC-Unet++ network model proposed in this paper plays an important role in road maintenance and traffic safety.

Keywords: road cracks; drone; feature fusion; SEnet; Blurpool; HC-Unet++

1. Introduction

As an important part of the transportation hub, roads provide the most basic facilities for the entire transportation network. Road cracks are a common phenomenon of road surface damage. After the road is completed, due to natural damage factors such as longterm exposure to the sun [1], rain erosion, and human factors such as frequent rolling of vehicles, construction materials, and construction quality, the road appears to incur varying degrees of crack damage. With the development of cracks, the overall structure of the road will gradually change [2], which will affect the service life and safety of the road to a certain extent, and even cause traffic accidents in severe cases. In order to reduce the cost [3] and prolong the service life of the road, it is necessary to survey and repair the road as soon as possible. Therefore, finding an accurate and effective method for identifying road cracks and repairing cracks in time is of great significance to the structural safety of roads and traffic safety.



Citation: Cao, H.; Gao, Y.; Cai, W.; Xu, Z.; Li, L. Segmentation Detection Method for Complex Road Cracks Collected by UAV Based on HC-Unet++. *Drones* **2023**, *7*, 189. https://doi.org/10.3390/ drones7030189

Academic Editor: Seokwon Yeom

Received: 16 February 2023 Revised: 5 March 2023 Accepted: 7 March 2023 Published: 10 March 2023



Copyright: © 2023 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/).

In the early days, many scholars proposed the use of traditional image processing methods to solve the task of crack identification, such as the use of typical digital image processing algorithms to extract crack features to identify cracks [4] and using multi-threshold image segmentation methods to reduce computational cost and improve segmentation accuracy [5]. For example, Jin et al. [6] proposed a detection and segmentation algorithm for mixed cracks, using the threshold method of histogram to obtain the approximate location of cracks. Li et al. [7] used the improved Otsu threshold segmentation algorithm and adaptive iterative threshold segmentation method to identify airport runway cracks containing runway markings. Li et al. [8] proposed a new unsupervised multi-scale fusion crack detection algorithm that extracts candidate cracks in the image through the minimum intensity path. Additionally, a multivariate statistical hypothesis-testing crack assessment model was developed to detect pavement cracks. Xu [9] and others used the second-order differential operator edge detection algorithm to identify cracks in infrared images at different times. Zhao et al. [10] combined the improved Canny edge detection algorithm with the edge filter of road surface edge detection to effectively detect edge cracks in road surface images while eliminating noise interference. Liang et al. [11] used wavelet technology to detect crack edges and eliminated background noise interference by searching and analyzing the maximum value of wavelet coefficients. At the same time, a threshold method was used to judge whether there were cracks on the pavement. Peggy et al. [12] performed a separable 2D continuous wavelet transform for several scales based on the continuous wavelet transform. The propagation between scales was analyzed by searching for the maximum value of the wavelet coefficients in order to determine whether there were cracks. Cheng et al. [13] built an end-to-end diagnostic mechanism based on continuous wavelet transform, one which adaptively captured features through automatic feature extraction. Although traditional image processing methods can identify and extract cracks to a certain extent, this method has poor generalization performance and can only be applied to specific contexts. Due to the complexity of road cracks, in the course of the crack identification process, whether or not the feature extraction is feasible, there is no guarantee that a crack will be detected.

With the development of deep learning, many scholars have proposed using convolutional neural network image processing methods to identify cracks. In addition, road crack images collected by UAV [14] can be more challenging due to lighting conditions, viewpoints, and scales. Researchers used a Gaussian noise residual network to extract crack features [15] and tried to use MVMNet's multiple detection method to identify cracks [16]. Zhu et al. [17] proposed a feature fusion enhancement module coupled with a convolutional network attention mechanism. It was used to improve the interaction between feature maps and strengthened the dependence between feature channels, so as to achieve the identification of cracks. Cha et al. [18] utilized sliding windows to divide the image into blocks and used a CNN convolutional network to predict whether cracks existed in the block. Fan et al. [19] proposed a new threshold method based on a CNN convolutional network to extract cracks in classified color images. Sadrawi et al. [20] proposed to use the lenet network model to identify cracks, and the model finally classified the cracks in the image into horizontal, vertical and massive cracks. Huang et al. [21] used FCN to extract feature hierarchy to detect cracks in subway shield tunnels. Zou et al. [22] utilized an endto-end trainable deep neural network to fuse the multi-scale deep convolutional features learned in the layered convolution stage to capture the line structure to better detect cracks. However, these convolutional neural networks cannot solve the following three problems well in the segmentation of cracks: 1. As shown in Table 1a, there are many complex background interferences such as zebra crossings of different colors, stains, manhole covers, etc. During the segmentation process, this interference information will affect the extraction and segmentation of fracture features by the network. 2. As shown in Table 1b, road cracks usually appear in irregular shapes, and the conventional convolutional neural network cannot fully capture the crack feature information in the feature extraction stage. 3. As shown in Table 1c, in the process of fracture segmentation, the down-sampling of each

Original
Image: Constraint of the second seco

layer will cause information loss, resulting in the phenomenon of fracture discontinuity in

 Table 1. Traditional network segmentation results.

the segmentation result.

In order to be able to completely identify irregularly shaped cracks, Yuan et al. [23] designed a RDA detail-attention module for crack detection, one which enhanced the segmentation of irregular cracks by accurately locating the spatial location of cracks. Billah et al. [24] proposed an encoder-based deep network architecture to strengthen the extraction of clear features of cracks to effectively find the exact location of cracks and identify cracks with irregular shapes. Li et al. [25] designed a multi-scale feature fusion method. A multi-scale parallel structure was obtained through various sampling rates and pooling methods. This structure can obtain more receptive fields to improve the ability of the network to identify disordered cracks. Considering the particularity of road crack segmentation, a deep parallel feature fusion module is proposed in this paper. This module is located in the deep layer of the network. The purpose is to obtain more receptive fields so that the network can effectively locate cracks and judge their shapes, so as to enhance the ability of network segmentation to identify irregular cracks.

In order to enhance the extraction of important information and suppress the interference of useless information, many scholars have tried different methods. Yang et al. [26] proposed the AFB attention fusion block, which was used to replace the original skip connection to enhance feature extraction. Zhang et al. [27] combined MobileNets with the convolutional attention module of CBAM. Firstly, the residual structure of MobileNetV2 was introduced to eliminate the accuracy drop caused by separable convolution in the depth direction, and then CBAM was embedded into the convolutional layer to enhance the effect of important information. Yang et al. [28] proposed a UAV-supported edge computing method that was able to integrate different levels of feature map information into low-level features. This allowed the network to remove the complexity of the background and the inhomogeneity of the crack intensity. Qiao et al. [29] proposed the scSE attention mechanism. This module was divided into upper and lower branches. Each branch obtained a different matrix and multiplied the input image and finally stitched it to achieve the recalibration of the feature map. Hu et al. [30] proposed the SEnet attention mechanism. Through the compressed weight matrix, different weights are assigned to different positions in the channel to help the network obtain important information. Based on the complexity of the background when detecting road cracks, this study uses the SEnet attention mechanism and uses it in the decoding and encoding stages deep in the network structure. The feature map is corrected by the obtained one-dimensional weight vector to achieve the purpose of eliminating the interference of non-fracture information and enhancing the feature extraction of fracture information during feature extraction.

In order to solve the fracture discontinuity phenomenon that occurs during the fracture segmentation process, Xiang et al. [31] adopted a pyramid module to divide the feature map into different sub-regions. Crack information was extracted from the global view by aggregating contextual information in different regions to enhance the continuity of pavement crack detection. Han et al. [32] proposed a jump-level round-trip sampling structure to solve the problem of interruption of continuous cracks in the segmentation process by improving the ability of different receptive fields to perceive information at different scales. Jiang et al. [33] proposed a segmentation framework with an enhanced graph network branch to improve crack segmentation. In this paper, due to the loss of part of the information caused by the maximum pooling operation in the Unet++ network downsampling, the segmentation cracks are not continuous in the segmentation results. Therefore, the Blurpool pooling operation is introduced, which can alleviate the shift-equivariance to the greatest extent, greatly reduce the loss of some crack features when the feature map is down-sampled, and solve the crack discontinuity in the segmentation.

Based on the problems of complex background, irregular crack shape and discontinuous crack segmentation in the process of road crack segmentation. We propose a road crack network segmentation model based on HC-Unet++. The contributions of this paper are as follows:

1. A UAV-based road crack dataset is constructed, one which contains 1040 images of road cracks with complex backgrounds and irregular crack shapes. These images are precisely annotated and used to train a network model to solve the problem of road crack segmentation.

2. We propose a deep parallel feature fusion module. This module operates on the network's deepest layer and can obtain a larger receptive field, making global feature extraction possible. This module is partitioned into two parallel branches; each branch extracts crack features using Conv, BN, and Relu operations and stitches the extracted feature maps. The spliced feature map has more comprehensive irregular crack features, thereby enhancing the network's capacity to segment irregular cracks.

3. The SEnet attention mechanism is introduced and used in the deep encoding and decoding stages of the network. First, the input feature map is compressed into a one-dimensional vector through the spatial dimension. Then after the 1*1 convolution, channel feature learning is performed to weight the one-dimensional vector. Finally, the one-dimensional weight of channel attention information is learned, which is corrected in combination with the input character map. The redundant irrelevant information contained in the output feature map is reduced, and the important crack feature information is increased.

4. Refers to the Blurpool pooling procedure. This pooling operation can eliminate aliasing to the greatest extent, return the original incorrect output to its correct location, and achieve translation invariance to a significant degree. The Blurpool pooling operation replaces the maximum pooling in the original network, making the output robust to small input translations and minimizing the loss of crack features during downsampling. The problem of crack discontinuity in the segmentation result has been resolved.

5. Experimental results show that compared to other advanced methods, the network model proposed in this article on the home-made road cracks dataset has stronger generalization capabilities and higher accuracy. In generalization experiments, its multiple indicators have also achieved excellent results. Therefore, the HC-Unet++ network has been more versatile in the segmentation of road cracks, making it more efficient and cost-effective.

2. Materials and Methods

2.1. Data Acquisition

The experimental dataset of this experiment was taken by the team researchers on Shaoshan South Road and Furong South Road in Changsha City with a DJI Mini3Pro drone (as shown in Figure 1), with a resolution of 5472×3468 . When dataset is collected, time synchronization and control of the drone is very important [34,35]. During the shooting process, we hovered the UAV 3 to 5 m above the ground and set the viewing angle to 87 degrees in order to capture close-up images of cracks. This type of crack image includes more individual crack targets. When photographing distant cracks, we set the flight altitude of the drone between 8 and 10 m. Currently, the UAV has a two-lane field of vision, and the obtained image depicts a crack scene that is relatively complete. From the acquired photos, a total of 1040 images with complex backgrounds and irregular crack shapes were selected as the route crack dataset for research. Among them, 734 pictures contained complex backgrounds such as zebra crossings, manhole covers, stains, scratches, etc., and 813 pictures had different crack shapes. Then the resolution of these images was adjusted to 512 imes 512 as input images, and these images were manually labeled using the Labelme tool. Afterwards, the background of the marked picture was black, the crack was white, and it was stored in the form of json. In training, we divided the dataset into training set, validation set and test set in a ratio of 8:1:1.



Figure 1. Collection of crack images.

2.2. *Methods*

2.2.1. HC-Unet++

In order to narrow the semantic gap between the feature maps from the encoder and decoder networks in the Unet network model, Zhou et al. [36] proposed a network segmentation model for Unet++. In this network model structure, the hollow Unet network was filled. It has connections at every point in the horizontal direction, so that features at different levels can be captured. Due to the different sensitivities of the receptive field at different depths, the shallow receptive field is more sensitive to small targets, while the deep receptive field is more sensitive to large targets. Additionally, splicing together through the feature concat can integrate the advantages of the two. The Unet++ network adds dense skip connections to reduce the semantic gap between feature maps, making it achieve good results in the general segmentation field.

However, in the field of road crack segmentation, because of the complex and changeable situation, a higher performance detection method is needed. In addition, in the actual segmentation, it is found that the cracks always appear in various irregular shapes. Additionally, it is accompanied by different complex backgrounds that interfere with the segmentation process, such as: zebra crossing backgrounds of different colors, water stains, stains, scratches, etc. In addition, due to the loss of information caused by the traditional down-sampling method of the network, the segmentation results show discontinuity of cracks, that is, the interruption of continuous crack segmentation. Based on these three questions, this paper uses Unet ++ as the backbone network to propose a HC-Unet++ network model (as shown in Figure 2a) to perform semantic segmentation on road cracks. In this model, we propose a deep parallel feature fusion module, as shown in Figure 2b, which can solve the problem of difficult segmentation of disordered fractures. The SEnet attention mechanism is introduced, as shown in Figure 2c, which can eliminate the interference of complex background in the road crack image, thereby improving the accuracy of segmentation. Replace the maximum pooling operation in the original network with the Blurpool pooling operation, as shown in Figure 2d. This operation reduces the loss of some fracture features during down-sampling, thereby reducing fracture discontinuity in the segmentation results. The following sections will provide more details.



Figure 2. Network structure diagram: (**a**) is the network structure of HC-Unet++. (**b**) is a flow chart of the DPFFB module. (**c**) is a flowchart of the SEnet attention mechanism. (**d**) is a workflow diagram of Blurpool pooling. (**e**) is the meaning of the corresponding modules in (**a**–**c**).

2.2.2. Deep Parallel Feature Fusion Module

In road crack segmentation, there will be various cracks with irregular shapes, which makes it difficult for the network to extract complete crack features, which brings certain difficulties to the actual segmentation. In order to improve the ability of the network to segment irregular cracks, it is necessary to add a special feature map processing module deep in the network. This module is used to obtain a larger receptive field and more deep semantic information to enhance the sensitivity of the network to the characteristics of road cracks. Therefore, the shape of the crack can be effectively judged, and the irregularly-shaped road crack can be identified.

However, in the Unet++ network, there is a lack of special processing of feature maps deep in the network, resulting in unsatisfactory segmentation capabilities for irregular cracks. In order to solve this problem, we propose a deep parallel feature fusion module (as shown in Figure 2b) to be placed into the bottom layer of the network (as shown in Figure 2a). It can be seen that the module placed at the bottom layer of the network can have a larger receptive field and be more sensitive to crack features. Therefore, the position of the road crack can be more accurately located and the shape of the crack can be judged. When the input feature map enters the module, the module will divide the feature map into upper and lower branches and perform independent and identical operations to obtain more complete crack features. It makes the network judge the shape of the crack more clearly.

The detailed operation steps of the deep parallel feature fusion module are as follows: (1) First, the input feature map is divided into upper and lower branches, and 1×1 convolution is performed on the upper and lower branches, respectively.

(2) Subsequently, the two branches perform BN normalization processing on the feature map, respectively. This will normalize the distribution of the data to the standard normal distribution, so that the input value of the subsequent activation function is in the sensitive area. The formula for BN processing is as follows:

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x_i \tag{1}$$

$$\sigma^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu)^2$$
(2)

$$y(i) = \gamma \frac{x_i - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \tag{3}$$

Among them, μ is the mean value in σ^2 a batch, is the variance in a batch, x_i is the input, y(i) is the output after the BN layer, γ and β is the learnable parameters, which will change with the gradient during the training process.

(3) After that, the upper branch and the lower branch, respectively, pass through the relu activation function to increase the nonlinear factor. The expression describing it is as follows:

$$y(i)' = \begin{cases} y(i) & y(i) > 0\\ 0 & y(i) \le 0 \end{cases}$$
(4)

Among them, y(i)' represents the output and y(i) represents the input.

The last two branches concatenate the extracted fracture features to strengthen the segmentation of irregular fracture shapes.

2.2.3. SEnet

In order to enhance the extraction of important features and reduce the interference of useless information, HU et al. [30] proposed the SEnet attention mechanism, which obtained a one-dimensional vector by global pooling the input feature map, and then sent the one-dimensional vector to the fully connected layer to make the one-dimensional vector have weight. Finally, the obtained one-dimensional weight was multiplied by the original input feature map to achieve the purpose of correcting the feature map. One of the difficulties in road crack detection is the elimination of complex backgrounds such as: zebra crossings of different colors, water stains, stains, shadows, etc. In order to effectively solve this problem, we introduced the SEnet attention mechanism, and replaced the original fully connected layer with a 1×1 convolutional layer to reduce the amount of calculation. At the same time, in the upsampling, in order to protect the integrity of the features, we set the scaling factor r to 2. Subsequently, the attention mechanism was placed in the deep layer of the network: $x^{3.0}$ between to $x^{4.0}$ and $x^{4.1}$ to $x^{3.1}$. In the deep layers of the network, the obtained 1D weight vector has a large receptive field. The one-dimensional weight is used to correct the input feature map to enhance the extraction of important features and suppress the interference of useless information, so as to achieve the purpose of eliminating complex backgrounds.

The specific implementation steps of the attention mechanism in this article are as follows:

(1) First, the input feature map *X* is globally pooled, and the one-dimensional feature vector $H'(1 \times 1 \times C)$ is generated by compressing *X* in the spatial dimension. This allows it to have per-channel global information. The H' calculation formula for is as follows:

$$H' = f_{sq}(X) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(i,j)$$
(5)

Among them, f_{sq} represents average pooling.

(2) The one-dimensional vector H' passes through the 1×1 convolution kernel so that the number of channels becomes 1/r. Then the activation function relu is entered to get the compressed one-dimensional vector H'' ($1 \times 1 \times C/r$). H'' ($1 \times 1 \times C/r$). The H'' calculation formula for is as follows:

$$H'' = \delta(w_1 a) \tag{6}$$

Among them, δ represents the activation function relu, and w_1 represents the dimensionality reduction parameter.

(3) The compressed one-dimensional vector is restored to the original number of channels H'' through a 1 × 1 convolution kernel, and it then enters the activation function sigmoid to obtain a weighted one-dimensional vector s (1 × 1 × C). Through (2), (3) two steps realize the interaction between channels, thus improving the computational efficiency of the network. The s calculation formula is as follows:

$$=\sigma(w_2a)\tag{7}$$

Among them, σ represents the activation function sigmoid, and w_2 represents the dimension-raising parameter.

S

(4) Finally, multiply the obtained one-dimensional weight s with the input feature map *X* to achieve the purpose of correcting the input feature map. This results in the output feature map *X'*. The *X'* calculation formula is as follows:

$$X' = f_{sacle}(s, X) = s_c X_c \tag{8}$$

2.2.4. Blurpool

Zhang et al. [37] found that modern convolutional networks were not displacement invariant. Commonly used downsampling methods such as maximum pooling and average pooling ignore the sampling theorem and do not insert low-pass filtering before downsampling. This will cause, when using these downsampling methods, small changes in the input that will cause the output value to fluctuate violently. This is not the desired result. In order to reduce the phenomenon wherein the output changes violently with the small displacement of the input, that is, to achieve displacement invariance to the greatest extent, the Blurpool pooling method is used to replace the maximum pooling in the original network. Its structure diagram is shown in Figure 2d, and its position in the network is shown in Figure 2a.

In the actual road crack identification process, the traditional maximum pooling operation will cause the use of each layer of down-sampling to cause the loss of some crack features, resulting in continuous crack interruptions in the segmentation results. In order to solve this problem, we introduce the Blurpool pooling method and use this operation to replace the original maximum pooling operation. The introduction of this operation can maximize translation invariance during downsampling to reduce the loss of some fracture features, thereby solving the phenomenon of fracture discontinuity in the segmentation results.

The specific implementation steps are as follows:

(1) Max operation with stride = 1. In this process, the operation has translation invariance and will not cause aliasing to information.

(2) Subsequently, a low-pass filter Blur is added before downsampling. Its function is to eliminate aliasing to the greatest extent and return the wrong output to the original position as much as possible. That is, the translation invariance is obtained to the greatest extent. The formula for translation invariance is as follows:

$$F(x) = F(shift_{\Delta h, \Delta w(x)}) \quad \forall (\Delta h, \Delta w)$$
(9)

Note, though that this formula is valid only when the translation amount is an integer multiple of N.

(3) Finally, downsample the module after low-pass filtering and output the result.

It can be seen that the aliasing of the downsampling has been greatly improved after adding the low-pass filter Blur. This strengthens the extraction of fracture features by the network during downsampling, and greatly reduces the discontinuity of fractures in the segmentation results.

2.3. Experimental Environment and Settings

2.3.1. Data Preparation

All tests in this study were performed on the same hardware and software platform. Table 2 is the hardware environment and software environment of this experiment.

Table 2. Experimental environment.

CPU	AMD EPYC 7543 32-Core Processor
ARM	80GB
Video memory	48GB
GPU	A40
OS	windows 11
PyTorch	1.11.0
Python	3.8
Ċuda	11.3
	CPU ARM Video memory GPU OS PyTorch Python Cuda

2.3.2. Training Methods

In order to avoid the mismatch between height and width, we adjust the height and width of the image to equal size, and the image input size is unified to 512×512 during training. The batch size is set to 2, the momentum parameter is 0.9, and the Adam optimizer is selected. The optimizer combines the advantages of two optimization algorithms, momentum and RMSProp, and can overcome the problem of sharp decrease of Adgrad gradient while automatically adapting to different learning rates for different parameters. This enables the network to avoid frequently updated parameters from being affected by a single outlier sample. Additionally, using cross-entropy loss as the loss function, a total of 300 epochs are trained. In the experiment, we divide the dataset of road crack detection into training set, verification set and test set with a ratio of 8:1:1. Table 3 contains the experimental parameters and settings. Table 3. Parameter Settings.

Size of Image	Batch_Size	Momentum	Initial lr	Optimizer	Iterations
512×512	2	0.9	e^{-4}	Adam	300

Figure 3 is the change of train loss and val loss with epoch. It can be seen from the figure that the network model tends to be stable at about 10 Epoch, and the network model converges quickly.



Figure 3. Train loss and val loss.

3. Experimental Results and Analysis

3.1. Experimental Evaluation Criteria

The evaluation index used in this experiment is composed of these five coefficients: *mIOU*, *mPA*, *mPrecision*, *dice*, and Hd95. In addition, in the following formula, *TP* is true (predicted result is a crack, and the actual result is a crack), *FP* false positive (predicted result is a crack, and the actual result is not cracked), and *FN* is false negative (the predicted result is non-cracked, and the actual result is a crack), *TN* is true negative (predicted result is non-crack, and actually is non-crack).

mIOU is the average intersection ratio, which represents the average intersection ratio of each class in this dataset, and its expression is as follows:

$$mIOU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FN + FP + TP}$$
(10)

mPA is the category average pixel accuracy rate, which represents the proportion of correctly classified pixels for each category, and then accumulates the average value. The expression about it is as follows:

$$mPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP + TN}{FN + TP + TN + FP}$$
(11)

mPrecision indicates the proportion of correct predictions in the samples that are predicted to be positive, and then accumulate and average, the expression is as follows:

$$mPrecision = \frac{1}{k+1} \sum_{i=0}^{k} \frac{TP}{FP+TP}$$
(12)

The value *dice* indicates the ratio of the intersection area between the predicted value and the real value to the total area, and is sensitive to internal filling. Its expression is as follows:

$$dice = \frac{2*TP}{TP + FN + TP + FP}$$
(13)

Hausdorff_distance is the maximum distance from one set to the nearest point of another set, and is sensitive to the segmented boundary. Its expression is as follows:

$$h(A,B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a,b) \} \}$$
(14)

In order to eliminate the unreasonable distance caused by some outliers, we refer to the Hd95 index. It can arrange the distances of these closest points in descending order, and select the distance ranked as 5% as the final value of Hd95.

3.2. Module Performance Analysis

This section tests the performance of each module we built in detail, divided into Section 3.2.1, Section 3.2.2 and Section 3.2.3. They respectively test the effectiveness of the deep parallel feature fusion module, the effectiveness of the SEnet module, and the effectiveness of the Blurpool module.

3.2.1. Effectiveness of Deep Parallel Feature Fusion Module

In this section, we use the self-made road crack dataset to verify the effectiveness of the deep parallel feature fusion module. We will use the SPP [38], ASPP [39] modules in comparison with the deep parallel feature fusion module, duly embedding them into the Unet++ network according to the position of DPFFB in the HC-Unet++ network. Table 4 is the comparison result of these feature extraction modules. It can be seen from the table that when the feature extraction block is embedded in the network, mIOU is improved and Hd95 is reduced by a certain value. Compared with SPP and ASPP, the mIOU obtained by embedding the DPFFB module is higher, and the Hd95 value is smaller. This shows that in the segmentation of road cracks, the performance of DPFFB is better than that of SPP and ASPP modules, which shows that this module is more suitable for the road crack segmentation task of this research. However, the embedding of the DPFFB module also brings a certain number of parameters, which inevitably increases the training time of the network.

Table 4. Comparison of feature extraction blocks.

Networks	Unet++	SPP+Unet++	ASPP+Unet++	DPFFB+Unet++
mIOU	70.39%	71.26%	71.39%	72.44%
Params	47.19 M	47.49 M	47.71 M	48.24 M
Hd95	12.62	10.16	10.03	8.16

3.2.2. Effectiveness of SEnet

To evaluate the effectiveness of the SEnet, we validate the SEnet module with the self-made road crack dataset. We will choose CA [40], CBAM [41] attention mechanism and SE attention mechanism for comparison. Additionally, we will imitate the location of SEnet in the HC-Unet++ network and embed them in the Unet++ network. Table 5 is the comparison result of these attention mechanisms. 3It can be seen from the table that after embedding the attention mechanism, the mIOU of the network model has been improved and Hd95 has been reduced. Compared with CBAM, although SEnet lacks the feature map correction of spatial attention, its data is slightly better than CBAM, and the overall parameter amount is also less than CBAM. Compared with CA, SEnet not only has better data than CA, but also has 0.78 M less parameters than it. This proves that the SEnet mechanism is more suitable for the road crack segmentation in this experiment.

Networks	Unet++	CBAM+Unet++	CA+Unet++	SE+Unet++
mIOU	70.39%	72.98%	72.84%	73.14%
Params	47.19 M	47.85 M	48.13 M	47.35 M
Hd95	12.62	8.31	9.03	7.96

Table 5. Comparison of Attention Mechanisms.

3.2.3. Effectiveness of Blurpool

To evaluate the effectiveness of Blurpool, we used the self-made road crack dataset to validate the Blurpool module. We conducted a comparative experiment between the Blurpool and the traditional Maxpool, and Figure 4 shows the loss training curve and accuracy curve of the two. It can be seen from the figure that when Blurpool is used instead of Maxpool, the convergence speed of network model training is accelerated and the accuracy is also improved, which shows that compared with Maxpool, Blurpool is more suitable for the network model in this paper.



Figure 4. Comparison of Maxpool and Blurpool.

3.3. Ablation Experiments

To evaluate the performance of the method proposed in this paper, we conduct ablation experiments. In the ablation experiment, we use a self-made road crack dataset for experiments and use Unet++ as the backbone structure of the network. On this basis, one or more methods proposed in this paper is added to compare and form ablation experiments to further prove the effectiveness of the deep parallel feature fusion module, SEnet module, and Blurpool module. The experimental results are shown in Table 6.

Table 6. Ablation H	Experiment Results
---------------------	--------------------

Number	Method	mIOU (%)	mPA (%)	mPrecision (%)	Hd95	Param
1	HC-Unet++	76.32	82.39	85.51	5.05	48.40 M
2	DPFFB+SE+Maxpool	75.12	81.12	84.69	5.83	48.40 M
3	SE+Blur	74.86	80.33	83.82	6.71	47.35 M
4	DPFFB+Blur	73.69	78.69	82.47	7.23	48.24 M
5	DPFFB+Maxpool	71.82	77.41	81.57	8.16	48.24 M
6	SE+Maxpool	72.54	76.20	82.05	7.96	47.35 M
7	Blur	71.16	75.21	81.94	11.26	47.19 M
8	Unet++	70.39	73.50	80.73	12.62	47.19 M

From Table 6 it can be seen that: DPFFB, Blurpool, and SEnet can all improve the accuracy in the process of road crack identification, and the HC-Unet++ model that combines these three methods performs the best. From the comparison between the fourth group and the seventh group, it can be seen that, after adding the DPFFB module, the Hd95 index changes significantly. This shows that after adding the DPPFB module, the model's ability to locate fractures is enhanced, and relatively complete fracture characteristics can be obtained, which improves the network's ability to identify irregular fractures. At the same time, the embedding of this module also brings some parameters. From the comparison between the third group and the seventh group, it can be seen that after adding the SEnet attention mechanism, mIOU, mPrecision, and mPA are significantly improved, and Hd95 is also significantly reduced. It shows that this module effectively eliminates the interference of complex background in the depth of the network, but it inevitably increases the number of parameters while improving the segmentation ability of the model. From the comparison between the fourth group and the fifth group, it can be seen that, after replacing Maxpool with Blurpool pooling, mIOU, mPrecision, and mPA have a slight increase and Hd95 has a slight decrease. Ultimately, the replacement of this module does not increase the computational load of the network.

These eight sets of experiments fully demonstrate the contribution of deep parallel feature fusion, SEnet, and Blurpool to the model's accuracy. It also shows that the HC-Unet++ proposed by us is more suitable for the detection of road cracks than Unet++.

In addition, we also compared the effects of DPFFB+SE+Maxpool, SE+Blur, and DPFFB+ Blur to further analyze the performance of this method, as shown in Table 7. In Figure a of Table 7, the Unet++ network is affected by the zebra crossing background, and some zebra crossings are misjudged as road cracks, while the other four networks perform relatively well. In Figure c of Table 7, Unet++ not only misses the judgment of road cracks, but also misjudges the manhole cover in the picture as a crack. After adding the DPFFB module and the SE module, this phenomenon of missed judgment and misjudgment disappears. This is because the network is more sensitive to fracture characteristics, can identify irregular fractures, and has improved anti-interference ability against complex backgrounds. However, the information is partially lost during downsampling, which makes the segmentation cracks appear discontinuous. After replacing Maxpool with Blurpool, the fracture interruption phenomenon of model segmentation is reduced, and the identified fractures are more fine and complete in comparison. This proves that the Blurpool module can alleviate the displacement variability of output information, reduce the loss of information during downsampling, and solve the problem of discontinuous crack segmentation. In the SE+Blur network in Figure d of Table 7, since the DPFFB module is erased, the network loses the special operation module for feature extraction. This leads to the reduction of the ability to divide irregular cracks in the network, so there is a serious phenomenon of missed judgments. When the DPFFB module is added, this missed judgment phenomenon disappears, which further proves the ability of the DPFFB module to identify irregular cracks. In Figure b of Table 7, when the SE module is erased from the HC-Unet++ network, the model misjudges it as a crack due to the black scratch background. After adding SEnet attention, the one-dimensional weight recalibrates the feature map to reduce the extraction of interference information. That is, the influence of the black scratch background is eliminated, which proves the ability of the SENet module to eliminate complex backgrounds.

		Detectio	n Result	
original				
Ground truth	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~	t		+
Unet++	/\) 	
DPFFB+SE+Maxpool	~~~			
DPFFB+Blur	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			
SE+Blur	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			
HC-Unet++	~~~~			
	(a)	(b)	(c)	(d)

Table 7. Visual comparison of the test results.

3.4. Comparsion of HC-Unet++ with Other Methods

To further analyze the performance of HC-Unet++, we compared it with state-of-theart network models. These network models are: BC-Dunet [42], U2-Net [43], CS2-Net [44]. Extremec3net [45], DCNet [46]; the experimental results about them are shown in

Table 8. In the table, we calculated the mIOU, mPA, mPrecision, dice, and Hd95 of different models for detecting road cracks. These five methods performed well on our self-made dataset, and their average mPA exceeded 80% and reached 80.39%. Among them, the mPrecision value of U2-Net surpassed 85.51% of HC-Unet++ network and reached 85.64%. However, compared with these five advanced network models, HC-Unet++ performed better, especially since the Hd95 value was only 5.05, which shows that the edge detection ability of the HC-Unet++ model in this paper is very good. In addition, the mIOU, mPA, and dice values of HC-Unet++ are all better than the other five networks, and the mPrecision value is only 0.13% lower than the mPrecision value in U2-Net. Overall, experimental

data show that our proposed HC-Unet++ network model is more suitable for road crack segmentation than some advanced network models.

Number	Method	mIOU (%)	mPA (%)	mPrecision (%)	Dice (%)	Hd95
1	HC-Unet++	76.32	82.39	85.51	70.26	5.05
2	BC-Dunet [42]	72.41	78.59	79.38	61.19	9.82
3	U2-Net [43]	73.28	80.63	85.64	63.74	11.68
4	CS2-Net [44]	73.19	79.50	82.73	64.51	7.34
5	Extremec3net [45]	74.76	81.99	81.98	67.84	9.57
6	DCNet [46]	72.53	81.24	83.75	63.49	12.58

Table 8. Comparison with advanced networks.

3.5. Generalization Experiments

In order to verify that our network model has good generalization ability, we used [47] concrete crack conglomerate dataset and Crack500 crack dataset, respectively, and used them in FCN [48], Unet [49], Unet++, HC-Unet++ models for training. Among them, the Concrete Crack Conglomerate Dataset has a total of 10,993 images and 1000 crack images are selected for training, and Crack500 has a total of 476 images and is used for training. They are divided into a training set, a verification set and a test set, in a ratio of 8:1:1; the experimental environment for training is the same as the experimental configuration. The experimental results are shown in Table 9. From the perspective of data, the performance of our HC-Unet++ model on these two different datasets is better than the other three comparative network models. Additionally, some of the indicators of the experimental results are even slightly better than the experimental data on the self-made dataset.

Dataset	Method	mIOU (%)	mPA (%)	mPrecision (%)	Dice (%)	Hd95
	HC-Unet++	77.23	86.45	85.91	71.38	3.17
Concrete Crack	FCN [48]	69.38	80.13	79.19	60.64	14.89
Conglomerate	Unet [49]	71.06	82.54	83.64	62.52	11.68
0	Unet++	73.71	81.67	82.91	67.98	9.98
	HC-Unet++	76.91	84.04	87.49	69.54	4.68
Crack 500	FCN [48]	70.41	79.65	78.91	59.94	13.29
	Unet [49]	73.95	83.24	81.03	65.23	10.68
	Unet++	73.83	82.94	84.57	64.26	9.35

Table 9. Generalization experiments.

In order to better demonstrate the good generalization performance of the HC-Unet++ model, we selected some renderings of the experiments in this section for comparison, as shown in Table 10. It can be seen from the figure that the performance of the FCN network is the worst. As shown in Figure b–f of Table 10, there are varying degrees of misjudgment and missed judgment which are caused by the lack of global context information in the network. The overall performance of the Unet and Unet++ networks is roughly the same, and both of them have slightly missed or misjudged. The HC-Unet++ network performs the best. It can be seen that most of the prediction maps are basically consistent with the label maps. This further demonstrates the excellent capabilities of our model on these two datasets. It shows that our network has good generalization ability on other datasets.

		Crack 500	0 Concrete Crack Conglomerate Dataset			k taset
original						~~~
Ground truth				$ \rightarrow $		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
HC-Unet++						~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Unet++			 }			
Unet [49]						\sim
FCN [48]						~~~
	а	b	с	d	e	f

Table 10. Generalization experiment effect diagram.

4. Discussion

To segment road cracks, in this paper we proposed the HC-Unet++ network model. In this study, we construct a new road crack dataset and use HC-Unet++ to train the network. Experiments show that our proposed HC-Unet++ network model is effective for segmenting road cracks. To a certain extent, it can solve the complex background of road cracks, irregular cracks and other problems. Nonetheless, we still need more research:

(1) HC-Unet++ networks are relatively large, so the training takes a relatively long time. How to reduce the parameters of the network without affecting the accuracy is a problem we need to solve in the future.

(2) As shown in Figure 5, in the process of HC-Unet++ network segmentation: When the UAV is flying at a high altitude, cracks will appear in relatively small forms in the image, and the result of model segmentation at this time is poor. In the future, we will need to add an efficient module to effectively segment out small cracks.

(3) There is a lack of quantification of cracks, and physical parameters such as the length, width, and area of road cracks cannot be obtained. In order to quantify cracks, we need to develop an effective method to quantify cracks in the future.



Figure 5. Segmentation of fine cracks.

5. Conclusions

This paper proposes HC-Unet++ road crack segmentation recognition technology. First of all, the road-crack dataset of this experiment was constructed by using camera shooting and UAV aerial photography, and the label processing was performed using the labelme tool. Then it was input into the HC-Unet++ network model for training. HC-Unet++ used Unet++ as the basic network structure, and embedded a deep parallel feature fusion module to improve the sensitivity of the network to crack features. This enabled the network to obtain relatively complete fracture information, thereby identifying irregular fractures. After adding the SEnet attention mechanism to eliminate the complex background interference in the road cracks, and replacing the Maxpool with the Blurpool module, the variability of displacement was greatly alleviated. This increased the network's extraction of some fracture features in the down-sampling of each layer, making the fractures obtained by network segmentation more continuous. The experimental results show that the HC-Unet++ network model achieves an average intersection ratio of 76.32%, an average pixel accuracy rate of 82.39%, an average precision rate of 85.51%, and Hausdorff 95 is 5.05. As well, in the generalization experiment, the performance of our network model is still stable, which shows that the HC-Unet++ network has good adaptability and provides data value for road maintenance and traffic safety.

Road crack segmentation is still an important research direction in the engineering field of image recognition technology, a task which is of great significance for prolonging the service life of roads and reducing traffic safety hazards. Although the segmentation and detection of road cracks based on convolutional neural network has achieved outstanding results, further improvement is still needed. The next step of this research will be to think about how to compress the network scale without affecting the segmentation accuracy and the physical quantification of the road cracks.

Author Contributions: H.C.: Methodology, Writing—original draft, Conceptualization. Y.G.: Software, Data acquisition, Investigation. W.C.: Model guidance. Z.X.: Validation, Project administration. L.L.: Visualization, Writing—Review and Editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation in China (Grant No. 61703441); in part by the Key Project of Education Department of Hunan Province (Grant No. 21A0179); in part by the Changsha Municipal Natural Science Foundation (Grant No. kq2014160); in part by Hunan Key Laboratory of Intelligent Logistics Technology (2019TP1015).

Data Availability Statement: All self-made datasets for this study are available by contacting the authors.

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

References

- Hu, G.X.; Hu, B.L.; Yang, Z.; Huang, L.; Li, P. Pavement crack detection method based on deep learning models. Wirel. Commun. Mob. Comput. 2021, 2021, 5573590. [CrossRef]
- Ren, J.; Zhao, G.; Ma, Y.; Zhao, D.; Liu, T.; Yan, J. Automatic Pavement Crack Detection Fusing Attention Mechanism. *Electronics* 2022, 11, 3622. [CrossRef]
- Johnson, A.M. Best Practices Handbook on Asphalt Pavement Maintenance; Minnesota Technology Transfer/LTAP Program, Center for Transportation Studies: Minneapolis, MN, USA, 2000.
- 4. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under concept drift: A review. *IEEE Trans. Knowl. Data Eng.* 2018, 31, 2346–2363. [CrossRef]
- Xing, Z. An improved emperor penguin optimization based multilevel thresholding for color image segmentation. *Knowl.-Based Syst.* 2020, 194, 105570. [CrossRef]
- Tang, J.; Gu, Y. Automatic crack detection and segmentation using a hybrid algorithm for road distress analysis. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013; pp. 3026–3030.
- Peng, L.; Chao, W.; Shuangmiao, L.; Baocai, F. Research on crack detection method of airport runway based on twice-threshold segmentation. In Proceedings of the 2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC), Qinhuangdao, China, 18–20 September 2015; pp. 1716–1720.
- Li, H.; Song, D.; Liu, Y.; Li, B. Automatic pavement crack detection by multi-scale image fusion. *IEEE Trans. Intell. Transp. Syst.* 2018, 20, 2025–2036. [CrossRef]
- 9. Xu, D.; Zhao, Y.; Jiang, Y.; Zhang, C.; Sun, B.; He, X. Using improved edge detection method to detect mining-induced ground fissures identified by unmanned aerial vehicle remote sensing. *Remote Sens.* **2021**, *13*, 3652. [CrossRef]
- Zhao, H.; Qin, G.; Wang, X. Improvement of canny algorithm based on pavement edge detection. In Proceedings of the 2010 3rd international congress on image and signal processing, Yantai, China, 16–18 October 2010; pp. 964–967.
- Liang, S.; Sun, B. Using wavelet technology for pavement crack detection. In ICLEM 2010: Logistics for Sustained Economic Development: Infrastructure, Information, Integration, Proceedings of the International Conference of Logistics Engineering and Management (ICLEM) 2010, Chengdu, China, 8–10 October 2010; Zhang, J., Xu, L., Zhang, X., Yi, P., Jian, M., Eds.; American Society of Civil Engineers: Reston, VA, USA, 2010; pp. 2479–2484.
- Subirats, P.; Dumoulin, J.; Legeay, V.; Barba, D. Automation of pavement surface crack detection using the continuous wavelet transform. In Proceedings of the 2006 International Conference on Image Processing, Atlanta, GA, USA, 8–11 October 2006; pp. 3037–3040.
- 13. Cheng, Y.; Lin, M.; Wu, J.; Zhu, H.; Shao, X. Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network. *Knowl.-Based Syst.* **2021**, *216*, 106796. [CrossRef]
- 14. Liu, W.; Quijano, K.; Crawford, M.M. YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 8085–8094. [CrossRef]
- Wang, B.; Yan, Z.; Lu, J.; Zhang, G.; Li, T. Explore uncertainty in residual networks for crowds flow prediction. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–7.
- Hu, Y.; Zhan, J.; Zhou, G.; Chen, A.; Cai, W.; Guo, K.; Hu, Y.; Li, L. Fast forest fire smoke detection using MVMNet. *Knowl.-Based* Syst. 2022, 241, 108219. [CrossRef]
- 17. Zhu, W.; Zhang, H.; Eastwood, J.; Qi, X.; Jia, J.; Cao, Y. Concrete crack detection using lightweight attention feature fusion single shot multibox detector. *Knowl.-Based Syst.* 2023, 261, 110216. [CrossRef]
- Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* 2017, 32, 361–378. [CrossRef]
- Fan, R.; Bocus, M.J.; Zhu, Y.; Jiao, J.; Wang, L.; Ma, F.; Cheng, S.; Liu, M. Road crack detection using deep convolutional neural network and adaptive thresholding. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 474–479.

- Sadrawi, M.; Yunus, J.; Abbod, M.F.; Shieh, J.-S. Higher Resolution Input Image of Convolutional Neural Network of Reinforced Concrete Earthquake-Generated Crack Classification and Localization. *IOP Conf. Ser. Mater. Sci. Eng.* 2020, 931, 012005. [CrossRef]
- 21. Huang, H.-W.; Li, Q.-T.; Zhang, D.-M. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn. Undergr. Space Technol.* **2018**, *77*, 166–176. [CrossRef]
- Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. Deepcrack: Learning hierarchical convolutional features for crack detection. IEEE Trans. Image Process. 2018, 28, 1498–1512. [CrossRef]
- 23. Yuan, G.; Li, J.; Meng, X.; Li, Y. CurSeg: A pavement crack detector based on a deep hierarchical feature learning segmentation framework. *IET Intell. Transp. Syst.* 2022, *16*, 782–799. [CrossRef]
- Billah, U.H.; Tavakkoli, A.; La, H.M. Concrete crack pixel classification using an encoder decoder based deep learning architecture. In Proceedings of the Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, 7–9 October 2019; Proceedings, Part I 14. pp. 593–604.
- Li, C.; Wen, Y.; Shi, Q.; Yang, F.; Ma, H.; Tian, X. A pavement crack detection method based on multiscale Attention and HFS. Comput. Intell. Neurosci. 2022, 2022, 1822585. [CrossRef]
- Yang, L.; Fan, J.; Huo, B.; Li, E.; Liu, Y. A nondestructive automatic defect detection method with pixelwise segmentation. *Knowl.-Based Syst.* 2022, 242, 108338. [CrossRef]
- Zhang, Y.; Huang, J.; Cai, F. On bridge surface crack detection based on an improved YOLO v3 algorithm. *IFAC-Pap.* 2020, 53, 8205–8210. [CrossRef]
- Yang, J.; Li, H.; Zou, J.; Jiang, S.; Li, R.; Liu, X. Concrete crack segmentation based on UAV-enabled edge computing. *Neurocomput*ing 2022, 485, 233–241. [CrossRef]
- 29. Qiao, W.; Liu, Q.; Wu, X.; Ma, B.; Li, G. Automatic pixel-level pavement crack recognition using a deep feature aggregation segmentation network with a scse attention mechanism module. *Sensors* **2021**, *21*, 2902. [CrossRef]
- Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
- Xiang, X.; Zhang, Y.; El Saddik, A. Pavement crack detection network based on pyramid structure and attention mechanism. *IET Image Process.* 2020, 14, 1580–1586. [CrossRef]
- 32. Han, C.; Ma, T.; Huyan, J.; Huang, X.; Zhang, Y. CrackW-Net: A novel pavement crack image segmentation convolutional neural network. *IEEE Trans. Intell. Transp. Syst.* 2021, 23, 22135–22144. [CrossRef]
- 33. Chen, J.; Yuan, Y.; Lang, H.; Ding, S.; Lu, J.J. The Improvement of Automated Crack Segmentation on Concrete Pavement with Graph Network. *J. Adv. Transp.* 2022, 2022, 2238095. [CrossRef]
- Liu, W.; Xia, X.; Xiong, L.; Lu, Y.; Gao, L.; Yu, Z. Automated vehicle sideslip angle estimation considering signal measurement characteristic. *IEEE Sens. J.* 2021, 21, 21675–21687. [CrossRef]
- Rehak, M.; Skaloud, J. Time synchronization of consumer cameras on Micro Aerial Vehicles. *ISPRS J. Photogramm. Remote Sens.* 2017, 123, 114–123. [CrossRef]
- 36. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. Unet++: A nested u-net architecture for medical image segmentation. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, 20 September 2018; Proceedings 4. pp. 3–11.
- Zhang, R. Making convolutional networks shift-invariant again. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 10–15 June 2019; pp. 7324–7334.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 37, 1904–1916. [CrossRef]
- 39. 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]
- Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
- 41. 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.
- 42. Liu, T.; Zhang, L.; Zhou, G.; Cai, W.; Cai, C.; Li, L. BC-DUnet-based segmentation of fine cracks in bridges under a complex background. *PLoS ONE* **2022**, *17*, e0265258. [CrossRef]
- Qin, X.; Zhang, Z.; Huang, C.; Dehghan, M.; Zaiane, O.R.; Jagersand, M. U2-Net: Going deeper with nested U-structure for salient object detection. *Pattern Recognit.* 2020, 106, 107404. [CrossRef]
- 44. Mou, L.; Zhao, Y.; Fu, H.; Liu, Y.; Cheng, J.; Zheng, Y.; Su, P.; Yang, J.; Chen, L.; Frangi, A.F. CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging. *Med. Image Anal.* **2021**, *67*, 101874. [CrossRef]
- 45. Park, H.; Sjösund, L.L.; Yoo, Y.; Bang, J.; Kwak, N. Extremec3net: Extreme lightweight portrait segmentation networks using advanced c3-modules. *arXiv* 2019, arXiv:1908.03093.
- Li, F.; Li, W.; Gao, X.; Liu, R.; Xiao, B. DCNet: Diversity convolutional network for ventricle segmentation on short-axis cardiac magnetic resonance images. *Knowl.-Based Syst.* 2022, 258, 110033. [CrossRef]
- 47. Bianchi, E.; Hebdon, M. Concrete Crack Conglomerate Dataset; University Libraries, Virginia Tech: Blacksburg, VA, USA, 2021.

- 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.
- Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Proceedings, Part III 18. pp. 234–241.

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.