MDPI

Article

GGMNet: Pavement-Crack Detection Based on Global Context Awareness and Multi-Scale Fusion

Yong Wang [1,*], Zhenglong He [1,2], Xiangqiang Zeng [1,2], Juncheng Zeng [3], Zongxi Cen [1,2], Luyang Qiu [4], Xiaowei Xu [3] and Qunxiong Zhuo [4]

1   State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; hezhenglong22@mails.ucas.ac.cn (Z.H.); zengxiangqiang21@mails.ucas.ac.cn (X.Z.); cenzongxi23@mails.ucas.ac.cn (Z.C.)
2   College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
3   Fujian Expressway Science & Technology Innovation Research Institute Co., Ltd., Fuzhou 350001, China
4   Fujian Luoning Expressway Co., Ltd., Fuzhou 350001, China
*   Correspondence: wangy@igsnrr.ac.cn; Tel.: +86-10-6488-8179

Abstract: Accurate and comprehensive detection of pavement cracks is important for maintaining road quality and ensuring traffic safety. However, the complexity of road surfaces and the diversity of cracks make it difficult for existing methods to accomplish this challenging task. This paper proposes a novel network named the global graph multiscale network (GGMNet) for automated pixel-level detection of pavement cracks. The GGMNet network has several innovations compared with the mainstream road crack detection network: (1) a global contextual Res-block (GC-Resblock) is proposed to guide the network to emphasize the identities of cracks while suppressing background noises; (2) a graph pyramid pooling module (GPPM) is designed to aggregate the multi-scale features and capture the long-range dependencies of cracks; (3) a multi-scale features fusion module (MFF) is established to efficiently represent and deeply fuse multi-scale features. We carried out extensive experiments on three pavement crack datasets. These were DeepCrack dataset, with complex background noises; the CrackTree260 dataset, with various crack structures; and the Aerial Track Detection dataset, with a drone's perspective. The experimental results demonstrate that GGMNet has excellent performance, high accuracy, and strong robustness. In conclusion, this paper provides support for accurate and timely road maintenance and has important reference values and enlightening implications for further linear feature extraction research.

Keywords: pavement cracks; deep learning; attention mechanism; graph reasoning; multi-scale features fusion

1. Introduction

Roads serve as the foundation of contemporary economic growth and sustainability. The operational and structural status of roads are pivotal factors in shaping the economic landscape of a nation and are deemed essential criteria by the World Bank for assessing the competitiveness of national economies [1]. However, due to traffic loads, construction defects, and environmental conditions, road surfaces are susceptible to a variety of defects, the most prevalent of which are cracks. If these defects are not professionally repaired promptly, they can seriously affect the driving quality and safety of traffic. Therefore, it has become imperative to monitor and assess pavement conditions more frequently. Nevertheless, the departments and agencies of pavement management face the challenges of traditional inspection based on manual detection, which is not only inefficient and costly, but also relies heavily on human subjective factors. When the inspection staff are fatigued, they tend to misidentify or overlook crack information during detection, which could significantly impact the accuracy of crack detection. Therefore, there is an urgent need

for a high-efficiency and automated crack-detection method to support refined pavement maintenance operations.

For many years, crack detection has mostly relied on traditional image-processing methods, such as threshold segmentation [2–5], morphological method [6], wavelet transform [7,8], and artificial feature engineering [9,10]. However, these traditional techniques have many drawbacks, resulting in pavement-crack detection algorithms relying on traditional image-processing techniques have not been effectively applied in engineering practice. For example, threshold segmentation is highly susceptible to road surface background information and expert knowledge [2], and the detection results depend largely on the set hyperparameters and can only extract crack information that is significantly different from the background lightness. Compared to threshold segmentation, although the semi-automatic method based on morphology maintains the structure of the cracks and enhances the quality of the segmented image, it is still difficult to completely extract crack features with complex structure, and it remains sensitive to noise information [6]. The wavelet transform can effectively strike a balance between suppressing noise and portraying edge details [7], but it is less effective in extracting cracks with uneven signal strength. The artificial feature engineering techniques have improved the accuracy of crack extraction, and they are time-consuming and labor-intensive because of the requirement for manual feature design and exhibit poor robustness [9]. In conclusion, although traditional image-processing techniques can detect cracks, these methods demonstrate poor robustness and low accuracy and encounter difficulties in meeting the demand for high-precision and fully automatic intelligent detection of pavement cracks.

Deep learning has exhibited exceptional performance in various tasks both upstream and downstream of computer vision in recent years, such as image-level classification [11], object-level detection [12], and pixel-level segmentation [13,14]. Built upon these achievements, deep learning has been widely employed for pavement-crack detection and attained satisfactory results. Some researchers have embedded CNNs into edge detectors to enhance the accuracy of crack edge detection. For example, the HED [15] for extracting edges of images, consisting of CNNs combined with edge detectors, has achieved better results than have traditional edge detectors. The model named RCF [16], with deeper convolutional layers, obtains better performance on pavement-crack detection compared to HED. More recently, to achieve more robust results and superior performance, research on crack detection has increasingly shifted toward the end-to-end and pixel level. The FCN [17] introduced to crack detection effectively alleviates the issue of low efficiency in crack detection. The use of symmetrical encoder–decoder models, exemplified by U-Net [18] and SegNet [19], has led to further improvement in detection accuracy. Due to the complexity of road-surface backgrounds, segmentation models with larger receptive fields such as PSPNet [20], Deeplabv3+ [21], TransUNet [22], and SwinUNet [23] have been employed for crack detection. Additionally, some specialized deep neural networks for crack extraction have been designed. Zou et al. [24] constructed a multi-level fusion structure based on SegNet for crack segmentation. Liu et al. [25] proposed a network with CNNs and Transformer, which significantly enhanced the extraction effectiveness of cracks. Bai et al. [26] designed a dual-path crack extraction network, enhancing the ability to describe complex crack features. Zhang et al. [27] proposed a network based on deformable convolution to adapt to the morphology of cracks. However, the following issues are still common: (1) insufficient global contextual awareness—due to the diversity and complexity of road scenes, simple convolutions with inadequate global contextual awareness fail to capture the spatial correlations of crack features, resulting in significant impact from background noise on crack detection; (2) inadequate capability to integrate multi-scale features—on account of the diversity of crack structures, the results from a single scale often fail to accurately and comprehensively represent crack information, limiting the performance of the model.

In addressing these aforementioned issues, this paper introduces a pixel-level crack detection network (GGMNet). The results from experiments conducted on three public datasets demonstrate the excellent performance, high accuracy, and strong robustness of

GGMNet. Specifically, the DeepCrack dataset was utilized to evaluate the ability of GGM-Net in extracting cracks within complex scenarios, the CrackTree260 dataset was employed to demonstrate our model's proficiency in recognizing various types of crack information, and the Aerial Track Detection dataset was employed to assess the generalization capability of the GGMNet under different viewpoints. The primary contributions of this paper can be outlined as follows:

(1) A novel network for pavement-crack detection with excellent accuracy and strong robustness was constructed.

(2) In the context of complex backgrounds, the GC-Resblock was constructed to guide the model to focus more on crack-information extraction. Specialized for intricate crack structures, the GPPM was innovatively designed to effectively aggregate features of various sizes and shapes. Moreover, the MFF was constructed to reduce the probability of missing detection.

2. Methods

This Section provides a comprehensive explanation of the proposed GGMNet architecture, the details of each model component, and the employed deep supervision training strategy.

2.1. Model Overview

The proposed GGMNet is presented in Figure 1 and comprises the encoder, decoder, and multi-scale feature fusion module. Firstly, the network takes crack images as input into the encoder, which utilizes GC-Resblock to obtain the local spatial and global contextual information of cracks. Continuously, GPPM processes features from the encoder using graph reasoning operators and a pooling pyramid structure, enriching them with higher-dimensional and higher-order feature representations. Finally, the decoder gradually restores crack features to the original resolution and employs MFF to integrate feature information from different levels, thereby obtaining output results with rich spatial and semantic information.
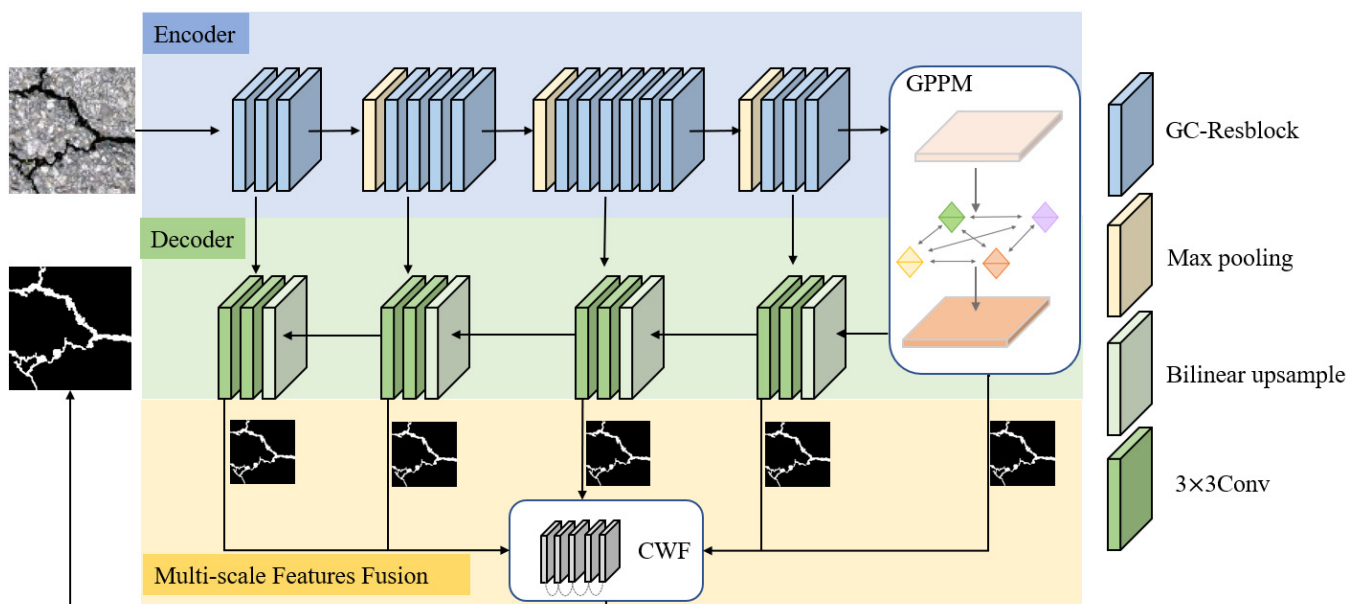

Figure 1. Network framework of the proposed GGMNet.

2.2. Global Contextual Res-Block

Realistic scenes of crack images frequently contain complex backgrounds, such as shadows, oil stains, debris, and garbage. If these types of noise are not suppressed, the performance of the model will be weakened. Therefore, a global contextual Res-block

(GC-Resblock) is proposed as the basic module of the encoder to guide the network to emphasize the identities of cracks. The structure is depicted in Figure 2.
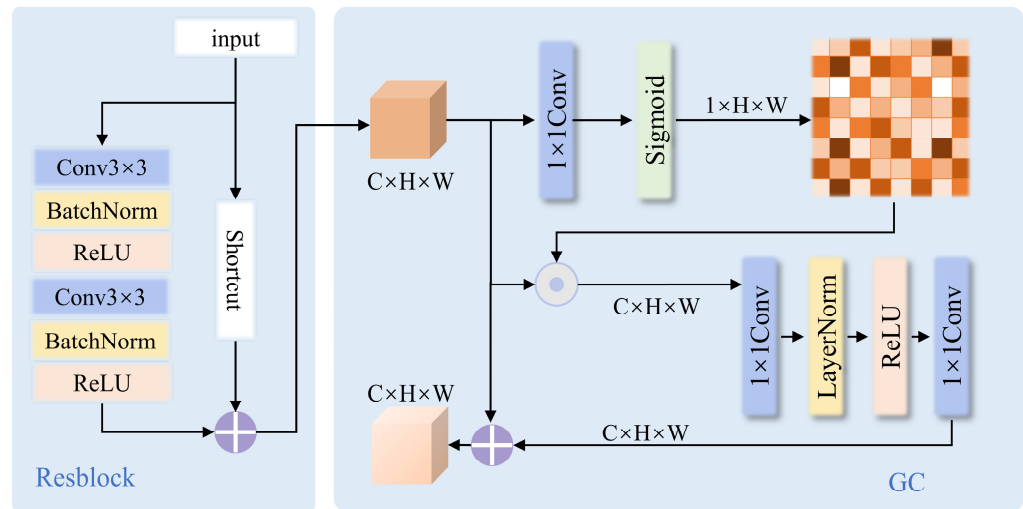

Figure 2. Architecture of the proposed GC-Resblock.

As indicated in Figure 2, GC-Resblock incorporates two components: a residual block [28] and a global contextual unit [29]. The residual block extracts local spatial information of crack features through convolutional operations and utilizes an identity mapping mechanism to aid in model training. The global semantic unit is utilized to grasp the overall information of the image, adaptively enhancing the output feature representation from the residual block.

(1) Residual block: The residual block consists mainly of two 3×3 convolutions and a residual connection. Batch normalization and rectified linear unit are employed subsequent to every convolutional layer to normalize hidden features and address the issue of nonlinear activation. The success of this module lies in its ability to preserve original features to a certain extent, thereby providing some guarantee for gradient backpropagation.

$$y = x \oplus f_{cbr}(f_{cbr}(x)) \tag{1}$$

$$f_{cbr}(x) = \delta_r(f_{bn}(f_{conv}(x))) \tag{2}$$

where f_{conv} represents 3×3 convolution, f_{bn} represents batch normalization, and δ_r represents the relu activation function.

(2) Global contextual unit: The features outputted by the residual block can represent shallow information but often contain significant noise that requires further processing. Therefore, this paper introduces the global contextual unit following the residual block. Initially, this module employs 1×1 convolution and the sigmoid activation function to produce pixel attention weights, where the highlighted regions denote cracks and lowlighted regions represent the background. Subsequently, the input maps are multiplied by the generated attention weights to obtain redistributed results. Further, convolutional operations are applied to recalibrate inter-channel relationships, yielding output results that incorporate spatial and channel attention processing.

$$y' = y \oplus f_{cw}(f_{pw}(y) \odot y) \tag{3}$$

$$f_{cw}(x) = f_{conv}(\delta_r(f_{\ln}(f_{conv}(x)))) \tag{4}$$

$$f_{pw}(x) = \delta_s(f_{conv}(x)) \tag{5}$$

where y and y' denote the features from the residual block and final features, respectively; f_{pw} and f_{cw} represent pixel operation and channel operation, respectively; f_{conv} represents

1×1 convolution; f_{ln} is layer normalization; and δ_s and δ_r denote sigmoid activation function and relu activation function, respectively.

2.3. Graph Pyramid Pooling Module

The use of GC-Resblock effectively mitigates the expression of irrelevant information and enhances the performance of the model in complex backgrounds. However, due to the uneven force strength on road surface, the cracks exhibit diverse structures. The network faces challenges in learning crack features of varying shapes and sizes. Therefore, to meet the requirement of extracting complex crack structures, this paper establishes the graph pyramid pooling module (GPPM), as depicted in Figure 3.
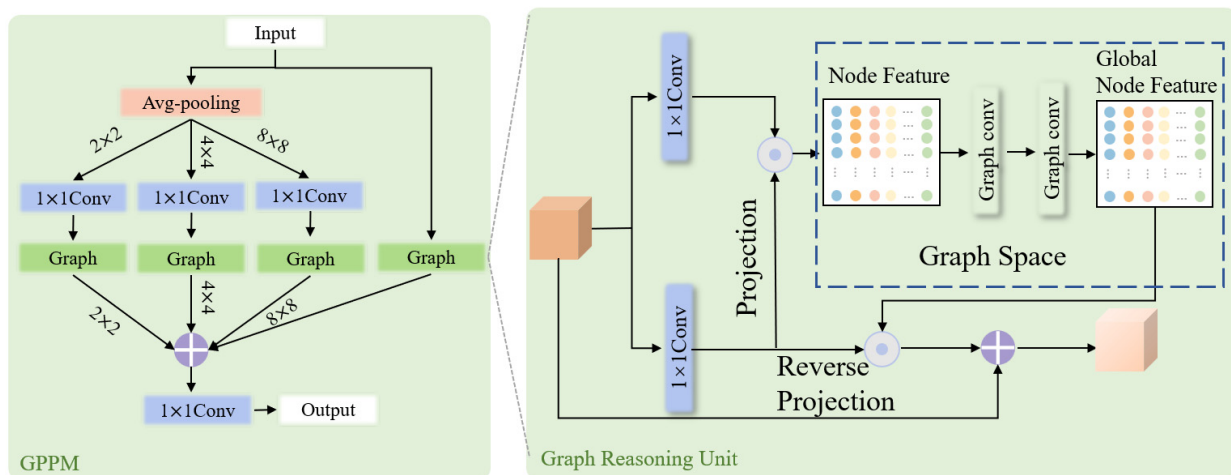

Figure 3. Framework of the proposed GPPM.

As illustrated in Figure 3, the GPPM comprises two components: a pooling pyramid module [30] and a graph reasoning unit [31]. The pooling pyramid module operates multi-scale features by pooling layers of different sizes, thereby aggregating crack information of different sizes. The graph reason block aims at perceiving global contextual information, capturing the relationships between disjoint regions with irregularly shaped relationships.

(1) Pooling pyramid module: The feature map undergoes average-pooling operations using various kernel sizes, followed by a 1×1 convolution. To preserve the original feature, one path is not pooled. Then, these feature maps are introduced into the graph reasoning unit to gain more contextual information, which is followed by elementwise addition and 1×1 convolution to obtain the output feature.

(2) Graph reasoning unit: The traditional convolutions can only handle pixels in the neighborhood, and they often struggle to effectively capture long-range global relationships between distant regions and require the use of multiple stacked convolutional layers. The advantage of graph convolutions is the ability to directly capture the contextual information of the entire graph. Thus, the graph reasoning unit is embedded in the pooling pyramid module for establishing the relationship of distant regions. As shown in Figure 3, the detailed operation is described below.

Projection: Before engaging in comprehensive graph-based relational inference, a prerequisite involves the transformation of features from the coordinate space, facilitating their projection and mapping onto the graph space. As shown in Figure 4, in contrast to feature map $g \in \mathbb{R}^{C \times H \times W}$ in the coordinate space, the projected feature map $G \in \mathbb{R}^{C \times N}$ in the graph space stores the features through the nodes. The projection function is learned by two 1×1 convolutions followed by elementwise multiplication.

$$G = f_{conv1}(g) \odot f_{conv2}(g) \tag{6}$$

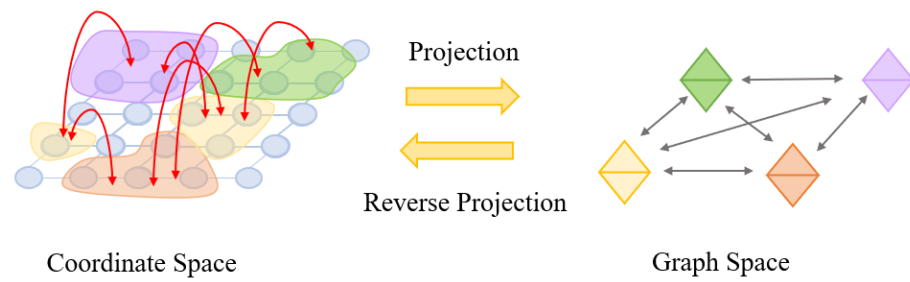where f_{conv1} and f_{conv2} both represent 1×1 convolution.

Figure 4. Schematic diagram of spatial reasoning.

Graph reasoning: The nodes denote the semantics of the original feature and facilitate the identification of relationships in distant and irregular regions. To grasp the attributes of the related nodes, the contextual relationship between each pair of nodes is represented and reasoned through the application of graph convolutions.

$$G' = f_g(G) \odot A \tag{7}$$

where f_g denotes the state update function of nodes in graph convolution, A represents the node adjacency matrix, and these two parameters are both learnable.

Reverse projection: The final step involves mapping the output features to return to the original coordinate space following the reasoning of relationships. Reverse projection is very similar to the projection.

$$G'' = G \oplus (f_{conv3}(G) \odot G') \tag{8}$$

where f_{conv2} and f_{conv3} are mutually inverse matrices.

In conclusion, the GPPM aggregates multi-scale features by a pooling pyramid module and captures the relationship of arbitrarily shaped cracks with a graph reasoning unit. This module enhances the ability of GGMNet to identify various sizes and shapes of cracks.

2.4. Multi-Scale Feature Fusion

To reduce the probability of missing detection and to ensure the accuracy of detecting pavement cracks, the multi-scale feature fusion (MFF) is constructed to assemble the feature maps of layers and avoid missing contextual and spatial information. As depicted in Figure 1, the feature maps of each layer are upsampled to the scale of 256×256 and subsequently introduced into the channel-weighting fusion unit (CWF), which is designed to learn and assign the weights of each channel. Compared to previous studies [32,33], we presume that the importance of each feature is different, and the relationship of these maps is not explored, the effective complementary knowledge will be overlooked, but redundant information will be retained. As shown in Figure 5, the channel-weighting fusion unit is composed of convolution, pooling and sigmoid activation function. The specific computational formula for this unit is as follows:

$$z' = z \odot f_{pcs}(z) \tag{9}$$

$$f_{pcs}(x) = \delta_s(f_{conv}(f_p(x))) \tag{10}$$

where z and z' represent input maps and output maps, respectively; f_{conv} represents 1×1 convolution; f_p represents global average pooling operation; and δ_s denotes sigmoid activation function.

Through the channel-weighting fusion unit, the same weight is shared among various spatial positions within the feature channel, while feature weights for different channels are redistributed. The useless channel information will be suppressed, and important channel information will be prominently expressed.
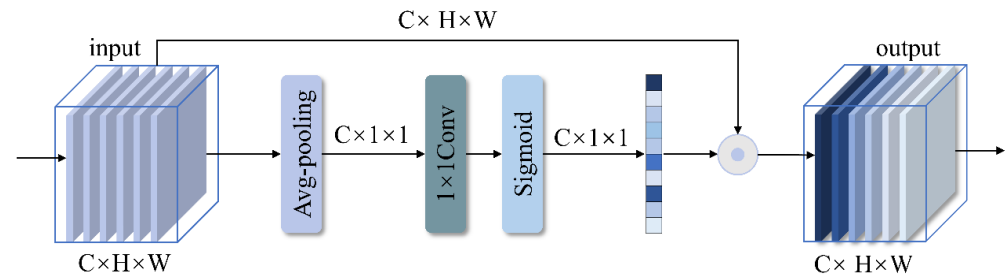
Figure 5. Framework of the proposed CWF.

2.5. Loss Function

The BCE loss function exhibits superior efficacy in image segmentation missions, indicating the disparities between ground truth and the predicted result [34]. However, the issue of significant imbalance between positive and negative samples exists in crack detection, and if only the BCE loss function is chosen, the model may fail to obtain the global optimal effect. Thus, the Dice loss function [35], which is designed to lighten the imbalance issues, is introduced into the training process.

$$l_{(Y,Y^*)} = l_{BCE} + l_{Dice} \tag{11}$$

$$l_{BCE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y^* \cdot \log Y_p + (1 - Y_p^*) \cdot \log(1 - Y_p) \right) \tag{12}$$

$$l_{Dice} = 1 - \frac{2 \times TP}{2 \times TP + FP + FN} \tag{13}$$

where $l_{(Y,Y^*)}$, l_{BCE} and l_{Dice} denote the total loss function, BCE, and Dice loss function, respectively; N denotes the total number of pixels in the image; Y_p^*, Y_p represent true and predicted values at pixel P, respectively; TP denotes the true-positive samples predicted as positive by the network; FP represents the negative samples predicted as positive by the network; and FN is the positive samples predicted as negative by the network.

In addition, a deep supervision mechanism [36] is separately applied to each output layer, which aims at enhancing the network's segmentation accuracy and accelerating the convergence speed of segmentation.

$$L = \sum_{m}^{M} \alpha_m L_{side}(Y_m^*, Y_m) + L_{fuse}(Y^*, Y) \tag{14}$$

where M represents the count of output layers; α_m denotes the weight of each output layer; and L_{side} and L_{fuse} represent the loss of each output layer and the fused predicted result, respectively.

3. Experimental Dataset and Setup

3.1. Datasets

(1) DeepCrack: The DeepCrack dataset was collected by Y. Liu et al. [24], and it contains 537 images of concrete pavement cracks with different scenes and light conditions. All crack images in the dataset are 544×384 pixels. In particular, a substantial amount of noise exists in this dataset in the form of such imagery as shadows, oil stains, and different shapes of road debris.

(2) CrackTree260: The CrackTree260 dataset was collected by Q. Zou et al. [37] This dataset comprises 260 pavement crack images of size 800×600 pixels, with multiple crack types, such as transverse, longitudinal, mesh, and block. The cracks in this dataset show various sizes and shapes, and it contains a number of relatively narrow cracks.

(3) Aerial Track Detection: The Aerial Track Detection dataset was collected by Z. Hong et al. [38]. In contrast to the images in both of the above datasets, the crack images

acquired under the unmanned aerial vehicle perspective for this dataset include 4118 post-earthquake pavement cracks with each image size being 512×512 pixels. This dataset is applied for training and testing to verify the robustness of our network.

3.2. Parameter Setting

The GGMNet proposed relied on the PyTorch framework, and the NVIDIA RTX A5000 was employed to expedite the training of the model. The specific parameters of the training process are presented in Table 1. In this study, the datasets were partitioned into training sets, validation sets, and testing sets with a division ratio of 6:2:2. To mitigate the risk of overfitting, data augmentation techniques, including random horizontal–vertical flipping, random cropping, and random color mapping, were used during training.

Table 1. The parameter settings.

Item	Setting
Epoch	100
Batch size	4
Optimizer	Adam [39–41]
Initial learning rate	1×10^{-4} [39,41,42]
Minimum learning rate	1×10^{-6} [39,41,42]
Momentum	0.9 [39,41,42]
Learning rate decay type	cos [39]
GPU memory	24 GB
Image size	256×256
Loss function	BCE + Dice [39–42]
Data augmentation	Random horizontal–vertical flipping, random cropping, and random color mapping [39,41]

3.3. Evaluation Metrics

To objectively assess the performance of GGMNet, we applied four common semantic segmentation evaluation metrics, precision (P), recall (R), F1 score (F1), and intersection over union value (IOU), based on previous studies [32,33,42].

$$R = \frac{TP}{TP + FP} \tag{15}$$

$$P = \frac{TP}{TP + FN} \tag{16}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{17}$$

$$IOU = \frac{TP}{TP + FP + FN} \tag{18}$$

4. Experimental Results

To evaluate the effectiveness of GGMNet, we combined the DeepCrack, CrackTree260, and Aerial Track Detection datasets to conduct comparison experiments. Four evaluation metrics were employed to conduct precise quantitative analysis. Additionally, visualization of the results was conducted to qualitatively analyze the detection performance of both GGMNet and other mainstream networks.

4.1. Results for DeepCrack

Table 2 shows the quantitative crack detection results of GGMNet. To evaluate the accuracy and performance of GGMNet, we utilized several mainstream models. The results indicate that GGMNet exhibited outstanding performance compared to current mainstream networks for the DeepCrack dataset, with a precision of 83.63%, a recall of 90.93%, an F1

score of 87.13%, and an IOU value of 77.19%. Except for a slightly lower recall, all other metrics were at their optimal values.

Table 2. Comparison of the methods' P, R, F1, and IOU for DeepCrack/%.

Method	Code (accessed on 16 May 2024)	P	R	F1	IOU
HED [15]	https://github.com/s9xie/hed	78.78	88.12	83.19	71.21
RCF [16]	https://github.com/yun-liu/RCF	79.36	89.14	83.97	72.37
DeepCrack [24]	https://github.com/qinnzou/DeepCrack	79.63	87.92	83.57	71.77
U-Net [18]	https://github.com/milesial/Pytorch-UNet	79.15	90.29	84.35	72.94
SegNet [19]	https://github.com/vinceecws/SegNet_PyTorch	79.43	88.31	83.63	71.88
PSPNet [20]	https://github.com/hszhao/PSPNet	69.50	82.87	75.60	60.77
Deeplabv3+ [21]	https://github.com/VainF/DeepLabV3Plus-Pytorch	75.80	91.21	82.79	70.64
TransUNet [22]	https://github.com/Beckschen/TransUNet	78.04	91.00	84.02	72.45
SegFormer [43]	https://github.com/NVlabs/SegFormer	73.58	86.11	79.35	65.78
DMFNet [26]	https://github.com/Bsl1/DMFNet	76.71	90.56	83.06	71.03
CrackFormer [25]	https://github.com/LouisNUST/CrackFormer-II	81.15	**91.81**	86.15	75.68
GGMNet	https://github.com/hzlsdxx/GGMNet	**83.63**	90.93	**87.13**	**77.19**

Figure 6 shows the visualization of the qualitative results of each model. GGMNet obtained the superior visual performance compared with the other models. As we can see in the top row in Figure 6, all models could acquire acceptable detection results when the interference of the background was weak and the structure of the crack was simple. However, in addition to the GGMNet, HED, and DeepCrack, the other networks all omitted some details of cracks (see the red rectangular box). As seen in the second and third rows, only the GGMNet detected the crack information completely and without the problem of misdetection, but there is still a considerable number of results with missed crack detection (see the red rectangular box) and background error recognition (see the yellow rectangular box) for the other models. This is mainly because the proposed GGMNet uses the GC-Resblock in the encoder to suppress the background noises and highlight the crack feature expression, reduces the probability of misdetection of background information, and obtains coherent crack information. As observed from the fourth and fifth rows, only GGMNet had a satisfactory crack-detection performance when the structure of cracks was complex and the background noises interference was strong. Taking a closer look at the red rectangular box in fifth row, we can see that GGMNet alone extracted the cracks unlabeled, which is because GGMNet applies both the GPPM and MFF, enabling awareness of global semantic information and the spatial relationships of the microcrack. In conclusion, GGMNet obtained outstanding experimental results for different scenes and cracks of diverse scales and could effectively distinguish between the crack and background even though the interference of the background was strong.

4.2. Results for CrackTree260

To further validate the validity and generalizability of the GGMNet, we also performed experimental investigations using the CrackTree260 dataset. This dataset contains all crack types, and the structure of the cracks is more complex compared with the DeepCrack dataset. Specifically, this dataset incorporates some thin cracks. Table 3 exhibits the quantitative comparison results of each model. The precision, recall, F1 score, and IOU value of GGMNet were the highest, and these metrics of GGMNet significantly exceeded those of the others.
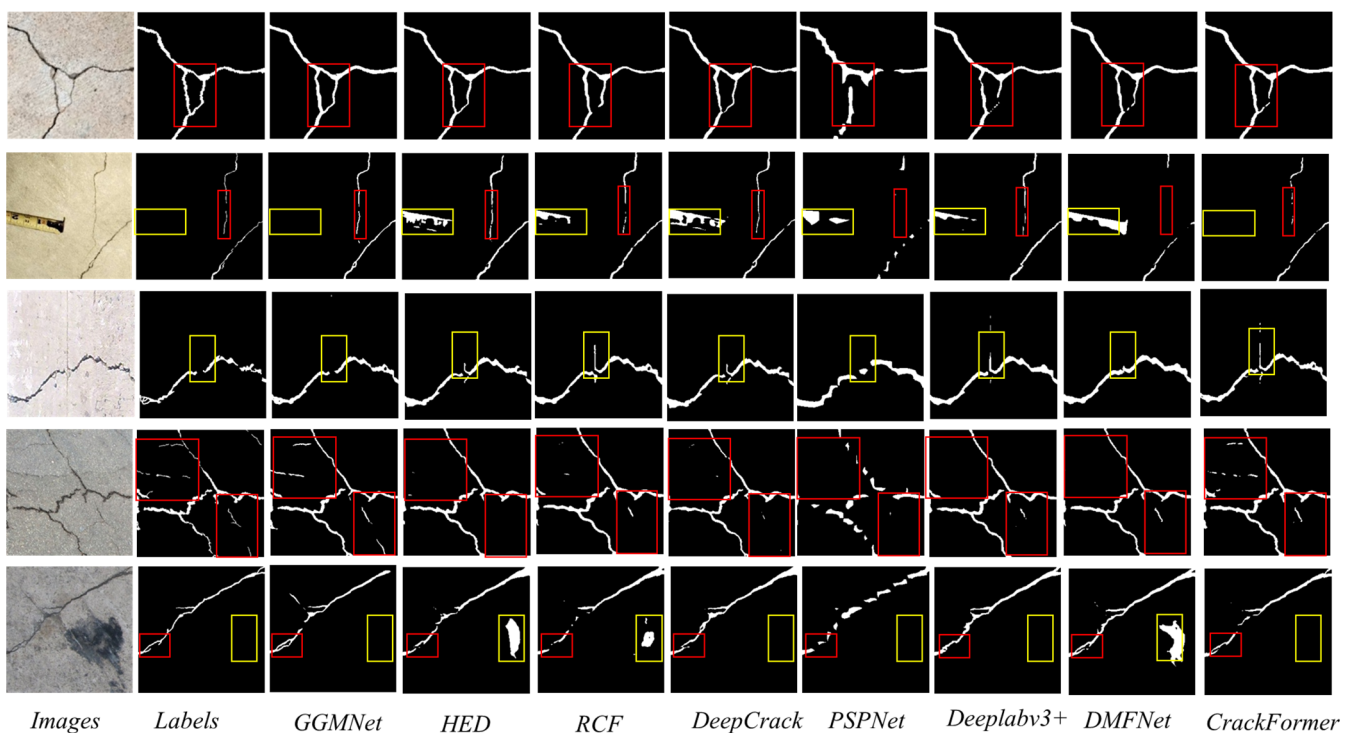
Figure 6. Visualization of the outcomes produced by diverse methods for DeepCrack.

Table 3. Comparison of the methods' P, R, F1, and IOU for the CrackTree260 dataset/%.

Method	P	R	F1	IOU
HED	74.10	73.85	73.97	58.70
RCF	73.72	72.45	73.08	57.58
DeepCrack	80.28	76.44	78.31	64.42
U-Net	85.21	81.68	83.41	71.54
SegNet	80.41	75.58	77.92	63.83
PSPNet	18.15	20.23	19.13	10.58
Deeplabv3+	40.81	72.07	52.11	40.81
GGMNet	**88.48**	**85.08**	**86.75**	**76.59**

Figure 7 shows the visualization of the qualitative results of each model. The outcomes of the proposed GGMNet were more accurate and complete. As we can see from the first and second rows in Figure 7, the other models except for GGMNet all missed detections when the structure of cracks was intricate (see the red rectangular box). Compared with that of additional labels, the visual performance of GGMNet was remarkably improved. This is because GGMNet with the GPPM and MFF can effectively aggregate multi-scale crack features and successfully capture the relationships of cracks across different regions and shapes. The third row shows that GGMNet displayed exceptional extraction abilities in microcracks and overcame the interference of background noises. In contrast, the other models all exhibited error in detection and omissions in extraction, and their visual performance was markedly inferior to that of GGMNet. In addition, we can find from the fourth and fifth rows that the GGMNet demonstrated robust performance even under uneven lighting conditions. This can be attributed to the fact that the GC-Resblock of the GGMNet guide network concentrates on the cracks and suppresses the other noises. In summary, the GGMNet showed excellent detection performance for the cracks of various sizes and shapes.
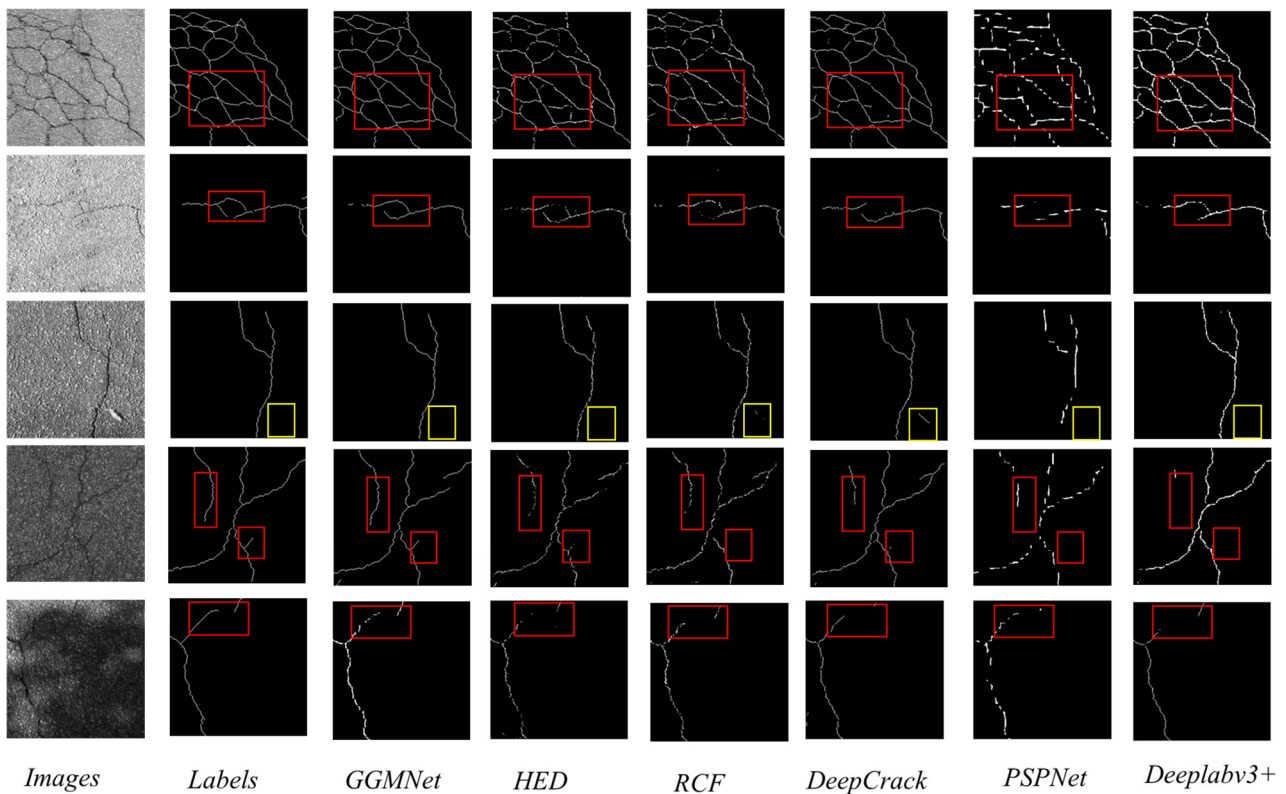
Figure 7. Visualization of the outcomes produced by diverse methods for CrackTree260.

4.3. Results for Aerial Track Detection

In contrast to the aforementioned two datasets, this dataset was acquired from the aerial viewpoint of drones. Due to the origin of the images from post-earthquake highway crack formations, where the cracks exhibit significant severity and the scene is relatively homogeneous, the crack information can be easily identified and extracted. The quantitative experimental findings of diverse segmentation models on the Aerial Track Detection dataset are shown in Table 4. As depicted in Table 4, the proposed GGMNet achieved 94.13% precision, 91.37% recall, 92.73% F1 score, and 86.45% IOU values. Compared to the other models, the GGMNet exhibited superior performance in the all evaluation metrics except for the recall. For recall, the GGMNet achieved suboptimal results, but this was only 0.01% lower than the result for DeepCrack.

Table 4. Comparison of the methods' P, R, F1, and IOU for the Aerial Track Detection dataset/%.

Method	P	R	F1	IOU
HED	86.34	85.88	86.11	75.61
RCF	91.44	88.67	90.03	81.87
DeepCrack	93.46	**91.38**	92.41	85.89
U-Net	91.23	89.10	90.15	82.07
SegNet	93.62	90.74	92.16	85.46
PSPNet	84.26	87.50	85.85	75.20
Deeplabv3+	89.16	87.19	88.16	78.83
GGMNet	**94.13**	91.37	**92.73**	**86.45**

Figure 8 qualitatively displays the visualization results of each model. All models achieved satisfactory performance, but except for GGMNet, the models still had some misdetection (the red rectangular box in the first row) and omissions in extraction (the red rectangular box in other rows). As depicted in Figure 8, it is evident that the crack-detection results of GGMNet were remarkably similar to the labels.
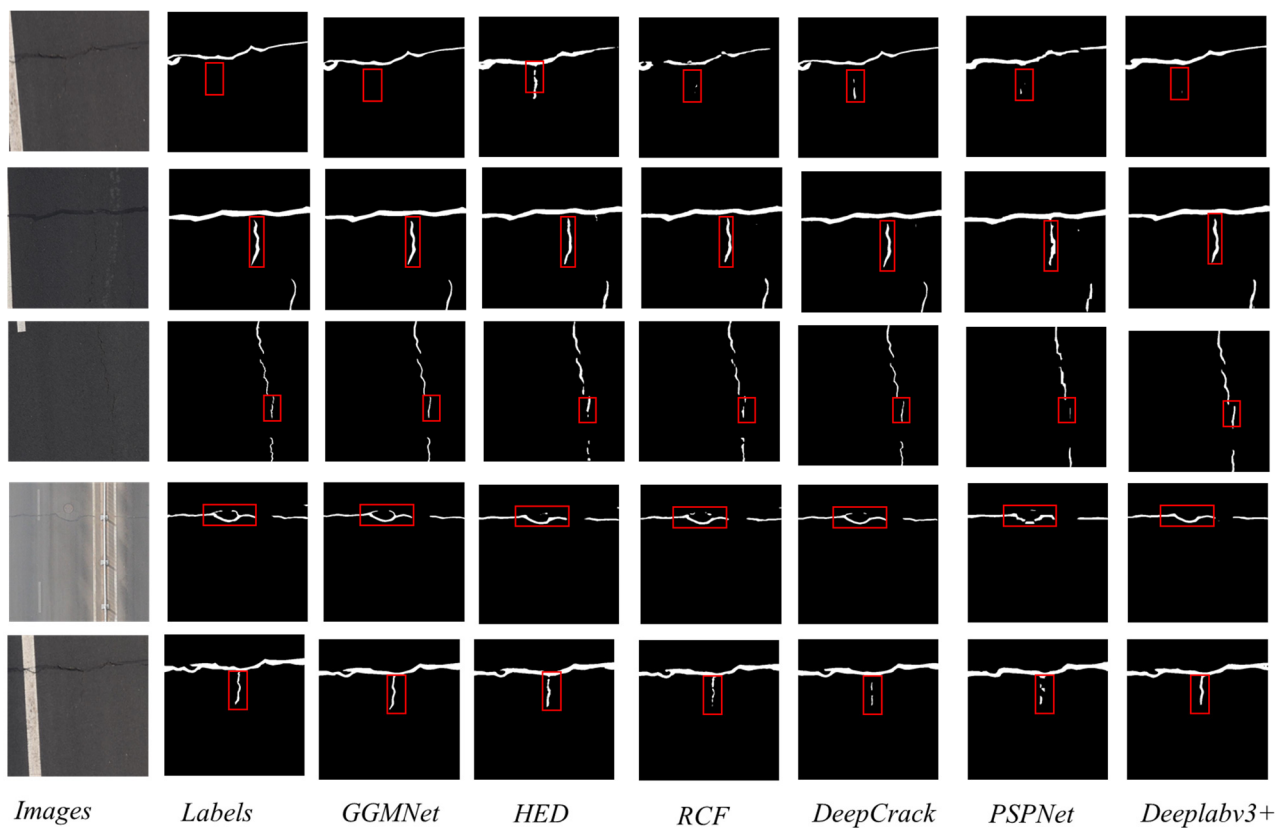
Images   *Labels*   *GGMNet*   *HED*   *RCF*   *DeepCrack*   *PSPNet*   *Deeplabv3+*

Figure 8. Visualization of the outcomes produced by diverse methods for Aerial Track Detection.

4.4. Experimental Conclusions

The results of the experiments on three publicly available datasets indicate that GGM-Net achieved the best performance. The results obtained on the DeepCrack dataset provide evidence that GGMNet exhibits excellent performance even in the presence of complex background information. The experiments on CrackTree260 demonstrate that GGMNet is effective in extracting cracks of various shapes and sizes, particularly thin cracks. The results for the Aerial Track Detection prove that the proposed GGMNet is capable of adapting to different perspectives and fields of view for crack detection. In conclusion, GGMNet is characterized by outstanding performance and strong robustness.

5. Discussions

5.1. Comparison of Effectiveness among Different Levels of GC-Resblock

To showcase the effectiveness of varying tiers of GC-Resblock, we conducted additional investigations into the functionality of this module using ablation experiments and feature visualization techniques on the DeepCrack dataset.

Table 5 displays the assessment findings for different levels of GC-Block for the DeepCrack testing sets. Compared with the No. 1 model, No. 5 acquired the optimal results, with the F1 score and IOU value improving by 1.43% and 2.22%, respectively. As we can see, the chosen evaluation metrics consistently exhibited incremental improvements from the No. 1 to the No. 5 model, indicating the effectiveness of incorporating GC-Resblock across different stages. Meanwhile, it can be observed that more significant improvements were obtained by incorporating this module in stage 1 and stage 4, resulting in increases of 0.57% and 0.81% for IOU value, respectively.

To further investigate the role of GC-Resblock, the feature maps were visualized before and after the GC-Resblock was applied at different levels. Figure 9 shows the visualized results, where different brightness levels indicate the model's attention to different regions. As shown in Figure 9, the feature maps all exhibited varying levels of luminosity changes

before and after the addition of GC-Resblock. To be more specific, a-b, c-d, and e-f show the brightness of cracks increased while the brightness of the background regions decreased. The top row in Figure 9 shows that the semantic information of cracks in f was more extensive compared to that in a. In conclusion, the evaluation metrics and visualizations all demonstrate that the GC-block guides the network to emphasize the identities of cracks while attenuating background interference.

Table 5. Assessment findings for different levels of GC-Resblock for DeepCrack/%.

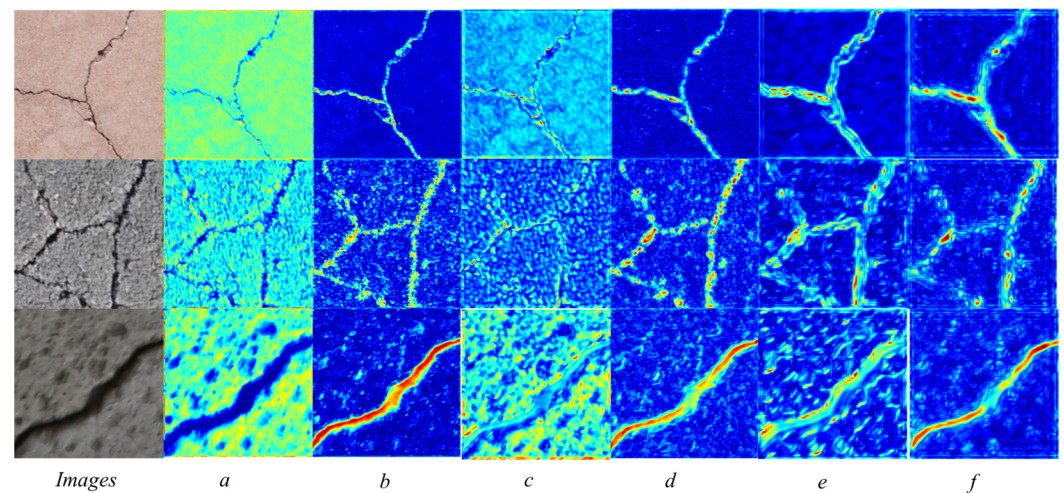| No.	Stage1	Stage2	Stage3	Stage4	F1	IOU
1					85.70	74.97
2	✓				86.07	75.54
3	✓	✓			86.33	75.95
4	✓	✓	✓		86.61	76.38
5	✓	✓	✓	✓	87.13	77.19

Images   a   b   c   d   e   f

Figure 9. Visualization outcomes for different levels of GC-Resblock. (**a**,**b**) Before and after addition of the first layer of the encoder. (**c**,**d**) Before and after the addition of the second layer of the encoder. (**e**,**f**) Before and after the addition of the third layer of the encoder.

5.2. Comparison of the Effectiveness among Different Multi-Scale Aggregation Schemes

Multi-scale aggregation approaches have been widely applied in tasks such as object detection and semantic segmentation, with related studies confirming their effectiveness [44–48]. In this study, the proposed GPPM enabled our network to perceive distant multi-scale crack features, resulting in a better representation of complex multidimensional crack features. To further demonstrate the superiority of this module, we compared it with other mainstream multi-scale aggregation approaches, and the comparative results are presented in Table 6, which indicate our module is more suitable for extracting complex crack features.

Table 6. F1 score and IOU value of the different multi-scale aggregation schemes for DeepCrack/%.

No.	Method	F1	IOU
1	ASPP [49]	86.09	75.58
2	DRB [46]	86.31	75.93
3	MFEM [44]	85.44	74.58
4	DCI [46]	86.90	76.82
5	GPPM	87.13	77.19

5.3. Comparison of Effectiveness among Various Feature Fusion Methods

To affirm the superiority of the CWF, we conducted a comprehensive comparison of various feature fusion methods on the DeepCrack testing set. As shown in Figure 10, the proposed CWF obtained the optimal F1 score and IOU value compared to the other methods. This is because features at different layers contain complementary and redundant information. If there are only output features from a single dimension, the results are often incomplete. If these features are concatenated without processing from different layers, this will lead to feature redundancy, and satisfactory results will not be acquired. Therefore, considering the contributions of features from different layers is of paramount importance. We devised a channel-weighting fusion module (CWF) that adaptively captures the weights of each channel, facilitating the propagation of informative features. The CWF proposed is more adaptive to crack detection in contrast to the SE module [50].
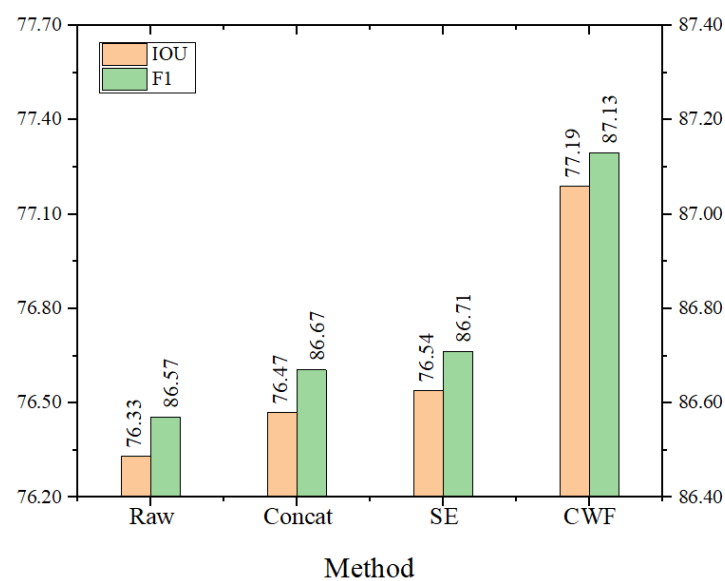

Figure 10. F1 score and IOU value of the different feature fusion methods for DeepCrack.

5.4. Ablation Experiments

To showcase the effectiveness of each component we proposed, the effects of removing GC-Resblock, GPPM, and MFF on the model performance are discussed, respectively. The findings from the performed ablation experiments on the DeepCrack dataset are displayed in Table 7.

Table 7. Ablation experiments for DeepCrack/%.

No.	Method	F1	IOU
1	Baseline	84.35	72.94
2	*w/o* GC-Resblock	85.70	74.97
3	*w/o* GPPM	86.64	76.42
4	*w/o* GRB(GPPM)	86.86	76.78
5	*w/o* MFF	86.57	76.33
6	GGMNet	87.13	77.19

The F1 score and IOU value of the model decreased significantly after removal of the GC-Resblock, which indicates that focusing on essential information and suppressing background noise is of paramount importance. The two metrics of GGMNet decreased to some extent after removal of the GPPM and MFF, respectively, which indicates that the aggregation and interaction of multi-scale information are also of great importance. In addition, for the GPPM module, discarding the graph reasoning unit negatively impacted

the model's performance, which indicates that capturing the relationships among different regions and extracting irregular spatial information is crucial for crack detection.

6. Conclusions

This paper introduces a novel pavement-crack detection network named GGMNet. Combined with three crack datasets, the experimental findings from quantitative assessment and qualitative analysis demonstrate that GGMNet exhibits excellent performance and strong robustness. This method will facilitate accurate and comprehensive pavement-crack detection, providing significant engineering significance for digital highway management and maintenance. Below are the specific contributions of this paper:

(1) An accurate and robust network, named GGMNet, is proposed for pavement-crack detection.

(2) A GC-Resblock was developed to guide the network to emphasize the identities of cracks while suppressing the background noises effectively.

(3) A GPPM was constructed to support the model to aggregate multi-scale features and capture the long-range dependencies of cracks.

(4) A MFF structure was designed to facilitate channel interaction and achieve feature complementarity across different layers.

Although the proposed GGMNet shows optimal detection performance, it has certain limitations. The model's parameter and computational complexity are slightly higher. Consequently, our future focus will be on simultaneously improving the model's accuracy and addressing speed considerations.

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

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Authors Juncheng Zeng and Xiaowei Xu are employed by the Fujian Expressway Science & Technology Innovation Research Institute Co., Ltd.; Authors Luyang Qiu and Qunxiong Zhuo are employed by the Fujian Luoning Expressway Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ragnoli, A.; De Blasiis, M.; Di Benedetto, A. Pavement Distress Detection Methods: A Review. *Infrastructures* **2018**, *3*, 58. [CrossRef]
2. Huang, W.; Zhang, N. A Novel Road Crack Detection and Identification Method Using Digital Image Processing Techniques. In Proceedings of the 2012 7th International Conference on Computing and Convergence Technology (ICCCT), Seoul, Republic of Korea, 3–5 December 2012; pp. 397–400.
3. Li, P.; Wang, C.; Li, S.; Feng, B. 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; Li, J., Ed.; IEEE: Piscataway, NJ, USA, 2015; pp. 1716–1720.
4. Xu, W.; Tang, Z.; Zhou, J.; Ding, J. Ieee Pavement Crack Detection Based On Saliency and Statistical Features. In Proceedings of the 2013 IEEE International Conference on Image Processing, Melbourne, Australia, 15–18 September 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 4093–4097.
5. Chambon, S.; Moliard, J. Automatic Road Pavement Assessment with Image Processing: Review and Comparison. *Int. J. Geophys.* **2011**, *2011*, 989354. [CrossRef]

6. Tanaka, N.; Uematsu, K. A Crack Detection Method in Road Surface Images Using Morphology. In Proceedings of the IAPR Workshop on Machine Vision Applications, Chiba, Japan, 17–19 November 1998; pp. 154–157.

7. Zhou, J.; Huang, P.; Chiang, F. Wavelet-Based Pavement Distress Detection and Evaluation. *Opt. Eng.* **2006**, *45*, 027007. [CrossRef]

8. Subirats, P.; Dumoulin, J.; Legeay, V.; Barba, D. IEEE Automation of Pavement Surface Crack Detection Using the Continuous Wavelet Transform. In Proceedings of the 2006 International Conference on Image Processing, Las Vegas, NV, USA, 26–29 June 2006; IEEE: Piscataway, NJ, USA, 2006; p. 3037.

9. Kapela, R.; Sniatala, P.; Turkot, A.; Rybarczyk, A.; Pozarycki, A.; Rydzewski, P.; Wyczalek, M.; Bloch, A. IEEE Asphalt Surfaced Pavement Cracks Detection Based on Histograms of Oriented Gradients. In Proceedings of the 2015 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES), Torun, Poland, 25–27 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 579–584.

10. Hu, Y.; Zhao, C. A Novel LBP Based Methods for Pavement Crack Detection. *JPRR* **2010**, *5*, 140–147. [CrossRef] [PubMed]

11. Yang, L.; Fan, J.; Huo, B.; Liu, Y. Inspection of Welding Defect Based on Multi-Feature Fusion and a Convolutional Network. *J. Nondestruct. Eval.* **2021**, *40*, 90. [CrossRef]

12. Yang, L.; Fan, J.; Liu, Y.; Li, E.; Peng, J.; Liang, Z. Automatic Detection and Location of Weld Beads With Deep Convolutional Neural Networks. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]

13. Liu, Y.; Shen, J.; Yang, L.; Bian, G.; Yu, H. ResDO-UNet: A Deep Residual Network for Accurate Retinal Vessel Segmentation from Fundus Images. *Biomed. Signal Process. Control* **2023**, *79*, 104087. [CrossRef]

14. Li, J.; Gao, G.; Yang, L.; Liu, Y. GDF-Net: A Multi-Task Symmetrical Network for Retinal Vessel Segmentation. *Biomed. Signal Process. Control* **2023**, *81*, 104426. [CrossRef]

15. Xie, S.; Tu, Z. Holistically-Nested Edge Detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1395–1403.

16. Liu, Y.; Cheng, M.-M.; Hu, X.; Wang, K.; Bai, X. Richer Convolutional Features for Edge Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3000–3009.

17. 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; IEEE: Piscataway, NJ, USA, 2015; pp. 3431–3440.

18. 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, Munich, Germany, 5–9 October 2015; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241.

19. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]

20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. IEEE Pyramid Scene Parsing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.

21. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; Volume 11211, pp. 833–851.

22. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306.

23. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*; Springer Nature: Cham, Switzerland, 2021.

24. Liu, Y.; Yao, J.; Lu, X.; Xie, R.; Li, L. DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation. *Neurocomputing* **2019**, *338*, 139–153. [CrossRef]

25. Liu, H.; Miao, X.; Mertz, C.; Xu, C.; Kong, H. CrackFormer: Transformer Network for Fine-Grained Crack Detection. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, BC, Canada, 11–17 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 3763–3772.

26. Bai, S.; Yang, L.; Liu, Y.; Yu, H. DMF-Net: A Dual-Encoding Multi-Scale Fusion Network for Pavement Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2023**, 1–16, *early access*. [CrossRef]

27. Zhang, Y.; Liu, C. Network for Robust and High-Accuracy Pavement Crack Segmentation. *Autom. Constr.* **2024**, *162*, 105375. [CrossRef]

28. 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; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.

29. Cao, Y.; Xu, J.; Lin, S.; Wei, F.; Hu, H. GCNet: Non-Local Networks Meet Squeeze-Excitation Networks and Beyond. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seoul, Republic of Korea, 27–28 October 2019; IEEE: Seoul, Republic of Korea, 2019; pp. 1971–1980.

30. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2117–2125.

31. Chen, Y.; Rohrbach, M.; Yan, Z.; Shuicheng, Y.; Feng, J.; Kalantidis, Y. Graph-Based Global Reasoning Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; IEEE: Long Beach, CA, USA, 2019; pp. 433–442.

32. Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1525–1535. [CrossRef]

33. Zhou, Y.; Chen, Z.; Wang, B.; Li, S.; Liu, H.; Xu, D.; Ma, C. BOMSC-Net: Boundary Optimization and Multi-Scale Context Awareness Based Building Extraction From High-Resolution Remote Sensing Imagery. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]

34. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.

35. Li, X.; Sun, X.; Meng, Y.; Liang, J.; Wu, F.; Li, J. Dice Loss for Data-imbalanced NLP Tasks. *arXiv* **2019**, arXiv:1911.02855.

36. Lee, C.-Y.; Xie, S.; Gallagher, P.; Zhang, Z.; Tu, Z. Deeply-Supervised Nets. In Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, PMLR, San Diego, CA, USA, 9–12 May 2015; pp. 562–570.

37. Zou, Q.; Cao, Y.; Li, Q.; Mao, Q.; Wang, S. Crack Tree: Automatic Crack Detection from Pavement Images. *Pattern Recognit. Lett.* **2012**, *33*, 227–238. [CrossRef]

38. Hong, Z.; Yang, F.; Pan, H.; Zhou, R.; Zhang, Y.; Han, Y.; Wang, J.; Yang, S.; Chen, P.; Tong, X.; et al. Highway Crack Segmentation From Unmanned Aerial Vehicle Images Using Deep Learning. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]

39. Sun, X.; Xie, Y.; Jiang, L.; Cao, Y.; Liu, B. DMA-Net: DeepLab With Multi-Scale Attention for Pavement Crack Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18392–18403. [CrossRef]

40. Guo, J.-M.; Markoni, H.; Lee, J.-D. BARNet: Boundary Aware Refinement Network for Crack Detection. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 7343–7358. [CrossRef]

41. Wang, Y.; Zeng, X.; Liao, X.; Zhuang, D. B-FGC-Net: A Building Extraction Network from High Resolution Remote Sensing Imagery. *Remote Sens.* **2022**, *14*, 269. [CrossRef]

42. Qu, Z.; Chen, W.; Wang, S.; Yi, T.; Liu, L. A Crack Detection Algorithm for Concrete Pavement Based on Attention Mechanism and Multi-Features Fusion. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 11710–11719. [CrossRef]

43. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.

44. Ullah, I.; Jian, M.; Hussain, S.; Lian, L.; Ali, Z.; Qureshi, I.; Guo, J.; Yin, Y. Global Context-Aware Multi-Scale Features Aggregative Network for Salient Object Detection. *Neurocomputing* **2021**, *455*, 139–153. [CrossRef]

45. Ma, C.; Zhuo, L.; Li, J.; Zhang, Y.; Zhang, J. Occluded Prohibited Object Detection in X-Ray Images with Global Context-Aware Multi-Scale Feature Aggregation. *Neurocomputing* **2023**, *519*, 1–16. [CrossRef]

46. Alam, M.S.; Wang, D.; Liao, Q.; Sowmya, A. A Multi-Scale Context Aware Attention Model for Medical Image Segmentation. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 3731–3739. [CrossRef] [PubMed]

47. Wang, B.; Ji, R.; Zhang, L.; Wu, Y. Bridging Multi-Scale Context-Aware Representation for Object Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 2317–2329. [CrossRef]

48. Niu, P.; Gu, J.; Zhang, Y.; Zhang, P.; Cai, T.; Xu, W.; Han, J. MDCGA-Net: Multi-Scale Direction Context-Aware Network with Global Attention for Building Extraction from Remote Sensing Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, 1–16. [CrossRef]

49. 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]

50. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; IEEE: Long Beach, CA, USA, 2019.