applied sciences

MDPI

Article

Automated Multi-Type Pavement Distress Segmentation and Quantification Using Transformer Networks for Pavement Condition Index Prediction

Zaiyan Zhang [1,2], Weidong Song [1,*], Yangyang Zhuang [2], Bing Zhang [1] and Jiachen Wu [1]

[1] School of Geomatics, Liaoning Technical University, Fuxin 123000, China; 2014800205@usth.edu.cn (Z.Z.); zhangbing@lntu.edu.cn (B.Z.); 13584931945@163.com (J.W.)
[2] College of Mining Engineering, Heilongjiang University of Science and Technology, Harbin 150022, China; 18103603386@163.com
* Correspondence: songweidong@lntu.edu.cn; Tel.: +86-1894-600-7768

Abstract: Pavement distress detection is a crucial task when assessing pavement performance conditions. Here, a novel deep-learning method based on a transformer network, referred to as ISTD-DisNet, is proposed for multi-type pavement distress semantic segmentation. In this methodology, a mix transformer (MiT) based on a hierarchical transformer structure is chosen as the backbone to obtain multi-scale feature information on pavement distress, and a mixed attention module (MAM) is introduced at the decoding stage to capture the pavement distress features across different channels and spatial locations. A learnable transposed convolution upsampling module (TCUM) enhances the model's ability to restore multi-scale distress details. Subsequently, a novel parameter—the distress pixel density ratio (*PDR*)—is introduced based on the segmentation results. Analyzing the intrinsic correlation between the *PDR* and the pavement condition index (PCI), a new pavement damage index prediction model is proposed. Finally, the experimental results reveal that the F1 and mIOU of the proposed method are 95.51% and 91.67%, respectively, and the segmentation performance is better than that of the other seven mainstream segmentation models. Further PCI prediction model validation experimental results also indicate that utilizing the *PDR* enables the quantitative evaluation of the pavement damage conditions for each assessment unit, holding promising engineering application potential.

Keywords: pavement distress segmentation; deep learning; transformer; ISTD-DisNet; pavement condition prediction

Citation: Zhang, Z.; Song, W.; Zhuang, Y.; Zhang, B.; Wu, J. Automated Multi-Type Pavement Distress Segmentation and Quantification Using Transformer Networks for Pavement Condition Index Prediction. *Appl. Sci.* **2024**, *14*, 4709. https://doi.org/10.3390/app14114709

Academic Editor: Adriana Brancaccio

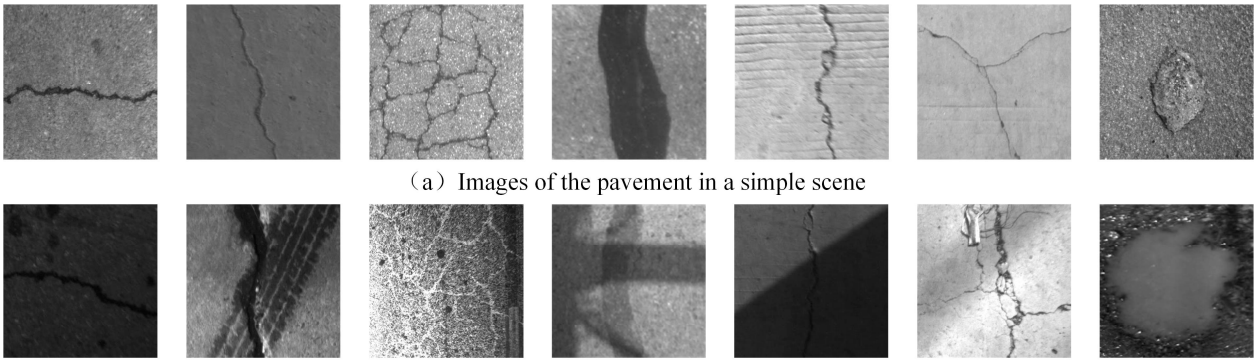
Received: 26 March 2024
Revised: 19 May 2024
Accepted: 27 May 2024
Published: 30 May 2024

1. Introduction

Pavement distress detection and damage condition evaluation have been hot issues in road maintenance decision-making [1,2]. Generally, road maintenance departments use mobile acquisition equipment to obtain pavement images, conduct manual visual inspection, and finally, obtain the pavement condition index (PCI) of the line evaluation unit [3]. This process has a long detection period, high cost, and strong subjectivity, which makes it difficult to meet the demand for the rapid and large-scale automatic detection and evaluation of pavement technical conditions [4]. The primary goal of this study is to design a pavement distress segmentation model for multiple types of pavement distress in complex scenarios based on the latest deep-learning techniques and quantify the pavement distress characteristics for the automated evaluation of pavement conditions.

As pavement cracks, potholes, and repairs in images are usually shown as linear and planar structures with shape variations, pavement damage detection can be regarded as a linear and planar detection task in computer vision [5–9]. Most of the early studies focused on crack detection algorithms based on digital image-processing techniques

and machine learning. The former are mainly represented by methods such as threshold segmentation [10], edge detection [11], morphological operation [12], minimal path selection [13], and region growth [14]. The latter, which can be predicted by learning the intrinsic knowledge of pavement crack data, are mainly represented by support vector machine (SVM) [15–17] and shallow artificial neural network (ANN) [18] models. Ideally (Figure 1a), if a single scene is captured and the distress has good continuity and high contrast, then the early methods based on low-level features can achieve a high accuracy in pavement distress detection [5]. However, during the acquisition of images of pavements, which is affected by factors such as the pavement type, lighting conditions, interferences, and stains, the above detection methods, despite being able to quickly obtain part of the information on the pavement lesions, cannot easily take into account the influence of multi-source noise with textural similarity, and the completeness, accuracy, and efficiency of the identification are not good. It is worth noting that most of the application scenarios are complex ones. Therefore, before carrying out the research described in this article, the "complex scene" (Figure 1b) in the scope of the subsequent experiments and discussions was defined as follows: (1) the scene contains different pavement types, multiple types of pavement distress, and a complex topology; (2) the scene has a pavement background with strong speckle noise that is complex and changeable, with a low target signal-to-noise ratio and poor spatial continuity of the target pixels; and (3) there may be shadows, occlusions, varied light intensity, and other factors that influence the recognition of pavement. These factors change, and there can be noise with textural similarity to pavement lesions [19]. Therefore, automatic pavement distress detection from CCD images of pavements acquired in real projects is still a challenging task.



（a）Images of the pavement in a simple scene

（b）Images of the pavement in complex Scenarios

Figure 1. The challenge of the automatic detection of multi-type pavement distress [20].

Recent theoretical research advances have shown that deep learning can autonomously acquire different levels of features and then construct high-level features from the low-level primitives to effectively solve complex fitting problems [21]. Influenced by this, deep-learning-based methods now play a dominant role in pavement damage detection [19,22]. Research on pavement damage detection based on deep convolutional neural networks (DCNNs) can be categorized into three types: image classification methods, target detection methods, and semantic segmentation methods. It is important to note that, compared to classification [23–25] and detection [26–28], pixel-level segmentation [5,6,29] can provide more accurate geometric target descriptions for a wide range of applications, such as geometric feature quantification of distress, severity classification, and quantitative assessment of pavement condition, which is the focus of this article. For example, Jenkins et al. [30] proposed a semantic segmentation algorithm for road cracks based on U-Net [31], but the model generalization was insufficient due to the lack of data (80 training images and 20 validation images). Subsequently, researchers have made a series of improvements based on U-Net [32,33] and achieved improved segmentation performances in crack segmentation test sets such as CRACK500 [34], CFD [35], and AigleRN [36]. In addition, to

address the problem of the small size of the existing crack segmentation datasets, a crack segmentation network known as ConnCrack, which integrates a generative adversarial network (GAN) and connectivity graphs, was proposed by Mei and Mustafa [37]. However, pavements do not have only one type of distress, and each type of distress has its own exclusive characteristics and apparent forms. For this reason, Lõuk et al. [38] applied a U-Net-like network architecture with different levels of contextual resolution to integrate more contextual information. The authors considered different types of pavement distress, but they only extracted the planar regions of the different types of distress. In addition, Zhang et al. [8] collected pavement images of urban roads in Montreal in Canada and produced a semantic segmentation dataset for potholes, patches, lanes, linear cracks, and mesh cracks. They also proposed and evaluated a method to automatically detect and classify the types of pavement distress using a convolutional neural network (CNN) and low-cost video data, where the model's detection and classification accuracies both reached 83.8%. It is important to note that, affected by the difficulty of dataset collection and the cost of labeling, until recently, there have been very few studies on the semantic segmentation of multiple types of pavement lesions in complex scenarios.

Although the above DCNN-based algorithms have achieved good distress segmentation performances, CNNs still do not perform well in extracting long-range contextual information, which is crucial for the model to understand distress with complex topological features in pavement images [39]. In addition, since the convolutional kernel size of the CNN (3×3) determines that the model can only understand input imagery with localized information, the small receptive field leads to discontinuous and false-positive predictions [40]. Fortunately, since the first successful application of the vision transformer (ViT) [41] in computer vision tasks, there has been an explosion of dense prediction research based on ViT models, which tokenize the input image and then utilize a self-attention mechanism to enable the model to comprehend the image with a global perspective [42,43]. However, these ViT-based improvements tend to significantly increase the computational cost. In addition, since the resolution of the positional encoding in the ViT model is fixed, it causes a degradation in the performance of dense prediction when the resolution of the test image is different from the resolution of the training image. For this reason, Xie et al. [44] proposed a hierarchical transformer encoder, replacing the positional encoding with a 3×3 convolution operation, which greatly improves the robustness of the model segmentation. However, the use of only a few simple multilayer perceptrons (MLPs) for decoding is likely to have reduced the model's ability to restore detailed information.

Quantitative assessment of pavement technical conditions is a prerequisite for the rational development of maintenance programs. After completing the pavement distress detection, the calculation results eventually need to be transformed into the detailed distress types and the geometric parameters of the distress [45], and then the PCI can be calculated [46]. It should be noted that, compared with research on the semantic segmentation of pavement distress, there have been relatively few quantitative studies on the segmentation results. Although a few crack quantization methods have realized the calculation of the width, length, and other eigenvalues of the cracks, the computation is time-consuming, and the results of the computation are affected by the impact of the complex scenario. For example, Hu et al. [4] performed crack classification and width calculation based on the segmentation results but failed to realize quantitative evaluation of the pavement condition. In addition, no research on pavement damage conditions based on quantitative features has been seen to date [47]. The main indices currently used to characterize the pavement condition include the pavement damage condition index (PCI), the existing pavement functional index (the present serviceability index, PSI), the present serviceability rating (PSR), and the pavement rating (pavement surface evaluation and rating (PASER)) [48]. Among these indices, the PCI is a quantitative pavement condition-based index, and all the other indicators can be categorized as qualitative pavement indices. Moreover, the PCI was developed by the U.S. Army Corps of Engineers, based on a large amount of measured inspection data and visual observations, and it is currently the most commonly used pave-

ment condition assessment index in the systems of highway pavement rehabilitation and maintenance management [49]. Although some research teams have proposed the use of neural network algorithms [50], decision tree algorithms [51], and integrated ANN and genetic programming (GP) algorithms [52] to predict pavement conditions, the limitation of these evaluation models is that they all rely on manual inspection of the damage. The inspector needs to calculate the area, length, and severity of the damage, and the outputs can then be used as inputs to these models to calculate the PCI. Therefore, there is a need for a fully automated pavement condition evaluation model, based on the output results of the pavement distress inspection models.

To address the abovementioned issues of multi-type pavement distress segmentation and automated prediction of pavement conditions in complex scenarios, based on the ISTD-PDS7 [20] proposed in our previous research, we first proposed a network based on a hierarchical transformer structure (named ISTD-DisNet) to extract the pavement distresses in an end-to-end manner. ISTD-DisNet adopts a hierarchical transformer encoder structure without positional coding in SegFormer to take into account the long-range contextual information on pavement distress at different scales. A mixed attention module (MAM) was introduced at the decoding stage to attenuate or even eliminate the interference of irrelevant information and enhance the representation of distress features in complex contexts; meanwhile, a transposed convolution upsampling module (TCUM) was constructed using transposed convolution with a learning capability, aiming to enhance the model's ability to restore details. The main contributions of this study are summarized below:

1. To enhance the model's multi-scale distress feature representation and detail recovery while modeling long-range dependencies and providing a global feature representation, we designed a multi-scale distress segmentation network—ISTD-DisNet—with an encoder–decoder transformer architecture.
2. Based on the ISTD-DisNet output, we analyzed the correlation between the *PDR* and PCI indices and developed a new pavement damage condition prediction model.
3. Extensive experiments were conducted on the ISTD-PDS7 benchmark dataset and evaluation units with different pavement conditions to verify the superiority of the ISTD-DisNet method proposed in this article in the task of multi-scale pavement distress segmentation in complex scenarios. In addition, a comprehensive comparative analysis was conducted to compare the results of the proposed automatic evaluation model for the pavement damage condition with those of manual visual discrimination.

The rest of this article is organized as follows. Section 2 outlines the network architecture of ISTD-DisNet, including the MAM and TCUM modules, and explains how they contribute to the feature enhancement and detail recovery. Section 3 describes the dataset and experimental setup and provides the comprehensive experimental results. Section 4 provides an insight into the methodology for constructing the pavement damage condition prediction model and validates the prediction model with an example. Section 5 summarizes the work.

2. Methods

2.1. Network Overview

As shown in Figure 2, the ISTD-DisNet architecture proposed in this article consists of an encoder/mix transformer (MiT) with a hierarchical transformer structure that does not require positional encoding, and a decoder that incorporates a hybrid attention mechanism and a TCUM module. ISTD-DisNet retains the coding structure of SegFormer. The size of the input image is first resized to $512 \times 512 \times 3$, and the input image then enters into the overlapping patch embedding (OPE) structure for feature extraction and downsampling. The obtained features are then input into the hierarchical transformer block module, where four hierarchical decoding operations decode the input features. After the four hierarchical decoding operations, feature maps of F1 = $128 \times 128 \times 64$, F2 = $64 \times 64 \times 128$, F3 = $32 \times 32 \times 320$, and F4 = $16 \times 16 \times 512$ are generated sequentially. In the decoding stage, the four feature maps are first upsampled to 1/4 of the input image size using the

MLP layer and then fed into the MAM module to capture the most significant semantic information on pavement lesions in different channels and different spatial locations. After performing the Concat operation, the merged feature maps then go through the TCUM module with learning capability to enhance the detail restoration ability of the feature information. Finally, a feature map with the resolution of H × W × Ncls is obtained, with Ncls = 2, where Ncls represents the number of semantic segmentation categories that contain the background. It is worth noting that the new model replaces the cross-entropy loss function in SegFormer with a mixed loss function to alleviate the adverse effects caused by the imbalance of the pixel share of the lesions in the dataset.

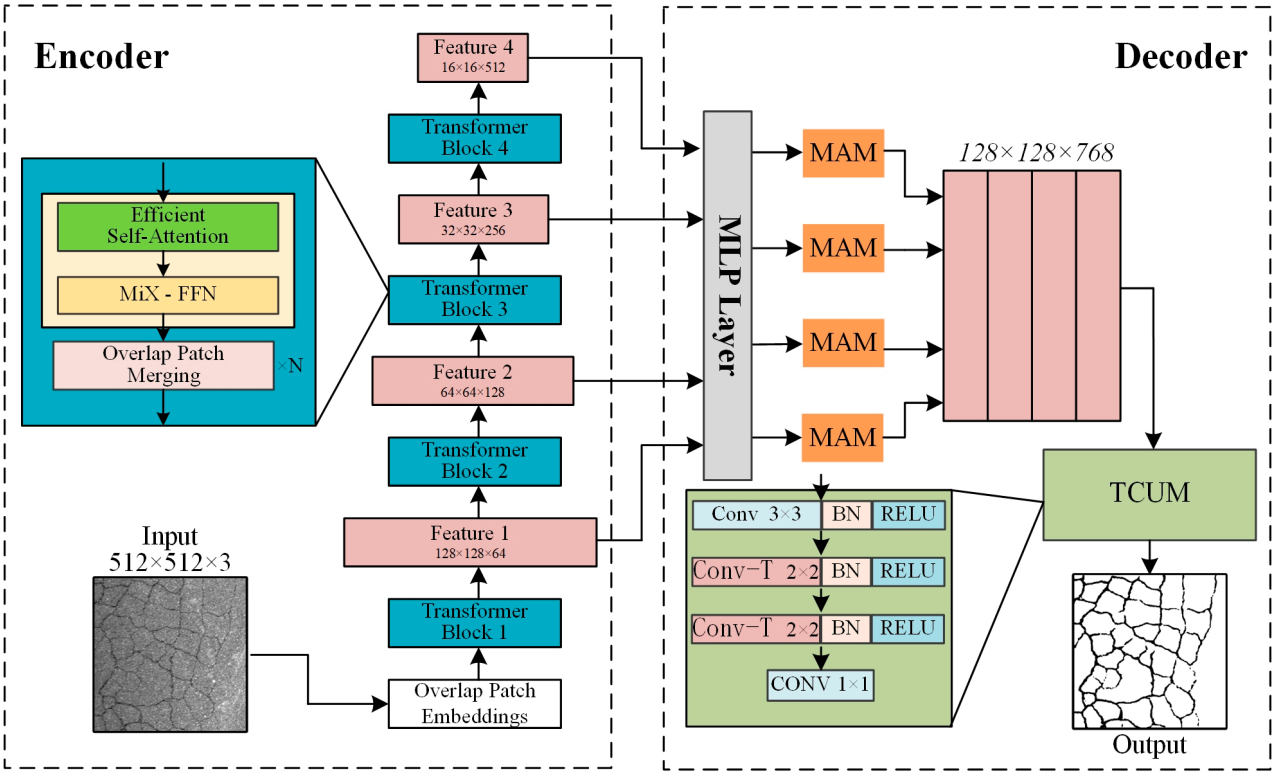


Figure 2. Structure of ISTD-DisNet.

2.2. Encoder

ISTD-DisNet retains the coding structure of SegFormer. As shown in Figure 2, the transformer module inside the encoder of SegFormer [44] uses OPE to extract features and downsample from the input image. Because the OPE module allows the slices to overlap with each other, the elements inside the slices are also connected to each other as a way to ensure the continuity and integrity of the elements. The OPE is then computed using a standard convolutional layer to spatially reshape the 2D features into 1D features. The features are then input into the efficient self-attention (ESA) layer and Mix-FFN layer for global linkage construction. To replace the positional encoding in the normal transformer, a 3 × 3 Conv is added between the two linear layers of the feed-forward network (FFN) to spatially fuse the positional information. The linear layers in the encoder are followed by layer normalization (LN), and the activation function is a Gaussian error linear unit (GELU). The transformer block uses multiple ESA and Mix-FFN layers to deepen the network, so as to extract rich details and semantic features. Slices of different sizes allow the input image to compute self-attention in the ESA layers at each scale, an operation that allows SegFormer to obtain global information more quickly and with more pure self-attention at each scale than previous networks based on CNNs that perform self-attention computation after integrating information from all the scales. ISTD-DisNet uses SegFormer MiT-B2 as the backbone, and the main hyperparameters of MiT-B2 are shown in Table 1.

Table 1. Main hyperparameters of MiT-B2.

Name	Number
Embed dims	[64, 128, 320, 512]
Num layers	[3, 4, 6, 3]
Num heads	[1, 2, 5, 8]
Patch size	[7, 3, 3, 3]
Strides	[4, 2, 2, 2]
Sr ratios	[8, 4, 2, 1]
MLP ratio	4

Although the transformer block in the encoding part of SegFormer allows each element to obtain the connection with the other elements, its overly simple decoding layer cannot accurately restore the encoded information, and simple bilinear interpolation upsampling and splicing would also lead to a large amount of lost detail information, which is crucial for the segmentation of small cracks. For this reason, in this study, we optimized the design of the decoder, which consists of: (1) using the MAM module to sequentially infer the attention weight map along two independent dimensions (channel and space), and then adding the attention map to the input feature map for adaptive feature optimization; and (2) in the last MLP layer, the upsampled portion is replaced with a transposed convolution with learning capability to enhance the model's ability to restore details.

2.3. Mixed Attention Module (MAM)

Images are fed into the network and, after encoding, a series of features with different information are generated. The features at the highest level have the strongest semantic characterization ability, while the features at the lowest level have the strongest edge information [5]. Decoding and fusing this semantic information directly results in the loss of many salient details due to the fact that the different channels and different spatial locations of the high-level features contribute differently to the computation. Inspired by Fu et al. [53], we designed the MAM module, which is divided into two parts—a channel attention mechanism and a spatial attention mechanism—to capture the pavement distress features in different channels and different spatial locations. The structure of the MAM module is shown in Figure 3.

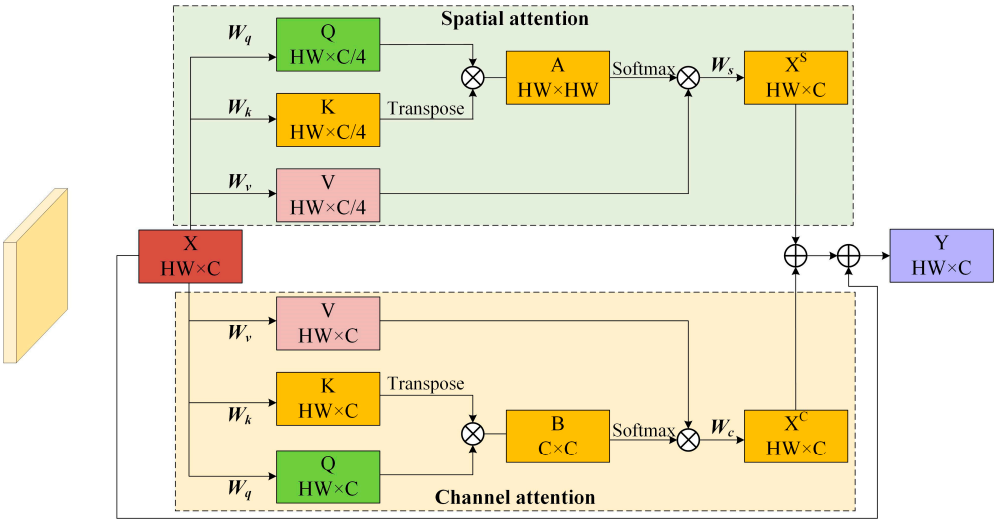


Figure 3. Mixed attention module.

2.3.1. Spatial Attention Mechanism

For the four feature maps $F_i, i \in [1, 2, 3, 4]$ extracted from MiT-B2, firstly, their width and height dimensions are expanded into one-dimensional vectors and transposed to obtain

a two-dimensional matrix $\boldsymbol{X} \in \boldsymbol{R}^{WH \times C}$. C is the number of channels of the feature map; H and W are the height and width, respectively; and HW is the number of multiplications of the height and width. Then, after three parallel fully connected layers $\boldsymbol{W}_q$, $\boldsymbol{W}_k$, and $\boldsymbol{W}_v$, the channels are dimensionalized to obtain the $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_q$, $\boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_k$, and $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_v$ matrices, respectively. Next, the correlation matrix is obtained using $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{K}^T$, where $\boldsymbol{A}_{ij}$ represents the inner product of row i in $\boldsymbol{Q}$ and row j in $\boldsymbol{K}$, i.e., the correlation of the vectors at two different spatial locations. Each row of the correlation matrix $\boldsymbol{A}$ is normalized using the Softmax function and constrained to be within (0,1). Finally, the correlation matrix $\boldsymbol{A}$ is multiplied by $\boldsymbol{V}$ and passes through a fully connected layer $\boldsymbol{W}_s$. The channel dimensions are recovered to obtain the spatial significance-enhanced feature map $\boldsymbol{X}^S = \boldsymbol{A}\boldsymbol{V}\boldsymbol{W}_s$. The final feature expression is:

$$\boldsymbol{X}^S = \sigma(\boldsymbol{X}\boldsymbol{W}_q(\boldsymbol{X}\boldsymbol{W}_k)^T)\boldsymbol{X}\boldsymbol{W}_v\boldsymbol{W}_s \tag{1}$$

where $\boldsymbol{W}_q$, $\boldsymbol{W}_k$ and $\boldsymbol{W}_v \in \boldsymbol{R}^{C \times C \times 4}$, $\boldsymbol{W}_s \in \boldsymbol{R}^{(C/4) \times C}$, and $\sigma(.)$ is the Softmax function.

2.3.2. Channel Attention Mechanism

The operation of the channel dimension is similar to the above, and the four feature maps $F_i, i \in [1, 2, 3, 4]$ are first expanded into one-dimensional vectors along the width and height dimensions and transposed to obtain $\boldsymbol{X} \in \boldsymbol{R}^{WH \times C}$ through three fully connected layers, and the outputs are $\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_q$, $\boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_k$, and $\boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_v$. As dimensionality reduction would bring about too much information loss, the algorithm proposed in this article does not reduce the dimensionality of the channel. The correlation matrix is then obtained by $\boldsymbol{B} = \boldsymbol{K}^T\boldsymbol{Q}$, where $\boldsymbol{B}_{ij}$ represents the inner product of column i in $\boldsymbol{K}$ and column i in $\boldsymbol{Q}$, i.e., the correlation of the two different channel vectors. Similarly, each column of the correlation matrix $\boldsymbol{B}$ needs to be normalized using the Softmax function, constrained to be within (0,1). Finally, after multiplying $\boldsymbol{V}$ with $\boldsymbol{B}$ and passing through a fully connected layer $\boldsymbol{W}_s$, the channel significance-enhanced feature map $\boldsymbol{X}^C = \boldsymbol{V}\boldsymbol{B}\boldsymbol{W}_s$ is obtained, and the final feature expression is:

$$\boldsymbol{X}^C = \boldsymbol{X}\boldsymbol{W}_v\sigma((\boldsymbol{X}\boldsymbol{W}_k)^T\boldsymbol{X}\boldsymbol{W}_q)\boldsymbol{W}_s \tag{2}$$

where $\boldsymbol{W}_q$, $\boldsymbol{W}_k$, $\boldsymbol{W}_v\boldsymbol{W}_s \in \boldsymbol{R}^{C \times C}$. Finally, the outputs of these two branches are merged. Considering the effect of the residual structure, the merged features are summed with the input $\boldsymbol{X}$ to generate the final feature map $\boldsymbol{Y} \in \boldsymbol{R}^{WH \times C}$:

$$\boldsymbol{Y} = \boldsymbol{X}^C \oplus \boldsymbol{X}^S \oplus \boldsymbol{X} \tag{3}$$

where "$\oplus$" denotes the summation of the feature maps at the element level. After transposition and recovery of the dimension expansion, $\boldsymbol{Y}$ is fed into the subsequent module.

2.4. Transposed Convolution Upsampling Module (TCUM)

A transposed convolution [54] is a kind of learnable convolution, which is different from an ordinary convolution in that it makes the size larger, i.e., it is a kind of upsampling. Inspired by this, to improve the model's ability to restore multi-scale distress details, we designed the learnable upsampling TCUM module based on the transposed convolution and replaced the MLP layer at the end of SegFormer. As shown in Figure 2, the TCUM module has one more 3×3 convolution, two batch normalization (BN) layers, and two rectified linear unit (RELU) activation functions compared to the original SegFormer MLP. That is, firstly, the fused feature maps are changed from $128 \times 128 \times 768$ to $128 \times 128 \times (768 \times 4)$ by the 3×3 convolution with padding of 1. After this, the feature map is sequentially subjected to two transposed convolution operations with BN, RELU, and a kernel size of 2. The size of the output feature map is $128 \times 128 \times (768 \times 16)$. Finally, a 1×1-sized convolution is used to predict the $128 \times 128 \times 2$ resolution segmentation mask using the fused features after two transposed convolutions.

2.5. Mixed Loss Function

Unlike natural imagery tasks such as semantic segmentation on the Pascal VOC2012 dataset, the percentage of target pixels in the existing publicly available pavement distress segmentation datasets is generally small (Figure 4). In the ISTD-PDS7 training set used in this study, the percentage of pavement distress pixels is only 3.17%. It can also be seen that the real pavement distress belongs to a few classes in the image, which leads to an imbalance in their classification and segmentation, and this makes it difficult for the model to effectively learn the features of the distressed regions.

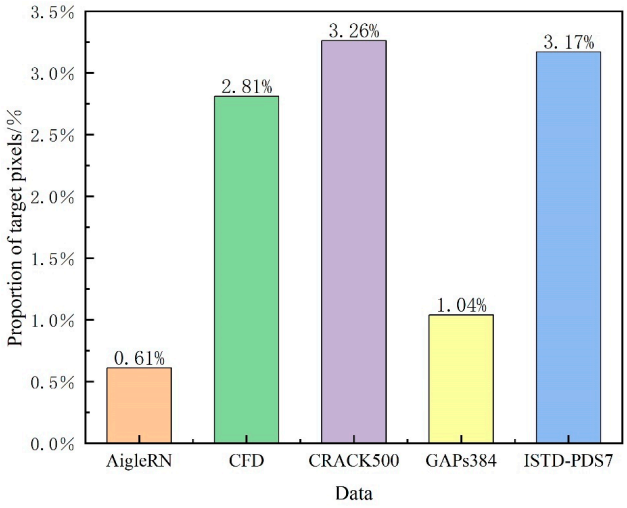


Figure 4. Target pixel proportion of different pavement distress segmentation datasets.

Zou et al. [5] found that, if the loss weight of cracks is directly increased, it leads to more false-positive predictions. Xu et al. [2] solved the problem of the imbalance in the percentage of crack pixels in the CRACK500 dataset by using a hybrid loss function combining focal loss [55] and dice loss [56], but the coefficient ratios of the two were not provided. As the training dataset used in this study includes multi-scale pavement lesions and the complexity of the scenes is high, to solve the problems of the class imbalance and segmentation accuracy, we combined the focal loss and dice loss and investigated the optimal coefficient ratios of the two to make this mixed loss function take into account the imbalanced classes and difficult sample challenges, while also improving the segmentation accuracy.

We adopted the focal loss function to alleviate the category imbalance between the distress pixels and background pixels. Specifically, it can improve the prediction ability of the model for a small number of samples by introducing an adjustable parameter to adjust the weight of the loss value for negative samples, which is calculated as follows:

$$L_{fl} = -\alpha(1 - p_t)^{\gamma} \log p_t \tag{4}$$

$$\alpha = \begin{cases} \alpha_t & \text{The current sample is a distress} \\ 1 - \alpha_t & \text{The current sample is the background} \end{cases} \tag{5}$$

where p_t represents the probability of a pixel being correctly classified, and p_t also reflects the difficulty of classification, where the larger the value of p_t, the higher the confidence of the classification, indicating that it is easier to classify the sample. The smaller the value of p_t, the lower the confidence of the classification, indicating that it is more difficult to classify the sample. α is the positive–negative sample balancing factor, α_t is the pre-set hyperparameter, and $\alpha_t = 0.5$. γ is the control-focusing parameter to make the loss more focused on difficult samples, which generally takes the value of 2 [56]. $(1 - p_t)^{\gamma}$ is the dynamic-modulating factor. The smaller the p_t value of the predicted distress pixel, the more inaccurate the prediction is, and the more the focal loss tends to regard this distress

as difficult to classify. At the same time, the larger the value of $(1-p_t)^{\gamma}$, the greater the contribution of the difficulty of classifying the distress type to the loss.

The dice loss is derived from the dice similarity coefficient, which is a metric function used to assess the similarity of samples, and it is a loss function based on segmentation modeling, which evaluates how good a model is by measuring the similarity between the model predictions and the true labels [56]. Specifically, the dice loss calculates the ratio of the intersecting part of the predicted label and the true label to the sum of its two parts as the accuracy of the model prediction. Therefore, the dice loss can solve the problem of the class imbalance and improve the prediction accuracy of the model for data with insignificant class boundaries, and it is calculated as follows:

$$L_{Dice} = 1 - \frac{2|X \cap Y|}{|X| + |Y|} \tag{6}$$

where X and Y denote the set of points contained in the real and predicted contour regions, respectively.

Based on the calculation of the above loss function, we can express the mixed loss function as follows.

$$L_{Mixed} = \lambda_1 L_{dice} + \lambda_2 L_{fl} \tag{7}$$

where λ_1 and λ_2 are the weight coefficients of the dice loss and focal loss. By adjusting λ_1 and λ_2, we can control the weights of each component of the hybrid loss, fully combining the advantages of the focal loss and dice loss, to solve the problems of the positive and negative class imbalance and segmentation accuracy. In this study, the mixed loss function was used to replace the cross-entropy loss function in the benchmark model, where $\lambda_1 = 0.7$ and $\lambda_2 = 0.3$. The selection of the weight coefficients is discussed in Section 3.5.

3. Results and Discussion

3.1. Implementation Details

- Parameter settings: We implemented the evaluation networks using the publicly available PyTorch1.7.0, which is well-known in this community. To improve the learning performance, we adopted the migration-learning method to train the proposed model. The pre-trained model selected the optimal weights of the backbone network trained on the Cityscapes dataset. The model training and testing were performed on a Windows 10 system using Python 3.6. The model training was divided into a freezing phase (50 EP) and a thawing phase (100 EP). The batch size of the freezing training phase was set to 4 and the learning rate was set to 1×10^{-4}. The batch size of the unfreezing process was set to 2 and the learning rate was set to 1×10^{-5}. The momentum and weight decay were set to 0.9 and 0, respectively. The model-training process was optimized using the AdamW optimizer with a learning rate of 1×10^{-4}. All the experiments described in this article were performed using a single GeForce RTX 3090 GPU.
- Datasets: To reveal the superiority of ISTD-DisNet, we choose the ISTD-PDS7 [20] benchmark dataset for the model training and testing. The original images in the ISTD-PDS7 dataset were acquired using a mobile acquisition vehicle. This dataset covers seven types of common pavement distress, with different scales for the different distress types. In addition, this dataset has high scene complexity and labeling fineness and contains a sufficient number of negative samples with textural similarity noise, such as shadows, water or oil stains, dropped objects, pavement appendages, etc. ISTD-PDS7 contains 6553 sample images of lesions and 11,974 negative sample images with interference noise. The size of the dataset after data enhancement (vertical flip, horizontal flip, flip, and transpose) was 30,475. The ratio of the ISTD-TR training set to the ISTD-VD validation set was 9:1. ISTD-TE (1000) and ISTD-CRTE (550) were used as the test sets to evaluate the performance in the multi-type distress task and crack segmentation task, respectively, in complex scenarios. Figure 5 shows examples of

seven types of distress and negative samples in different scenarios, with the first row showing the original images and the second row showing the corresponding labeled images. Finally, we evaluated the generalizability of the ISTD-DisNet model using the CRACK500 [34], CFD [35], AigleRN [36], and GAPs384 [34] datasets.

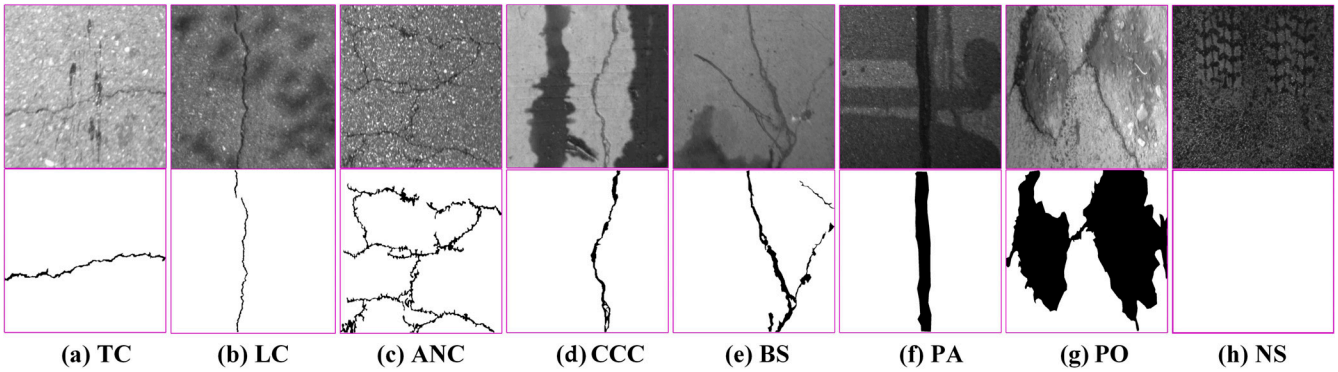


Figure 5. Example images of ISTD-PDS7, where "TC", "LC", "CCC", "ANC", "BS", "PA", "PO", and "NS" are the abbreviations for "transverse crack", "longitudinal crack", "cement concrete crack", "alligator network crack", "broken slab", "patch", "pothole", and "negative sample".

- Comparison algorithms: In this study, we evaluated the performance of ISTD-DisNet and seven mainstream semantic segmentation models on the ISTD-PDS7 dataset. The comparison algorithms included five CNN-based segmentation networks (SegNet [57], PSPNet [58], DeepLabv3+ [59], U-Net [31], HRNet [60]) and two transformer-based segmentation networks (Swin-Unet [42], SegFormer [44]). In addition, ablation experiments were performed to verify the effectiveness of the improved methods at three locations.
- Evaluation metrics: To provide a comprehensive evaluation, the *F*1 score and the mean intersection over union (*mIoU*) were used to quantitatively evaluate the performance of the different segmentation models. Specifically, the *F*1 score is the weighted harmonic average, which was used to measure the comprehensive performance of the model. The *IoU* is the ratio between the intersection and concatenation between the predicted results and the true labels, which measured the segmentation ability of the model in terms of the degree of overlap between the predicted and actual distress. *mIoU* is the average of the *IoU*. For each image, the *Precision* and *Recall* can be calculated by comparing the detected distress with the human-annotated ground truth. The metrics can be calculated by the following formulas:

$$\begin{aligned} Precision &= \frac{TP}{TP + FP} \\ Recall &= \frac{TP}{TP + FN} \\ F1 &= \frac{2 \times Precision \times Recall}{Precision + Recall} \end{aligned} \tag{8}$$

$$mIoU = \frac{1}{N}\sum_{k=1}^{N}\frac{TP_K}{TP_K + FP_K + FN_K} \tag{9}$$

where TP_K, FP_K, and FN_K represent the true positive, false positive, and false negative, respectively; these are measured by the class throughout the test set, where $N = 2$.

3.2. Quantitative Evaluation

We used the ISTD-TE and ISTD-CRTE test sets to compare and analyze ISTD-DisNet with the seven other current state-of-the-art semantic segmentation models. The comparison metrics used here are the segmentation accuracy, computational complexity, and inference efficiency, with the segmentation accuracy denoted by the *F*1 score and *mIoU*, the computational complexity denoted by the number of network parameters (Par), the com-

putational complexity (CC) denoted by GFLOPs, and the inference efficiency denoted by frames per second (FPS). Table 2 lists the quantitative comparison metrics for all the experimental models, where V16 = VGG16 [61], MV2 = MobileNetV2 [62], R-50 = ResNet50 [63], H-V2 = HRNetV2 [60], X = Xception [64], and ST = Swin transformer [42]. It is worth noting that the test results of the comparison models in Table 2 are from our previous benchmark assessment of the ISTD-PDS7 dataset [20].

Table 2. Quantitative evaluation on the ISTD-PDS7 validation and test sets. Bold indicates the best performing item.

Dataset	Metric	SegNet [57]	PSPNet [58]		DeepLabv3+ [59]		U-Net [31]	HRNet [60]	Swin-UNet [42]	SegFormer [44]	ISTD-DisNet
Attribute	Backbone	V16	MV2	R-50	MV2	X	V16	H-V2	ST	MiT-B2	MiT-B2
	Input size	416 × 416	473 × 473	473 × 473	512 × 512	512 × 512	512 × 512	480 × 480	224 × 224	512 × 512	512 × 512
	Par/M	16.32	**2.38**	46.71	5.813	54.709	24.89	29.538	27.18	27.35	28.94
	CC/GFLOPs	**601.78**	5.28	118.43	52.87	166.841	450.602	79.915	52.56	113.427	176.331
	Speed/FPS	57.91	**131.38**	75.53	100.32	47.05	26.99	23.32	110.95	34.75	28.35
ISTD-TE	*F1*/%	71.55	89.64	83.04	88.68	89.44	92.03	92.72	89.11	94.23	**95.51**
	mIoU/%	60.4	82.27	73.55	81.03	81.13	85.96	87.07	81.64	89.49	**91.67**
ISTD-CRTE	*F1*/%	71.44	73.61	74.57	81.98	79.25	85.91	85.07	79.75	87.14	**89.24**
	mIoU/%	57.05	63.5	65.00	72.66	67.5	77.5	76.43	70.12	79.12	**81.91**

From Table 2, it can be observed that, as far as the evaluation results for ISTD-TE and ISTD-CRTE are concerned, the performance of the different models in the crack segmentation task (ISTD-CRTE) is lower than that in the multi-type distress class dichotomous image segmentation task (ISTD-TE), which is because the structure of cracks is very complex and delicate. This requires the model to retain as much spatial information as possible, which is a challenge for most models. Compared to the other models, the proposed ISTD-DisNet achieves the most competitive performance in two evaluation metrics, and the original SegFormer takes second place in the ranking, which proves the superiority of the hierarchical transformer encoder in multi-scale distress feature extraction. In comparison, the *mIoU* values of Swin-UNet for ISTD-TE and ISTD-CRTE are decreased by 10.03 and 11.79, respectively. It is worth noting that, although Swin-UNet possesses a similar number of network parameters to SegFormer, it has a lower computational complexity (GFLOPs = 52.56) and a higher computational speed (FPS = 110.95). In addition, the convolution-based HRNet and U-Net achieve moderate performances, with the former performing better in the two-class semantic segmentation task and the latter performing better in the crack segmentation task. The two models also achieve the lowest FPS values and have a slow inference speed. SegNet, with the simple encoder–decoder architecture, performs the worst of all.

3.3. Visual Performance

To more intuitively present the distress segmentation performance of the different models in complex scenes, we selected seven pavement distress and negative sample images containing interferences in the test set, some of which contained interference noise such as oil stains, shadows, and tree branches. Figure 5 shows the segmentation results of the eight algorithms.

As shown in Figure 6, a visual inspection reveals that the proposed method outperforms the other methods in the task of the two-class semantic segmentation of multiple types of pavement lesions in different natural scenarios. The proposed method performs especially well in coping with different types and sizes of distress, varied illumination, and disturbances, with the most fine-grained and complete lesion extraction, with fewer false-positive predictions. From the prediction results of the different segmentation networks based on CNN structures, as shown in column 3 of Figure 6, SegNet only achieves coarse segmentation of the distress samples and is poor at handling the details and discontinuities. It also has difficulty extracting the planar regions of pits. In addition, from the segmentation results of PSPNet and DeepLabv3+ in columns 4–7 of Figure 6, it can be found that, as these segmentation networks use multi-scale pooling and cavity convolution, this can increase

the receptive field to a certain extent, but it leads to a large amount of spatial information loss, which results in a significant decrease in the small crack detection performance. In contrast, U-Net (column 8 of Figure 6) and HRNet (column 9 of Figure 6), which take into account the fusion of multi-scale feature information, show an improved ability to restore the details of the distress and are more suitable for the task of the intensive prediction of pavement distress; however, the inference speed is reduced (Table 2). Columns 10, 11, and 12 of Figure 6 demonstrate the potential of the transformer backbone-based feature extraction network structure for the intensive prediction task. In terms of the operation speed, Swin-UNet obtains the second best prediction speed, with an FPS of 110.95, due to the lower computational complexity and fewer parameters, but the model suffers from a loss of detail information due to the pure transformer operations, the insufficient edge information for crack extraction, and poor immunity to interference. In contrast, SegFormer uses a hierarchically structured transformer feature extraction network to output multi-scale feature maps, and it uses a 3×3 sized depthwise convolution instead of positional encoding for the position information. This avoids the problem of the degradation of the dense prediction performance due to the interpolation of the positional encoding when the test resolution is different from the training resolution. In addition, SegFormer uses the simplest MLP decoder for the feature information at different scales, and although it can effectively combine local attention and global attention, the model prediction is not sufficiently fine-grained compared to the method proposed in this article.

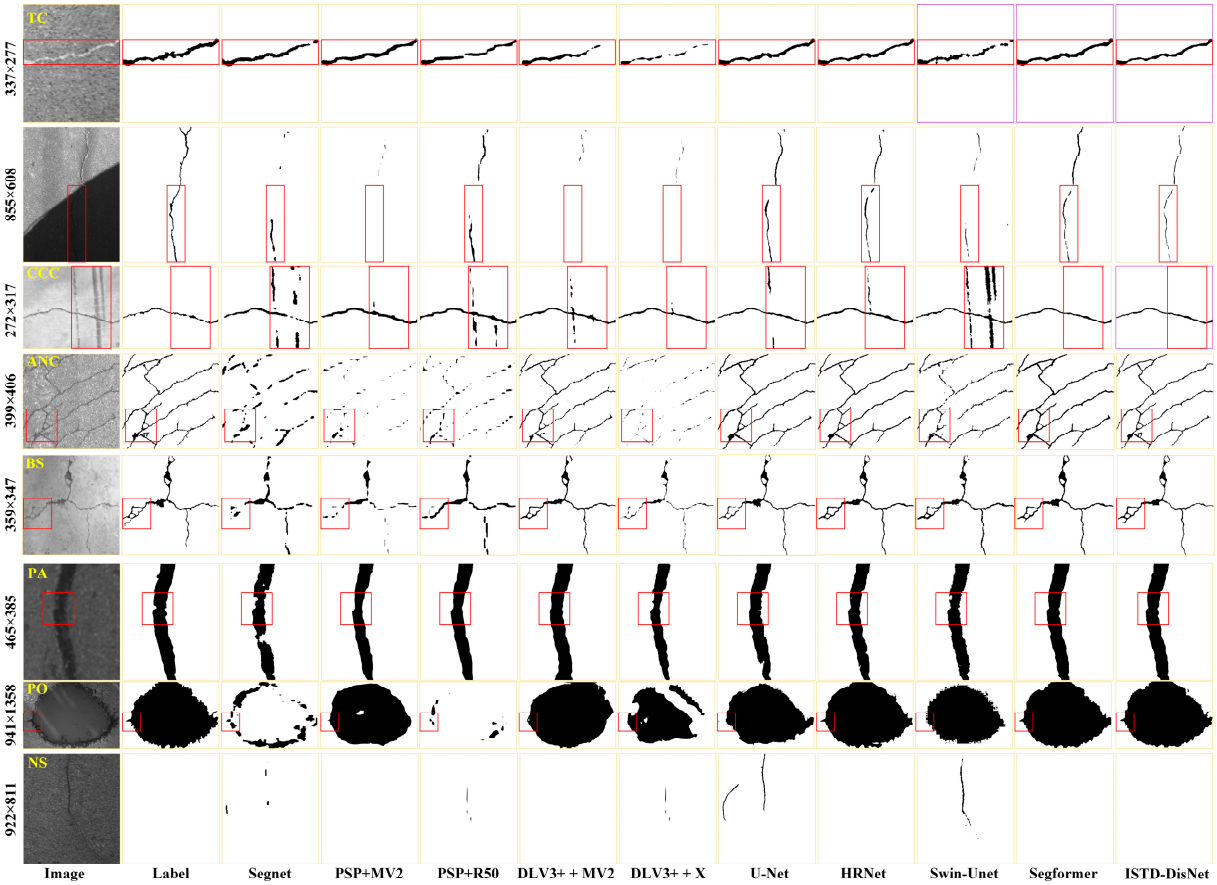


Figure 6. Segmentation results comparison for the seven classes of distress and distractors in complex scenes by the eight methods on the ISTD-PDS7. The red box is the focus area.

3.4. Ablation Study

To reveal the effectiveness of the different improvement methods in ISTD-DisNet, the MAM and TCUM modules and the mixed loss function ($\lambda_1 = 0.7$, $\lambda_2 = 0.3$) were replaced or removed, respectively. Table 3 shows that, as far as the effect of a single improvement

measure on the distress segmentation performance is concerned, the mixed loss function, which takes into account the positive and negative sample imbalance, contributes the most to the improvement in the *mIoU*. Compared with the baseline network, the model improves the *F*1 score and *mIoU* by 0.93% and 1.58%, respectively, without increasing the number of parameters, which suggests that the improvement strategy that focuses on the positive and negative sample imbalance is more effective for the pavement distress segmentation task. This is followed by the MAM module, and the smallest contribution is made by adding the learnable uploading module at the decoding stage (TCUM). The combination of TCUM and Mixed_Loss resulted in the greatest improvement in the model's segmentation performance when the two improved strategies were used in combination; the simultaneous use of the three improvement measures can improve the *F*1 score and *mIoU* of SegFormer by 1.28% and 2.18%, respectively. The results of the ablation experiments show that the three improvement measures in ISTD-DisNet can effectively improve the performance of pavement distress detection. With regard to the accuracy, computational complexity, number of parameters, and FPS, ISTD-DisNet has the highest accuracy and fewer parameters.

Table 3. Effectiveness of each module in ISTD-DisNet.

Method	*F*1/%	*mIoU*/%	CC/GFLOPs	Par/M	Speed/FPS
SegFormer	94.23	89.49	113.427	27.348	34.75
SegFormer + MAM	94.94	90.77	113.467	27.496	34.73
SegFormer + TCUM	94.83	90.49	176.316	28.767	28.35
SegFormer + Mixed_Loss	95.16	91.07	113.427	27.348	34.75
SegFormer + MAM + Mixed_Loss	94.84	90.51	113.467	27.496	34.73
SegFormer + TCUM + Mixed_Loss	95.20	91.12	176.316	28.767	28.35
SegFormer + MAM + TCUM	94.99	90.77	176.356	28.915	32.74
ISTD-DisNet	95.51	91.67	176.356	28.941	28.35

In addition, to more visually reveal the effects of the MAM and TCUM modules on the model's segmentation performance, we randomly selected different distress samples from the ISTD-PDS7 test set and used GradCAM to visually represent the feature maps of the TCUM module output locations in Figure 1, i.e., the output features of the MLP layer in the tail of the original SegFormer. GradCAM is a popular visualization method that takes gradient-weighted class-activation mappings in the form of heatmaps to be a visual representation and provides interpretability without compromising the accuracy. Darker areas in the visualization results indicate higher model responsiveness to the target class. The results are shown in Figure 7, where the first column contains the original images of the seven different distress categories; the second column contains the labels of the corresponding distress types; the third column contains the heatmaps of the corresponding locations of SegFormer; the fourth column contains the heatmaps of the corresponding locations after adding only the MAM module; the fifth column contains the heatmaps of the corresponding locations after adding only the TCUM module; and the sixth column contains the heatmaps of the corresponding locations for ISTD-DisNet. Observation of columns 3 and 4 reveals that the addition of the MAM module improves the model's response to the distressed areas and enhances the completeness of the predictions. In particular, for the cement cracks with expansion joints in the background (four rows), the MAM module is also effective in reducing the false-positive predictions of the model. Observation of columns 3 and 5 reveals that the addition of the TCUM module proposed in this article improves the model's ability to restore the details of the damage and solves the problem of the insufficient fine-grained segmentation of the baseline model. It should be noted that, when only the TCUM module is used, it leads to the incomplete extraction of some cracks and pits (two rows and five columns, two rows and seven columns). Observing columns 3 and 6, it can be seen that the simultaneous use of the MAM and TCUM modules

not only improves the model's response to the distressed area and its ability to restore details but also effectively increases the completeness of the crack extraction.

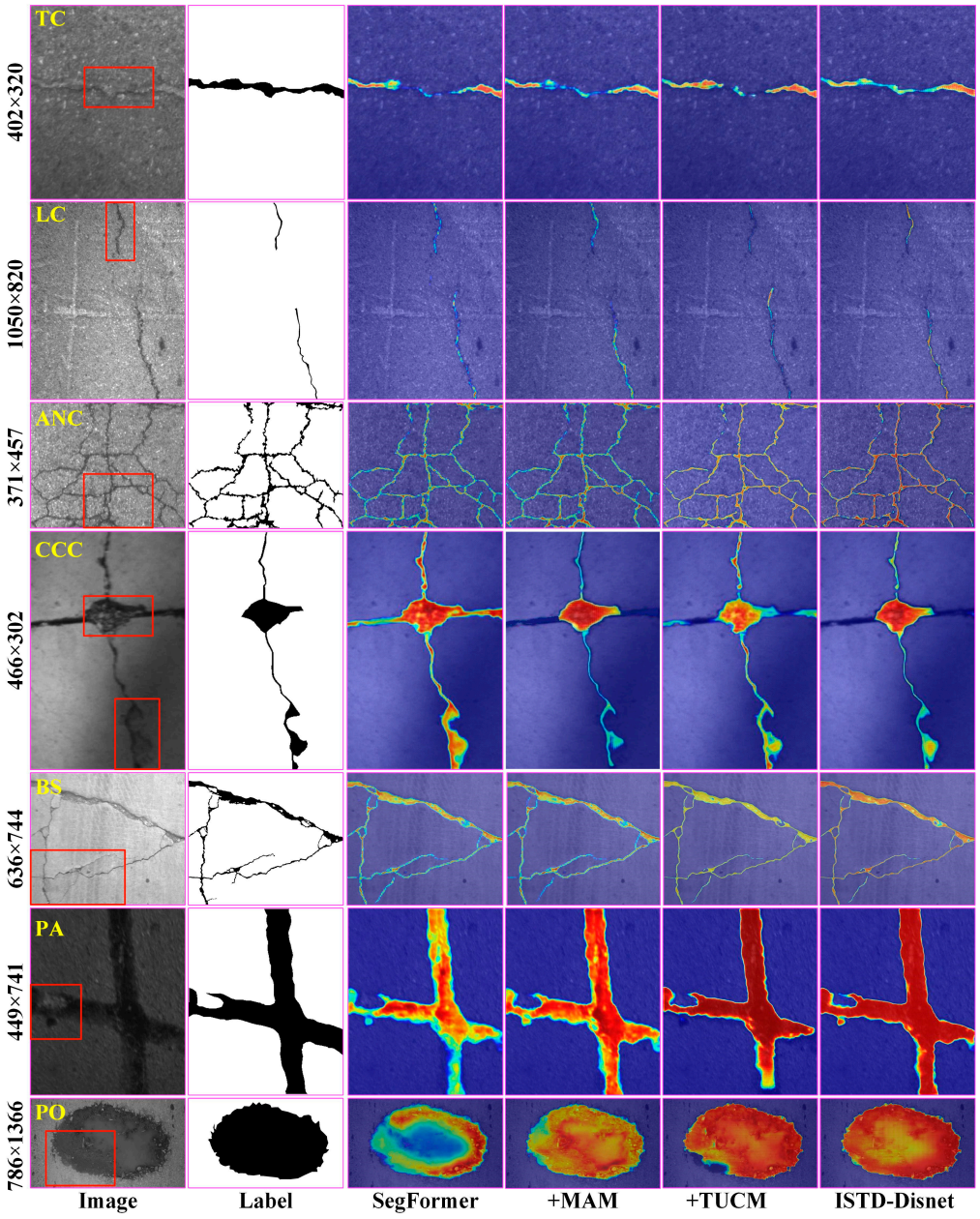


Figure 7. Heatmap visualization with the ISTD-TE dataset. The red box is the focus area.

3.5. Weight Coefficient Analysis for the Mixed Loss Function

In Section 2.5, we described how we combined the advantages of the dice loss and focal loss and weighted the sum of the two loss functions as a mixed loss function to address the challenges posed by the imbalanced classes and difficult samples. Given the different principles of the two loss functions, this section discusses the impact of the mixed loss function on the segmentation performance of models with different weighting coefficients. To this end, we set the values of (λ_1: λ_2) in Equation (7) to (1:0), (0.9:0.1), (0.8:0.2), (0.7:0.3), (0.6:0.4), (0.5:0.5), (0.4:0.6), (0.3:0.7), (0.2:0.8), (0.1:0.9), and (0:1), respectively. Figure 8 shows the loss curves of the validation set during ISTD-DisNet training with the different weighting coefficients, where Figure 8a is the initial phase of training (0–15 epochs) and Figure 8 is the unfrozen phase of training (50–150 epochs). It can be seen that, for the pavement distress two-class segmentation task with a small percentage of target pixels, the

loss curve fluctuates the most when only the dice loss function is used, i.e., the gradient changes drastically and the training is unstable. When using only the focal loss function, the oscillations are not obvious, but the convergence is slow. When (λ_1:λ_2) takes the value of (0.7:0.3), the loss curve fluctuation is the smallest, the loss value decreases the fastest, and the value of the convergence is the lowest, which indicates that the training process is more stable at this time and that training time can be saved.

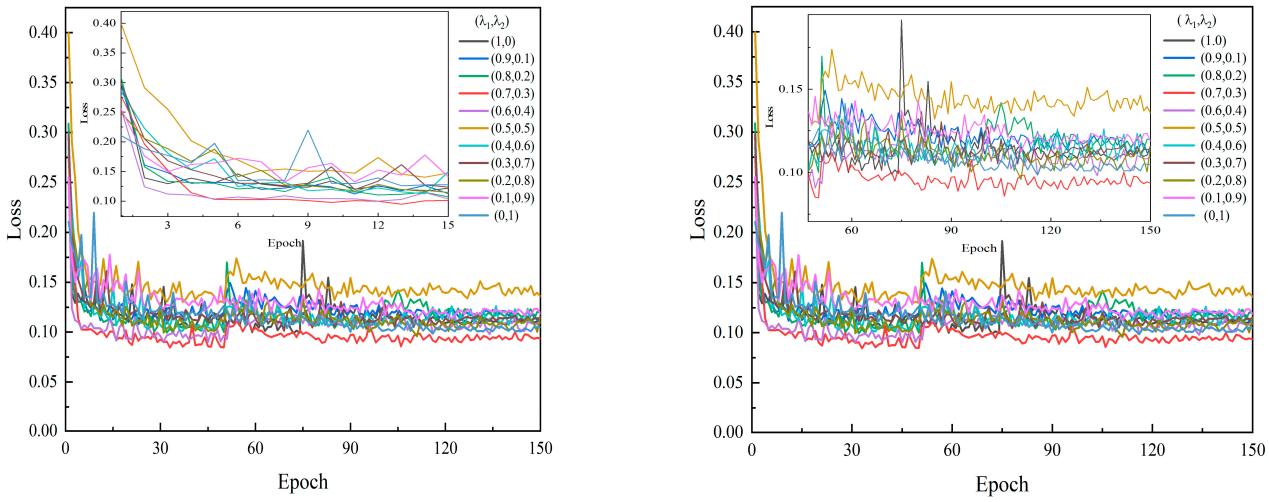


Figure 8. Loss curves during training of ISTD-DisNet: (**a**) the loss curve is amplified for the initial training phase, and (**b**) the loss curve at the end stage of the training is amplified.

Figure 9 demonstrates the variation in the *F*1 score and *mIoU* of ISTD-DisNet on the test set when different weighting coefficients are assigned to the two types of loss functions. It can be seen that, compared with the use of the focal loss alone, the use of the dice loss alone can better improve the model's distress segmentation performance. When the (λ_1:λ_2) ratio of the mixed loss function is (0.7:0.3), the model's segmentation performance is the best, indicating that, at this time, the mixed loss function is able to take into account the distribution of the number of samples, but at the same time, it is also able to better optimize the classification accuracy. Taken together, the combination of the focal loss and dice loss is a more effective way to solve the imbalance of positive and negative samples in pavement distress segmentation. For similar tasks, the weight coefficients can be selected and adjusted according to the characteristics of the dataset and the needs of the model, which can achieve better classification results.

3.6. Generalizability Analysis on Different Publicly Available Datasets

In order to verify the generalizability of the ISTD-DisNet model for pavement distress segmentation, we chose the original SegFormer as the baseline model for comparison and calculated the distress segmentation metrics of the two algorithms for the CRACK500 [34], CFD [35], AigleRN [36], and GAPs384 [34] datasets, which come from different filming equipment and scenarios. As shown in Table 4, compared to the original SegFormer, the *F*1 scores of the model proposed in this article are improved by 2.05%, 1.38%, 2.92%, and 2.19% on the four public test sets, and the mIoU scores are improved by 2.38%, 1.55%, 3.53%, and 2.23%, respectively. This shows that the model proposed in this article has superior robustness and generalizability. Although the mIoU of ISTD-DisNet on the public dataset is not higher than its score on the ISTD-CRTE test set, this is mainly because the labeling accuracy is not consistent.

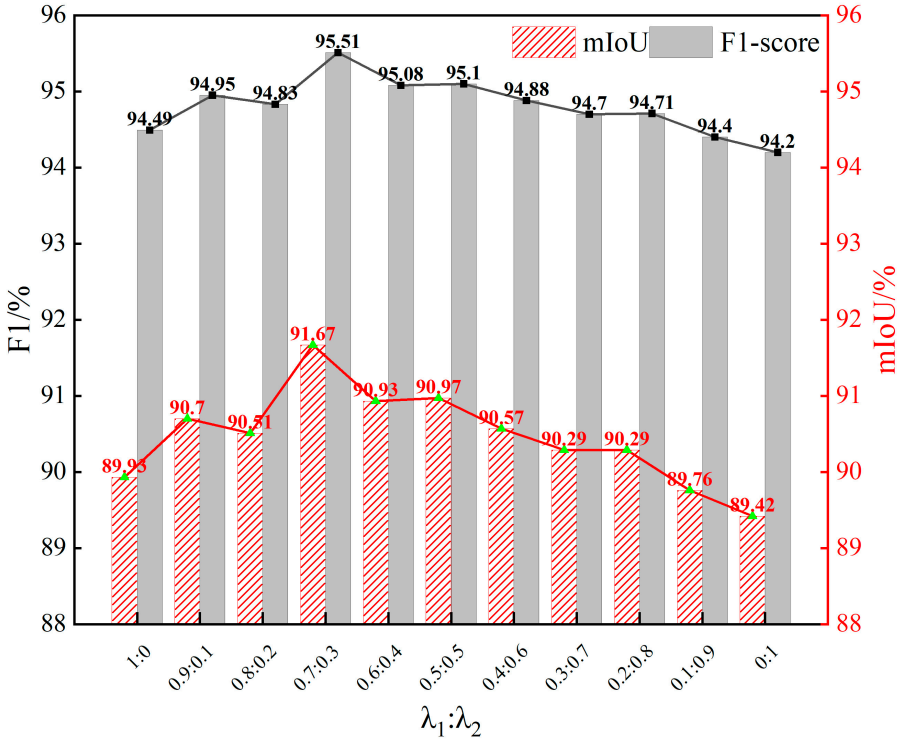


Figure 9. Variation in the segmentation performance of ISTD-DisNet with different values of λ_1:λ_2.

Table 4. Comparison of the crack segmentation results between ISTD-DisNet and SegFormer on four public datasets.

Dataset	Method	*F1*/%	*mIoU*/%	Equipment	Scenario
CFD	SegFormer	82.65	73.98	Smartphone	Urban roads in Beijing
	ISTD-DisNet	84.70	76.36		
CRACK500	SegFormer	81.66	72.72	Smartphone	Temple University campus road
	ISTD-DisNet	83.05	74.27		
AigleRN	SegFormer	71.53	62.45	Area-array camera	French humanoid path
	ISTD-DisNet	74.45	65.98		
GAPs384	SegFormer	70.46	61.91	Linear-array camera	German Autobahn
	ISTD-DisNet	72.65	64.14		

Figure 10 visualizes the segmentation results of the baseline and the model proposed in this article on the four datasets. Figure 10a shows the pavement crack images acquired by the different shooting devices, which have different resolutions, distress topologies, and texture noises. Figure 10b shows the corresponding labels of the original images. From Figure 10c,d, it can be seen that, compared with the baseline model, the ISTD-DisNet crack segmentation model proposed in this article has a better degree of completeness and refinement, which also indicates that the use of the MAM attention mechanism and transposed convolution in the encoding stage can effectively solve the problem of the baseline model's insufficient fine-grainedness for fine cracks and improve the model's detail restoration ability. In addition, for the GAPs384 dataset captured with a line-array CCD camera, the method proposed in this article can also accurately segment the multiple types of distress in the images. The results of the quantitative and qualitative comparative analysis of the generalization show that the segmentation method proposed in this article can accurately segment most of the pavement crack-type lesions and that the model has strong robustness.

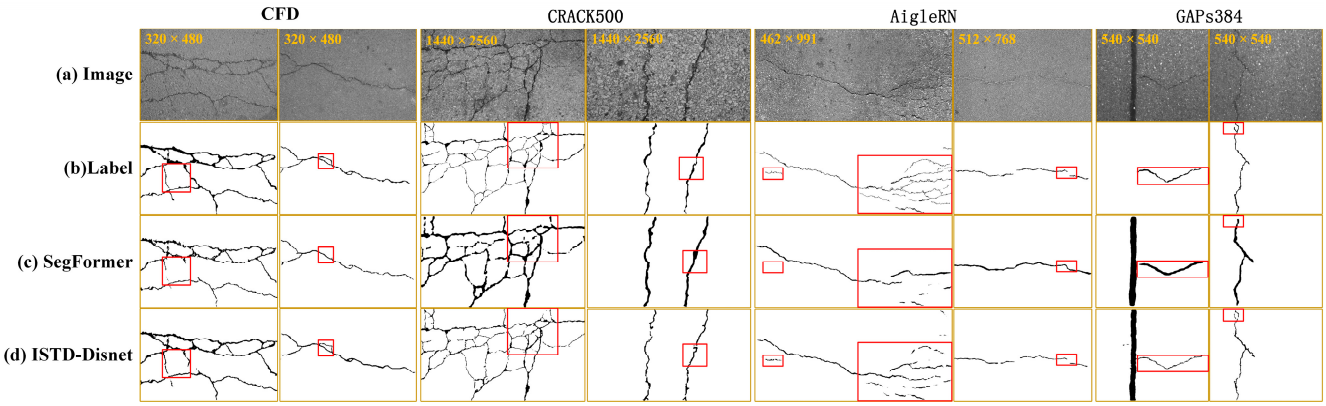


Figure 10. The qualitative experimental results of the generalization analysis. The red box is the focus area.

4. PCI Prediction Model Based on the ISTD-DisNet Outputs

4.1. PCI and Distress Pixel Density Ratio

(1) PCI assessment

Currently, road maintenance departments often refer to the internationally recognized PCI to develop maintenance programs. In the pavement condition evaluation task based on manual visual discrimination, Equation (10) is usually used to calculate the comprehensive pavement damage rate (DR) of each evaluation unit (100 m or 1000 m) of the route, and Equation (11) is used to calculate the PCI [6,49].

$$DR = 100 \times \sum_{i=1}^{i_0} \frac{w_i A_i}{A} \tag{10}$$

$$PCI = 100 - a_0 \times DR^{a_1} \tag{11}$$

where DR is the comprehensive damage rate of the pavement, which is the sum of the area of the various distress types and the percentage (%) of the pavement survey area; $PCI \in [0, 100]$; i is the type of distress; i_0 is the total number of types of distress; A_i is the area damaged by pavement distress in category i (m^2); A is the area of the surveyed pavement (the product of the surveyed length and width of the pavement, m^2); and w_i is the weighting of the pavement damage of the pavement distress in category i.

(2) Calculation of the distress *PDR*

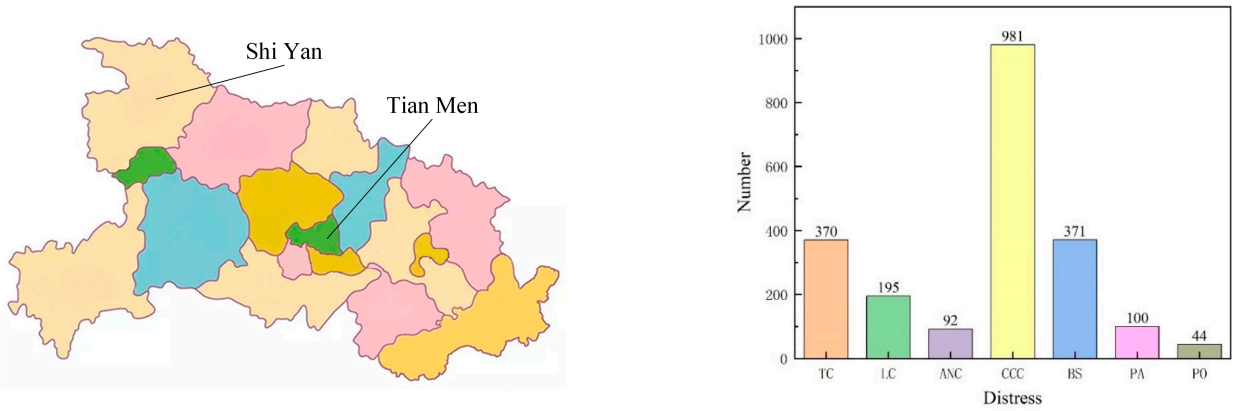
The manual visual discrimination process has the problems of a long detection period, high cost, and subjectivity [4]. Therefore, in the following, we describe how we calculated the distress *PDR* of each evaluation unit based on the output of the ISTD-DisNet model. We then describe how we investigated the correlation between the *PDR* and PCI and finally developed a new model for calculating the PCI. It should be noted that, due to the large differences in the background and distress type appearance of asphalt and cement concrete pavements, we analyzed and constructed the PCI calculation model for the two pavement types separately. If we assume that the examined line contains n 100 m evaluation units, then the optimal weighted model of ISTD-DisNet was used to extract all the distressed pixels in the image for each evaluation unit, and the *PDR* of the distressed pixels in that evaluation unit was calculated according to Equation (12):

$$PDR_i = \frac{\sum_{j=1}^{50} NP_j}{A_j \times 50} \tag{12}$$

where $i \in [1, n]$; NP_j is the total number of distressed pixels of image j in evaluation unit i; and A_j is the total number of pixels of image j. The ortho-corrected image resolution is $A_j = 3517 \times 2193$, which corresponds to an actual size of 3.2 m × 2.0 m.

4.2. Experimental Data

To build a robust and comprehensive PCI computational model, we screened 66 100 m evaluation units (42 asphalt, 26 cement concrete) by pavement type, distress type, and scenario from the 2023 sets of measured highway data from Hubei Province (Figure 11a) in China as the experimental data for the PCI computational modeling study. The statistics concerning the number of pavement distress types in the experimental data are shown in Figure 11b. Firstly, three experts in the field of distress inspection used manual visual discrimination to accurately outline the pavement lesions in the image of each evaluation unit and then calculated the PCI score of each evaluation unit, which was used as the true value for the model construction. ISTD-DisNet was then deployed on the experimental line, and the *PDR* of each evaluation unit was output. The results for the *PDR* and the PCI of the manual visual discrimination are shown in Table 5.



(**a**) (**b**)

Figure 11. Data-related information. From left to right: (**a**) data collection area; (**b**) plot of the quantity distribution of the various distress types in the experimental data, where "TC", "LC", "CCC", "ANC", "BS", "PA", and "PO" are the abbreviations for "transverse crack", "longitudinal crack", "cement concrete crack", "alligator network crack", "broken slab", "patch", and "pothole".

Table 5. Independent variables of the *PDR* and PCI.

Evaluation Unit ID	*PDR*	PCI (Visual Inspection Method)	Type of Pavement
1	0.00050	95.67	Asphalt pavement
2	0.00116	95.13	
3	0.00034	96.39	
4	0.00035	96.26	
5	0.00194	94.36	
6	0.00112	95.44	
7	0.00226	94.68	
8	0.00091	96.23	
9	0.00117	95.43	
10	0.00119	95.65	
11	0.00151	95.40	
12	0.00156	95.73	
13	0.00124	96.12	
14	0.00026	97.39	
15	0.00078	95.20	

Table 5. *Cont.*

Evaluation Unit ID	*PDR*	PCI (Visual Inspection Method)	Type of Pavement
16	0.00242	94.60	Asphalt pavement
17	0.00126	94.49	
18	0.00174	95.00	
19	0.00172	95.24	
20	0.00277	94.20	
21	0.00072	96.88	
22	0.00058	96.18	
23	0.00175	95.12	
24	0.00197	95.26	
25	0.00134	95.21	
26	0.00046	96.70	
27	0.00093	96.03	
28	0.00098	95.61	
29	0.00164	95.00	
30	0.00087	95.01	
31	0.00241	93.40	
32	0.00200	94.64	
33	0.00163	94.51	
34	0.00182	94.72	
35	0.00153	95.39	
36	0.00030	96.68	
37	0.00047	95.80	
38	0.00057	95.63	
39	0.00094	95.53	
40	0.00263	94.18	
41	0.00138	95.31	
42	0.00088	95.10	
43	0.00501	90.45	Cement concrete pavement
44	0.00025	96.48	
45	0.00170	92.95	
46	0.00029	95.72	
47	0.00116	94.26	
48	0.01719	86.58	
49	0.00269	92.64	
50	0.01674	85.53	
51	0.01694	86.00	
52	0.00293	91.77	
53	0.00206	93.35	
54	0.00101	96.64	
55	0.00744	89.40	
56	0.00221	91.50	
57	0.00032	91.65	
58	0.00883	89.36	
59	0.00323	91.11	
60	0.00029	94.88	
61	0.00858	89.94	
62	0.02228	86.15	
63	0.04044	83.30	
64	0.03881	81.47	
65	0.01107	87.37	
66	0.01580	85.93	

4.3. PCI Prediction Modeling

(1) Calculation of the distress *PDR*

To establish a PCI prediction model based on the output results of ISTD-DisNet, based on the data in Table 5, we calculated the *PDR* and PCI correlation coefficients according to

Equations (13) and (14) to determine the degree of correlation between the *PDR* and PCI indicators and analyze the feasibility of the automated prediction model construction.

$$\begin{cases} s_{PDR} = \sqrt{\dfrac{\sum_{i=1}^{i=n}(PDR_i - PDR_{mean})}{n-1}} \\ s_{PCI} = \sqrt{\dfrac{\sum_{i=1}^{i=n}(PCI_i - PCI_{mean})}{n-1}} \\ S_{PDR,PCI} = \dfrac{\sum_{i=1}^{i=n}(PDR_i - PDR_{mean})(PCI_i - PCI_{mean})}{n-1} \end{cases} \tag{13}$$

where $n = 66$ is the number of evaluation units; PDR_{mean} and PCI_{mean} are the arithmetic mean of the two samples, respectively; s_{PDR} and s_{PCI} are the standard deviation of the two samples, respectively; and $S_{PDR,PCI}$ is the covariance of *PDR* and *PCI*. The correlation coefficient $r_{PDR,PCI}$ is calculated according to Equation (14):

$$r_{PDR,PCI} = \frac{S_{PDR,PCI}}{s_{PDR} \times s_{PCI}} \tag{14}$$

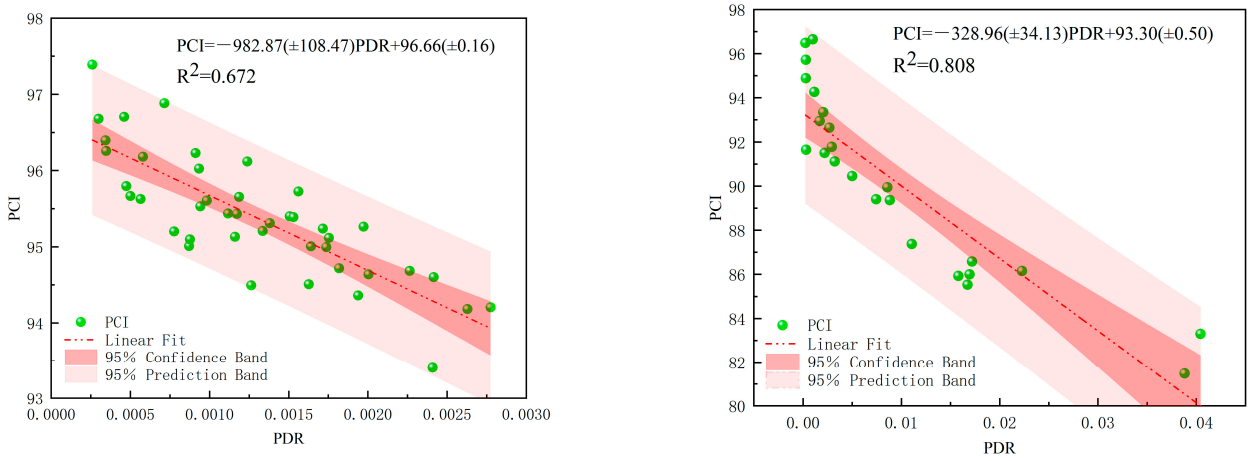
From Evans and Groot [65], it can be seen that, when $1 \geq |r_{PDR,PCI}| > 0.75$, the two have a strong correlation; when $0.7 \geq |r_{PDR,PCI}| \geq 0.3$, the two have a moderate correlation; and when $0.25 \geq |r_{PDR,PCI}| \geq 0$, the two have a weak correlation. Based on the data in Table 5, we calculated the *PDR* and PCI correlation coefficients of the 42 asphalt evaluation units and the 24 cement evaluation units by Equation (14) The results of the calculations are $r_{Asphalt} = -0.824$ and $r_{Cement\ concrete} = -0.965$. This indicates that the two have a strong correlation. That is, using the calculation results of ISTD-DisNet to calculate the distress *PDR* of the evaluation units and taking it as the independent variable of the pavement damage index calculation can respond to the degree of damage of the pavement and realize the construction of the PCI prediction model.

(2) The PCI prediction model

In this study, based on the *PDR* index, linear and nonlinear fitting methods were used to construct the PCI calculation model, and the performance of the model was evaluated using the coefficient of determination, R^2. R^2 can reflect the proportion of all the variations of the dependent variable that can be explained by the independent variable through the regression relationship, where the closer R^2 is to 1, the better the performance of the model. Figure 12 demonstrates the PCI computational model based on the simple linear fitting method, where it can be seen that, for the two types of pavement types, the coefficients of determination of the linear function model constructed based on the *PDR* are 0.672 and 0.808, respectively.

Nonlinear equations have a wide range of applications in electricity, mechanics, economic management, engineering technology, etc. [66]. It can be seen from Equation (11) that there is a nonlinear relationship between the PCI and the breakage rate DR, which belongs to a kind of exponential function containing two coefficients. Considering the large difference in the size of the *PDR* and DR values, we chose Equation (15) as the nonlinear fitting function of the *PDR* (independent variable) and PCI (dependent variable), where A_1 and A_2 are the two coefficients to be solved, and ε represents the correction parameters. The Levenberg–Marquardt (LM) algorithm [67] is used to solve the least-squares solution of the nonlinear function coefficients.

$$PCI = 100 - A_1 \times PDR^{A_2} + \varepsilon \tag{15}$$



(**a**) (**b**)

Figure 12. PCI prediction model based on a linear fitting algorithm: (**a**) asphalt; and (**b**) cement concrete.

Figure 12 shows the *PDR*-based PCI computational model after 29 iterations. It can be seen that, compared to the simple linear fitting model, the R^2 is improved by 0.036 and 0.144, respectively, using the PCI computational model. This indicates that the PCI computational model constructed using the nonlinear regression approach performs better.

4.4. Model Verification

To further evaluate the generalization ability of the PCI prediction model proposed in this article, the developed model was deployed for computation on 6 road sections with different road conditions, with a total of 89 100 m evaluation units (46 asphalt, 43 cement). Firstly, the PCI score of each evaluation unit was obtained by manual visual discrimination, and the total time was recorded. Secondly, all the images were analyzed using ISTD-DisNet to calculate the distress *PDR* for each evaluation unit. Finally, the extracted *PDR* features were used as the input values, and a nonlinear regression model (Figure 13) was used to calculate the PCI score of each evaluation unit and count the total computation time. Figure 14 shows the PCI calculation results of the two different methods. As can be seen from Figure 14, the PCI values calculated for each evaluation unit based on the distress *PDR* are in good agreement with the results of the corresponding manual visual discrimination. The maximum difference between the two methods is 7.33 and 8.93 for the asphalt and concrete pavements, respectively, and the mean absolute values of the differences are 2.48 and 3.92, respectively. It can be observed that the PCI obtained from the manual visual discrimination is lower than that calculated from the *PDR*-based PCI for nearly 80% of the evaluation units, which is because the training dataset used in this study did not contain all the distress types, such as the less common oiling, loosening, and exposed aggregate types. In addition, the efficiency of the PCI prediction model proposed in this article is about 24 times higher than that of manual visual discrimination. In summary, the computational model proposed in this article can realize the rapid calculation and ranking of pavement damage conditions, which has potential value in engineering applications.

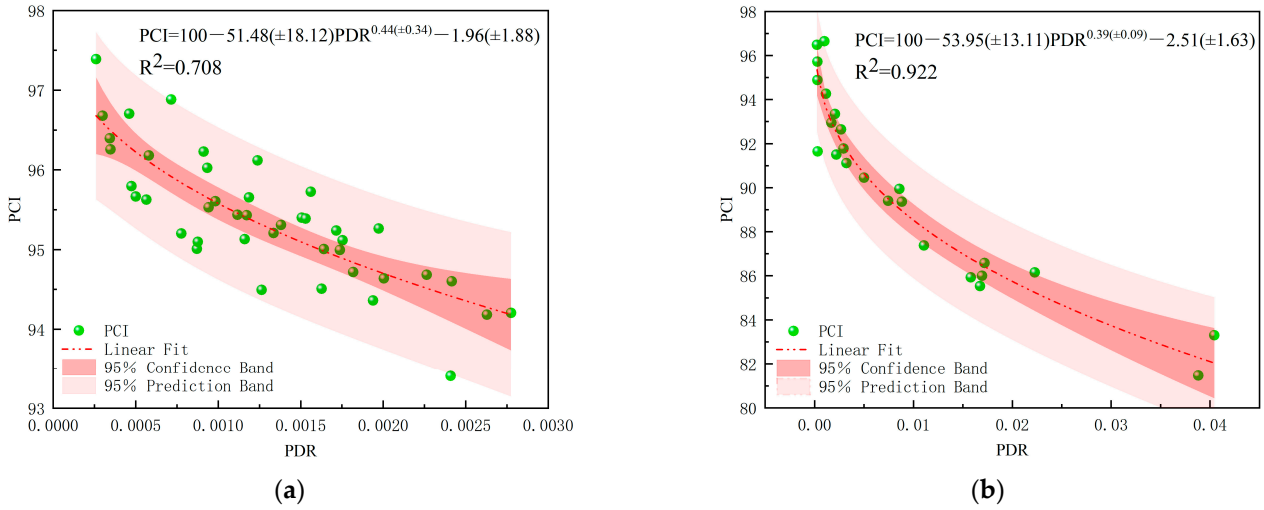


Figure 13. PCI prediction model based on the proposed algorithm: (**a**) asphalt; and (**b**) cement concrete.

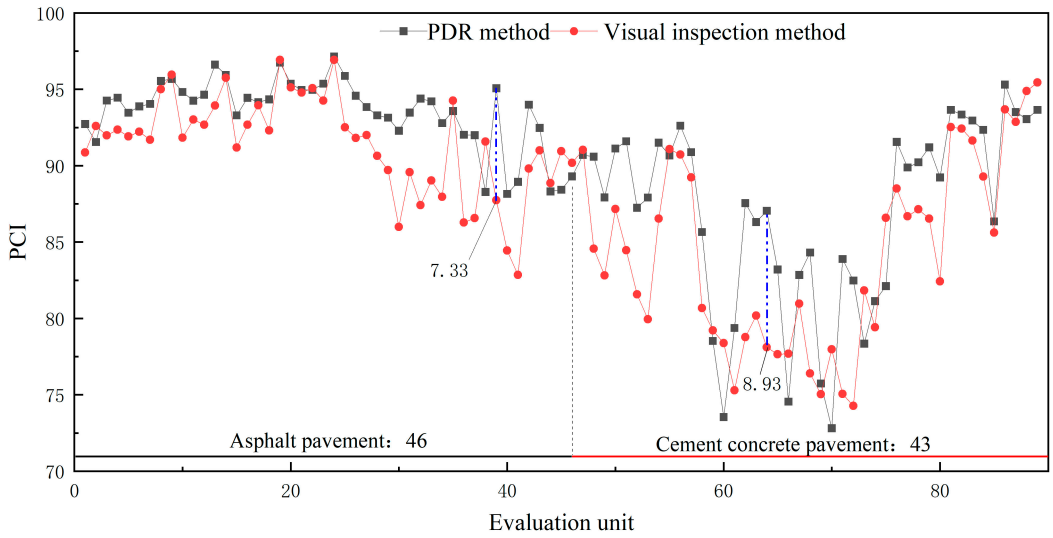


Figure 14. Comparison of the PCI calculation results based on the *PDR* and manual visual discrimination. The difference between the two is largest at the short blue line.

5. Conclusions

In this article, we have systematically investigated the tasks of binary semantic segmentation of multi-type pavement distress and the quantitative evaluation of pavement damage conditions in complex scenarios in terms of both research and application. To this end, we have proposed the ISTD-DisNet architecture for multi-type pavement damage detection based on the ISTD-PDS7 dataset presented in our previous research. Firstly, ISTD-DisNet realizes the extraction of different-scale damage feature information with the help of the SegFormer feature extraction network. Secondly, we designed the MAM module, which enables the model to pay more attention to the pavement distress features in different channels and different spatial locations in the decoding stage and reduces the false-positive prediction of the model. In addition, the TCUM module is used to cope with the challenge of restoring different distress details in complex scenes. Finally, a weighted and mixed loss function is constructed to alleviate the problem of the positive and negative sample imbalance by setting the focal loss and dice loss weights to improve the segmentation performance for distress features. The experimental results obtained on the ISTD-PDS7 dataset show that the developed ISTD-DisNet performs well in segmenting different types

of distress in complex scenarios. In addition, the ablation study and comparisons showed the effectiveness of the three main improvements proposed in this article in enhancing the accuracy of distress detection. The evaluation results obtained on four pavement crack datasets confirmed that the proposed method is highly robust in coping with crack detection in different scenarios. ISTD-DisNet may be applied to other surface damage segmentation tasks, such as the intelligent detection of bridge cracks and spalling.

In contrast the previous studies of pavement distress segmentation modeling, we proposed a fully automated pavement damage condition evaluation model based on the segmentation results. Firstly, pavement images of 66 100 m evaluation units were obtained from different road sections in Hubei Province, China, and the PCI of each evaluation unit was obtained by manual visual discrimination. Secondly, the distress segmentation results for the images of each evaluation unit were obtained using ISTD-DisNet, and the *PDR* was calculated. Finally, the prediction model of the PCI with the *PDR* was established. The model validation results showed that the prediction model proposed in this article is more consistent with the results obtained from manual visual discrimination and can realize the rapid calculation and ranking of pavement damage conditions, with potential value in engineering applications. In the future, other less common pavement distress samples will be collected in a targeted manner, and the use of pavement triple-point cloud data to generate depth images will be considered to increase the number of rutting distress samples related to depth [68], thus further optimizing the prediction accuracy of the evaluation model.

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

Funding: This research was funded in part by the National Natural Science Foundation of China under Grants 42071343 and 42204031, and in part by the Basic Scientific Research Expenses of Heilongjiang Provincial Universities, China, under Grant 2020-KYYWF-0690.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

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

References

1. Babkov, V.F. *Road Conditions and Traffic Safety*; Mir Publishers: Portsmouth, NH, USA, 1975.
2. Hamed, M.; Yaw, A.; William, G.B. Deep machine learning approach to develop a new asphalt pavement condition index. *Constr. Build. Mater.* **2020**, *247*, 118513.
3. Xu, C.; Zhang, Q.; Mei, L.; Chang, X.; Ye, Z.; Wang, J.; Ye, L.; Yang, W. Cross-Attention-Guided Feature Alignment Network for Road Crack Detection. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 382. [CrossRef]
4. Hu, G.; Hu, B.; Yang, Z.; Huang, L.; Li, P. Pavement Crack Detection Method Based on Deep Learning Models. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5573590. [CrossRef]
5. Zou, Q.; Zhang, Z.; Li, Q.; Qi, X.; Wang, Q.; Wang, S. DeepCrack: Learning hierarchical convolutional features for crack detection. *IEEE Trans. Image Process.* **2019**, *28*, 1498–1512. [CrossRef]
6. Song, W.; Jia, G.; Jia, D.; Zhu, H. Automatic pavement crack detection and classification using multiscale feature attention network. *IEEE Access* **2019**, *7*, 171001–171012. [CrossRef]
7. Stricker, R.; Aganian, D.; Sesselmann, M.; Seichter, D.; Engelhardt, M.; Spielhofer, R.; Gross, H.M. Road surface segmentation-pixel-perfect distress and object detection for road assessment. In Proceedings of the 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), Lyon, France, 23–27 August 2021; pp. 1789–1796.
8. Zhang, C.; Nateghinia, E.; Miranda-Moreno, L.; Sun, L. Pavement distress detection using convolutional neural network (CNN): A case study in Montreal, Canada. *Int. J. Transp. Sci. Technol.* **2021**, *11*, 298–309. [CrossRef]

9. Li, J.; Yuan, C.; Wang, X. Real-time instance-level detection of asphalt pavement distress combining space-to-depth (SPD) YOLO and omni-scale network (OSNet). *Autom. Constr.* **2023**, *155*, 105062. [CrossRef]
10. Cheng, H.; Shi, X.J.; Glazier, C. Real-Time image thresholding based on sample space reduction and interpolation approach. *J. Comput. Civ. Eng.* **2003**, *17*, 264–272. [CrossRef]
11. Ayenu-Prah, A.Y.; Attoh-Okine, N.O. Evaluating pavement cracks with bidimensional empirical mode decomposition. *EURASIP J. Adv. Signal Process.* **2008**, *2008*, 861701. [CrossRef]
12. He, Y.; Qiu, H.; Jian, W.; Wei, Z.; Xie, J. Studying of road crack image detection method based on the mathematical morphology. In Proceedings of the 2011 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Volume 2, pp. 967–969.
13. Amhaz, R.; Chambon, S.; Idier, J.; Baltazart, V. Automatic crack detection on Two-Dimensional pavement images: An algorithm based on minimal path selection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2718–2729. [CrossRef]
14. Zhang, D.; Li, Q.; Chen, Y.; Cao, M.; He, L.; Zhang, B. An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection. *Image Vis. Comput.* **2017**, *57*, 130–146. [CrossRef]
15. Li, N.; Hou, X.; Yang, X.; Dong, Y. Automation recognition of pavement surface distress based on support vector machine. In Proceedings of the 2009 Second International Conference on Intelligent Networks and Intelligent Systems, Tianjian, China, 1–3 November 2009; pp. 346–349.
16. Carvalhido, A.G.; Marques, S.; Nunes, F.D.; Correia, P.L. Automatic Road Pavement Crack Detection Using SVM. Master's Thesis, Instituto Superior Técnico, Lisbon, Portugal, 2012.
17. Ai, D.H.; Jiang, G.Y.; Siew Kei, L.; Li, C.W. Automatic Pixel-Level pavement crack detection using information of Multi-Scale neighborhoods. *IEEE Access* **2018**, *6*, 24452–24463. [CrossRef]
18. Hoang, N. An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable Filter-Based feature extraction. *Adv. Civ. Eng.* **2018**, *2018*, 7419058. [CrossRef]
19. Xu, Z.; Guan, H.; Kang, J.; Lei, X.; Ma, L.; Yu, Y.; Chen, Y.; Li, J. Pavement crack detection from CCD images with a locally enhanced transformer network. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *110*, 102825. [CrossRef]
20. Song, W.; Zhang, Z.; Zhang, B.; Jia, G.; Zhu, H.; Zhang, J. ISTD-PDS7: A Benchmark Dataset for Multi-Type Pavement Distress Segmentation from CCD Images in Complex Scenarios. *Remote Sens.* **2023**, *15*, 1750. [CrossRef]
21. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
22. Cao, W.; Liu, Q.F.; He, Z.Q. Review of Pavement Defect Detection Methods. *IEEE Access* **2020**, *8*, 14531–14544. [CrossRef]
23. Dung, C.V.; Sekiya, H.; Hirano, S.; Okatani, T.; Miki, C. A vision-based method for crack detection in gusset plate welded joints of steel bridges using deep convolutional neural networks. *Autom. Constr.* **2019**, *102*, 217–229. [CrossRef]
24. Xu, H.; Su, X.; Wang, Y.; Cai, H.; Cui, K.; Chen, X. Automatic bridge crack detection using a convolutional neural network. *Appl. Sci.* **2019**, *9*, 2867. [CrossRef]
25. Cha, Y.J.; Choi, W.; Oral, B. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [CrossRef]
26. Tran, T.S.; Tran, V.P.; Lee, H.J.; Flores, J.M.; Le, V.P. A two-step sequential automated crack detection and severity classification process for asphalt pavements. *Int. J. Pavement Eng.* **2020**, *23*, 2019–2033. [CrossRef]
27. Wu, Z.; Kalfarisi, R.; Kouyoumdjian, F.; Taelman, C. Applying deep convolutional neural network with 3D reality mesh model for water tank crack detection and evaluation. *Urban Water J.* **2020**, *17*, 682–695. [CrossRef]
28. Jeong, D. Road damage detection using YOLO with smartphone images. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 5559–5562.
29. Huang, H.; Li, Q.; Zhang, D. Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunn. Undergr. Space Technol.* **2018**, *77*, 166–176. [CrossRef]
30. Jenkins, M.D.; Carr, T.A.; Iglesias, M.I.; Buggy, T.W.; Morison, G. A deep convolutional neural network for semantic Pixel-Wise segmentation of road and pavement surface cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Roma, Italy, 3–7 September 2018; pp. 2120–2124.
31. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351.
32. Lau, S.; Chong, E.K.; Yang, X.; Wang, X. Automated pavement crack segmentation using U-Net-Based convolutional neural network. *IEEE Access* **2019**, *8*, 114892–114899. [CrossRef]
33. Escalona, U.; Arce, F.; Zamora, E.; Azuela, J. Fully convolutional networks for automatic pavement crack segmentation. *Comput. Sist.* **2019**, *23*, 451–460. [CrossRef]
34. 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.* **2019**, *21*, 1525–1535. [CrossRef]
35. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic road crack detection using random structured forests. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3434–3445. [CrossRef]
36. Chambon, S.; Moliard, J. Automatic road pavement assessment with image processing: Review and Comparison. *Int. J. Geophys.* **2011**, *2011*, 989354. [CrossRef]

37. Mei, Q.; Mustafa, G. A cost effective solution for pavement crack inspection using cameras and deep neural networks. *Constr. Build. Mater.* **2020**, *256*, 119397. [CrossRef]
38. Lõuk, R.; Riid, A.; Pihlak, R.; Tepljakov, A. Pavement defect segmentation in orthoframes with a pipeline of three convolutional neural networks. *Algorithms* **2020**, *13*, 198. [CrossRef]
39. Shuai, B.; Zuo, Z.; Wang, B.; Wang, G. Scene Segmentation with DAG-Recurrent Neural Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1480–1493. [CrossRef] [PubMed]
40. Li, R.; Zheng, Y.; Zhang, C.; Duan, C.; Wang, B.; Peter, M. Atkinson. ABCNet: Attentive bilateral contextual network for efficient semantic segmentation of Fine-Resolution remotely sensed imagery. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 84–98. [CrossRef]
41. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for image recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
42. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like pure transformer for medical image segmentation. *arXiv* **2021**, arXiv:2105.05537.
43. Wang, W.; Xie, E.; Li, X.; Fan, D.; Song, K.; Liang, D.; Lu, T.; Luo, P.; Shao, L. Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 548–558.
44. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Álvarez, J.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
45. Rakshitha, R.; Srinath, S. A Comprehensive Review on Asphalt Pavement Distress Detection and Assessment based on Artificial Intelligence. In Proceedings of the 2022 IEEE 9th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Prayagraj, India, 2–4 December 2022; Volume 2022, pp. 1–6.
46. Xu, P.; Zhu, X.; Yao, D.; Shi, C.; Qian, G.; Yu, H. Review on intelligent detection and decision-making of asphalt pavement maintenance. *J. Cent. South Univ. Sci. Technol.* **2021**, *52*, 2099–2117.
47. Yuan, J.; Ren, Q.; Jia, C.; Zhang, J.; Fu, J.; Li, M. Automated pixel-level crack detection and quantification using deep convolutional neural networks for structural condition assessment. *Structures* **2024**, *59*, 105780. [CrossRef]
48. Huang, Y. *Pavement Analysis and Design*, 2nd ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2004.
49. Shahin, M. *Pavement Management for Airports, Roads, and Parking Lots*; Springer: New York, NY, USA, 2006; Volume 501.
50. Eldin, N.; Senouci, A. A pavement condition-rating model using backpropagation neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **1995**, *10*, 433–441. [CrossRef]
51. Piryonesi, S.; El-Diraby, T. Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index. *J. Infrastruct. Syst.* **2020**, *26*, 04019036.1–04019036.23. [CrossRef]
52. Shahnazari, H.; Tutunchian, M.; Mashayekhi, M.; Amini, A. Application of soft computing for prediction of pavement condition index. *J. Transp. Eng. -ASCE* **2012**, *138*, 1495–1506. [CrossRef]
53. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual attention network for scene segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
54. Dumoulin, V.; Visin, F. A guide to convolution arithmetic for deep learning. *arXiv* **2016**, arXiv:1603.07285.
55. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 2980–2988.
56. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 565–571.
57. 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] [PubMed]
58. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
59. Chen, L.; Zhu, Y.; Papandreou, G.; Schroff, E.; 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; pp. 801–818.
60. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [CrossRef] [PubMed]
61. Simonyan, K.; Zisserman, A. Very deep convolutional networks for Large-Scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
62. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
63. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
64. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
65. Evans, G.; Groot, M. *Statistics*; Springer: New York, NY, USA, 1997.

66. Leonov, E.; Polbin, A. Numerical Search for a Global Solution in a Two-Mode Economy Model with an Exhaustible Resource of Hydrocarbons. *Math. Models Comput. Simul.* **2022**, *14*, 213–223. [CrossRef]
67. Yamashita, N.; Fukushima, M. On the rate of Convergence of the Levenberg-Marquardt Method. In *Topics in Numerical Analysis: With Special Emphasis on Nonlinear Problems*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 239–249.
68. Kodikara, J.; Sounthararajah, A.; Chen, L. Reimagining Unbound Road Pavement Technology: Integrating Testing, Design, Construction and Performance in the Post-Digital Era. *Transp. Geotech.* **2024**, *47*, 101274. [CrossRef]