buildings

MDPI

Article

Segmentation of Apparent Multi-Defect Images of Concrete Bridges Based on PID Encoder and Multi-Feature Fusion

Yanna Liao, Chaoyang Huang * and Yafang Yin

School of Electronic Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; liaoyn@xupt.edu.cn (Y.L.); yinyf@xupt.edu.cn (Y.Y.)
* Correspondence: chaoy6532@163.com

Abstract: To address the issue of insufficient deep contextual information mining in the semantic segmentation task of multiple defects in concrete bridges, due to the diversity in texture, shape, and scale of the defects as well as significant differences in the background, we propose the Concrete Bridge Apparent Multi-Defect Segmentation Network (PID-MHENet) based on a PID encoder and multi-feature fusion. PID-MHENet consists of a PID encoder, skip connection, and decoder. The PID encoder adopts a multi-branch structure, including an integral branch and a proportional branch with a "thick and long" design principle and a differential branch with a "thin and short" design principle. The PID Aggregation Enhancement (PAE) combines the detail information of the proportional branch and the semantic information of the differential branch to enhance the fusion of contextual information and, at the same time, introduces the self-learning parameters, which can effectively extract the information of the boundary details of the lesions, the texture, and the background differences. The Multi-Feature Fusion Enhancement Decoding Block (MFEDB) in the decoding stage enhances the information and globally fuses the different feature maps introduced by the three-channel skip connection, which improves the segmentation accuracy of the network for the background similarity and the micro-defects. The experimental results show that the mean Pixel accuracy (mPa) and mean Intersection over Union (mIoU) values of PID-MHENet on the concrete bridge multi-defect semantic segmentation dataset improved by 5.17% and 5.46%, respectively, compared to the UNet network.

Keywords: segmentation of bridge defects; deep learning; pid encoder; three-channel skip connection; multi-feature fusion; contextual information

Citation: Liao, Y.; Huang, C.; Yin, Y. Segmentation of Apparent Multi-Defect Images of Concrete Bridges Based on PID Encoder and Multi-Feature Fusion. *Buildings* **2024**, *14*, 1463. https://doi.org/10.3390/buildings14051463

Academic Editor: Humberto Varum

Received: 14 April 2024
Revised: 5 May 2024
Accepted: 14 May 2024
Published: 17 May 2024

1. Introduction

With the rapid expansion of infrastructure development, there is more and more research related to concrete bridges, such as for the ultimate span of steel tube concrete bridges [1], the application of holographic projection technology to bridge construction [2], and the bearing capacity of short concrete columns [3], while the research in this paper is about the apparent defect segmentation of concrete bridges. During the use of concrete bridges, frequent defects include crack, spallation, efflorescence, exposedbar, etc., which will jeopardize the stability and longevity of the bridge structure [4]. Conventionally, the detection of these bridge defects has depended on manual inspections, which are not only time-consuming but also prone to inaccuracies.

Recent advancements in image recognition and deep learning have revolutionized the field of bridge defect detection, particularly through semantic segmentation techniques [5–7]. Notably, UNet and its enhancements have demonstrated remarkable capabilities in the semantic segmentation of specific bridge conditions, such as cracks. For instance, Fan Liu et al. [8] proposed a parallel attention PA-UNet network in order to reduce the interference of noise in crack images. Similarly, Zhao et al. [9] employed a variable-scale convolution kernel to adaptively respond to the size and distribution of road cracks. Furthermore,

Hongguang Chu and team [10] proposed a refined cascade segmentation approach, leveraging transformer and coordinate attention mechanisms to improve the representation of minor crack features. Liang Dong and associates [11] utilized an enhanced ResNet-14 and RS-UNet model for the effective and precise identification of concrete bridge cracks, emphasizing feature and model integration.

While these studies focus predominantly on crack detection, it is imperative to acknowledge that cracks represent merely the initial stage in the progression of bridge defects. The infiltration of corrosive agents from the environment can lead to structural and reinforcement corrosion, eventually manifesting as more severe conditions such as spallation and exposed rebar [12,13]. Therefore, the detection and segmentation of multiple defects in concrete bridges are of critical importance [14].

Current methodologies for the segmentation of multiple defects in concrete bridges face challenges due to the interference among different defect types, the resemblance between the background and the defects, and the defects' diverse and complex textures, shapes, and sizes. This often results in suboptimal segmentation outcomes. This paper introduces several novel contributions to address these challenges:

1. A three-branched PID Encoder is proposed, incorporating the PID (Proportional–Integral–Differential) algorithm, BiSeNet-v2's dual-network regulation, and the concept of feature information decoupling. This encoder features two auxiliary regulation branches and employs a PID aggregation enhancement module (PAE) to emulate PID control in the spatial domain. It introduces self-learning regulation parameters to enhance deep contextual information mining capabilities, thereby offering richer semantic and detailed information for the decoding phase.
2. The Multi-Feature Fusion Enhancement Decoding Block (MFEDB) is introduced in the decoding stage to amalgamate different layers of feature information through three-channel skipping connections. This approach facilitates the network's ability to discern various defect sizes and scenes, thereby enhancing the accuracy of segmentation in cases of background similarity and the delineation of smaller lesions.

2. Related Work

2.1. Multi-Branch Networks

The advent of the BiSeNet series marks a significant evolution in Convolutional Neural Networks (CNNs) [7,15,16], steering toward multi-branch architectures to enhance network performance. Notably, Yu et al. [16] introduced the BiSeNet V2, a bilateral segmentation network featuring two branches of varying depths and feature map dimensions. This design aims to separately parse detailed and contextual information. A novel bilateral bootstrap aggregation layer at the juncture of these branches seamlessly integrates their complementary outputs, thereby augmenting the model's proficiency in depicting complex scenes.

Building on this, Xu et al. [17] innovatively merged the principles of PID (Proportional–Integral–Differential Control) algorithm with the CNN architecture, abstractly addressing the inherent overshoot issues associated with dual-branch networks. These problems, such as the distortion of object boundaries by adjacent pixels and the overshadowing of smaller objects by larger neighboring ones, are critically analyzed through the PID control lens. The proposed three-branch network structure, inspired by PID logic, is posited as an effective remedy for these overshooting challenges.

Despite the divergent analytical lenses applied to their network structures, both examples underscore a common goal: they strive to detach and independently analyze the network's characteristic information. This approach not only enriches the understanding of network behaviors in complex scenarios, but also opens avenues for further refinement of CNN architectures for enhanced performance.

2.2. Feature Information Decoupling

CNNs have achieved remarkable success in the field of computer vision, and one of the important developments is the concept of feature information decoupling. This concept revolutionizes network architecture by segregating feature representations into distinct levels, thereby enriching the network's interpretability of data across various dimensions. The merits of this approach include:

Enhanced representational capacity: by segregating features into detail, context, and semantic categories, CNNs achieve a nuanced understanding of images. This decomposition allows for a more comprehensive capture of image attributes, from granular details to overarching themes, bolstering the network's ability to represent complex structures and contents.

Improved generalizability: decoupling feature information curtails the tendency of overfitting, facilitating a model's applicability to novel, unseen datasets. This stratification of learned features—spanning from details to semantics—ensures that the network's learning is robust and adaptable, enhancing its predictive performance across diverse scenarios.

Superior handling of complex scenes: feature information decoupling empowers CNNs to adeptly navigate intricate image scenes, including occlusions and variable lighting conditions. Detail-oriented features pinpoint subtle nuances, context-based features decode environmental cues, and semantic insights unveil the deeper narrative of images, together providing a holistic scene analysis.

Modular network design: segmenting feature information promotes a modular approach to network design, with each module specializing in a different aspect of information processing. This modularity simplifies debugging, optimization, and scaling of the network, making it more agile and maintainable.

In essence, feature information decoupling transforms CNNs into more versatile, effective tools for a broad spectrum of image processing tasks. This innovation not only elevates the performance of CNNs, but also broadens their applicability.

2.3. PID Algorithm

The PID (Proportional–Integral–Differential) algorithm represents a cornerstone in control theory, encapsulating proportional, integral, and differential components to adjust the controller's output through three distinct operations, as delineated in Equation (1):

$$U(t) = Kp \times e(t) + Ki \times \int e(t)\mathrm{d}t + Kd \times \frac{\mathrm{d}e(t)}{\mathrm{d}t} \quad (1)$$

where $U(t)$ signifies the PID algorithm control output, $e(t)$ denotes the discrepancy between the actual and desired outputs, and Kp, Ki, and Kd are the tuning parameters for the proportional, integral, and differential terms, respectively.

Predominantly utilized across dynamic systems in robotics [18], chemical processing [19], and power systems [20] due to its straightforwardness and efficacy, the PID algorithm has been ingeniously applied to image denoising [21], stochastic gradient descent [22], and numerical optimization [23], yielding considerable enhancements of conventional techniques. Addressing the challenges of multi-defect segmentation in concrete bridges, this work innovatively integrates PID algorithm principles with the concept of feature information decoupling. We introduce a novel three-branch encoder architecture, as illustrated in Figure 1, that effectively segregates feature information. This architecture mimics a PID controller in the spatial domain, where each branch—Proportional (P) for high-resolution detail, Integral (I) for contextual information with extended dependencies, and Differential (D) for deep semantic insight—tackles distinct aspects of image data, thereby enhancing segmentation accuracy and efficiency.

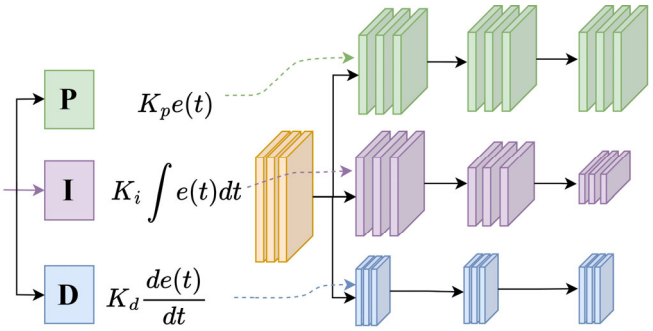

Figure 1. Design conceptualization.

3. Network Architecture Design

The architecture of the PID-MHENet network is illustrated in Figure 2. It employs an encoder–decoder framework comprising a PID encoder, a decoder, and a skip connection organized into five stages (stage 1 to stage 5). The PID encoder features a unique three-branch structure: proportional, integral, and differential branches. Initially, in stage 1, the input image undergoes a convolution operation to extract shallow features. This is followed by downsampling and further convolution in stage 2. Subsequently, the process advances through the PID encoding blocks in stages 3 to 5. Each block is dedicated to encoding different types of information: semantic, contextual, and detail, enhancing the network's ability to capture and integrate varied information streams. The PAE module further augments contextual information by integrating details and semantic insights, thereby enriching feature representation.

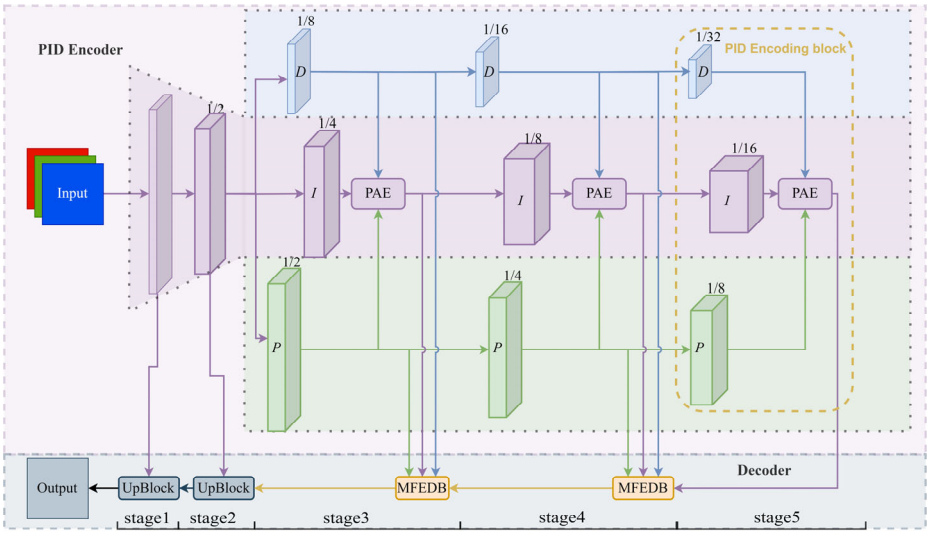

Figure 2. PID-MHENet network structure. The PID-MHENet network consists of a PID encoder and a decoder. The PID encoder consists of a proportional branch (light green part in the figure), an integral branch (light purple part in the figure), and a differential branch (light blue part in the figure).

Different-sized feature layers outputted from the PID encoder are channeled into the decoder via skip connections. This section includes the MFEDB and an upconvolution block (UpBlock). The MFEDB of stage 4 and stage 3 splices and fuses the feature maps of the current stage with the hybrid feature maps of the high level, performs the multi-feature hybrid enhancement to reconstruct the input feature maps, after that it sequentially enters the UpBlock of stage 2 and stage 1, and, finally, outputs the segmentation results to generate the final segmentation map.

3.1. PID Encoder Design

In complex multi-target segmentation scenarios, conventional encoders often extract features that lack variety, risking the loss of context and detail in deeper network layers. The challenge of accurately segmenting multiple anomalies on concrete bridges, especially under conditions of similar defect backgrounds or defect features, underscores the limitations of traditional approaches. These often resort to incorporating attention mechanisms in deeper layers or leveraging shallow, high-resolution information to mitigate segmentation inaccuracies. In response, this paper introduces a novel approach by integrating a PID algorithm abstraction with feature information decoupling to conceive the PID encoder. Displayed in the upper segment of Figure 2, this encoder is structured into proportional, integral, and differential branches, spread across five horizontal stages. Stages 1 and 2 consist of standard encoding blocks, while stages 3 to 5 feature PID encoding blocks. Notably, each feature layer's label in Figure 1 indicates its size relative to the original image, providing a clear quantification of information reduction at each stage. The proportional, integral, and differential branches, alongside the PAE module for enhanced information fusion, are detailed further below.

3.1.1. Three-Branch Network Architecture Design

Integral Branch: illustrated in light purple in Figure 2, the integral branch serves as the core of the PID encoder, primarily focusing on the extraction of contextual information. This branch is vital for interpreting long-distance dependencies among defected pixels. It features a sequence of convolutional kernels (size 3, stride 1, padding 1) and a max pooling layer with a stride of 2, adopting an inverted pyramid shape to ensure gradual progression. The design processes the original 512 × 512 image through four downsampling and five channel-deepening steps, resulting in a 32 × 32 feature map. The branch's specific architecture, excluding the PAE module, is detailed in Table 1.

Table 1. Integral branching structure.

Layer	Integral Branching_Layer	Output_Size
Stage 1	Conv3-64 × 2	512 × 512
Stage 2	MaxPool2d Conv3-128 × 2	256 × 256
Stage 3	MaxPool2d Conv3-256 × 3	128 × 128
Stage 4	MaxPool2d Conv3-512 × 3	64 × 64
Stage 5	MaxPool2d Conv3-512 × 3	32 × 32

Note: Conv3 is a 2D convolution with a convolutional kernel of 3, and MaxPool2d is a maximum pooling layer with a step size of 2.

Proportional Branch: this branch, visualized in light green in Figure 2, operates through a series of singular convolutions at each stage (kernel size of 3, stride of 2, padding of 1), leading to feature layers populated by 256 channels (refer to Table 2). The design of the proportional branch is described as "thick and long"; "thick" refers to its enlarged feature map area, being double the size of the main branch's feature layer, thereby offering a higher resolution, while "long" indicates an increased channel count, enabling the branch to encapsulate a richer spatial context. The proportional branch can capture richer spatial details in the image, such as the details of cracks and the edges of exposedbars, so that the network can more accurately localize the lesions, a process that is crucial for the detail resolution and boundary detection of defects.

Table 2. Proportional branching structure.

Layer	Proportional Branching_Layer	Output_Size
Stage 3	Conv3-256	256 × 256
Stage 4	Conv3-256	128 × 128
Stage 5	Conv3-256	64 × 64

Note: Conv3 is a 2D convolution with a convolutional kernel of 3.

Differential Branch: illustrated in light blue in Figure 2, each stage in this branch adopts a convolutional kernel size of 3, a stride of 1, and a padding of 1, adhering to a "thin and short" principle. "Short" reflects a reduced channel count in the feature layer, while "thin" indicates a diminished feature layer area, the latter being half the size of the main branch's feature layer, as illustrated in Table 3. Distinguished by its frequent convolution and pooling operations, the differential branch boasts an extended receptive field, crucial for the extraction of broader semantic content. This includes insights into the overarching structural integrity of bridges and the differentiation of defect types, thereby facilitating a more accurate defect classification. This aspect is particularly beneficial for identifying surface-level conditions, such as efflorescence and spallation, which manifest as flaky deteriorations.

Table 3. Differential branching structure.

Layer	Differential Branching_Layer	Output_Size
Stage 3	MaxPool2d Conv3-64 × 2	64 × 64
Stage 4	MaxPool2d Conv3-64 × 2	32 × 32
Stage 5	MaxPool2d Conv3-64 × 2	16 × 16

Note: Conv3 is a 2D convolution with a convolutional kernel of 3, and MaxPool2d is a maximum pooling layer with a step size of 2.

3.1.2. PAE Module Design

The PID algorithm's application within early-stage industrial control often necessitates manual parameter adjustments in response to feedback, a process exemplified in scenarios such as rotating inverted pendulums and balance cars. Addressing this, our approach integrates the proportional–integral–Differential parameters (kp, ki, kd) as learnable components within the CNN architecture, allowing for automatic adjustment in relation to network loss. Through Equation (2) during forward propagation, kp, ki, and kd are transformed into kp', ki', and kd', respectively, ensuring parameter tuning remains within an optimal range to enhance feature interpretation across the three branches.

$$Kn' = \frac{e^{kn}}{e^{kp} + e^{ki} + e^{kd}}, n = p, i, d \tag{2}$$

Figure 3 unveils the PAE module, innovatively applying the PID algorithm's three-channel regulation concept to bolster feature fusion. This module enriches contextual understanding by leveraging the proportional branch's detailed information and the differential branch's semantic insights, addressing the original network's limitation in deep contextual information extraction. The process begins with the detail information feature map (f_1) undergoing downsampling and a 3 × 3 convolution for preliminary extraction. Similarly, the semantic information feature map (f_3) is processed through a 3 × 3 convolution and upsampling. The contextual information feature map (f_2) is then refined via a Sigmoid function. Subsequently, each branch's feature map is scaled by the corresponding kp', ki', and kd' tuning parameters, culminating in a composite feature map through

element-wise summation and fusion with the original contextual data, resulting in the output F_{PAE}.

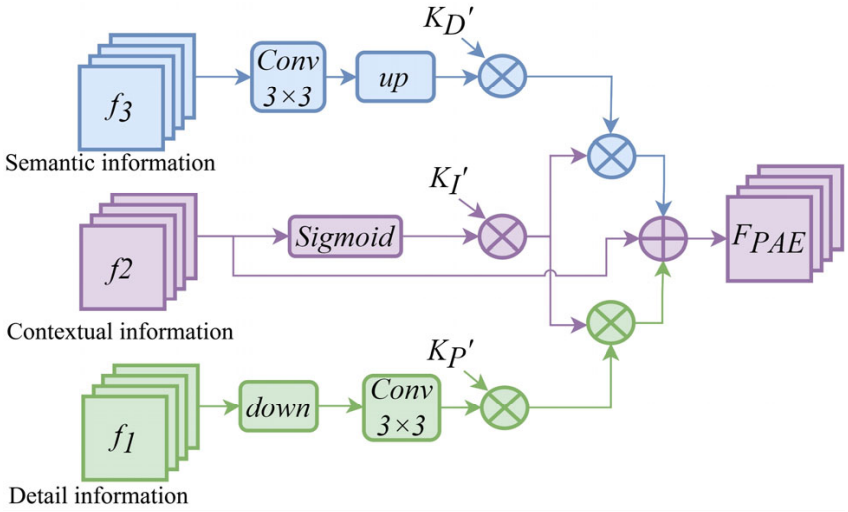

Figure 3. PAE module structure. The notation within the figure includes "⊕" for element-wise addition, and "⊗" for element-wise multiplication.

Incorporating the PAE module significantly deepens the network's insight into defect-related boundary nuances, texture variations, brightness levels, and background disparities. This enhancement is instrumental in elevating the extraction of contextual features.

3.2. Decoder Design

The PID-MHENet network's deeper layers utilize a three-channel skip connection, integrating the output feature maps from the proportional, integral, and differential branches into the decoder. This integration aims to amalgamate a broader spectrum of features with varying resolutions. Specifically, the decoder's stage 3 and stage 4 are crafted with the Multi-Feature Enhanced Decoding Block (MFEDB), optimizing the capture of layered feature information. This architecture enables the network to discern lesions and scenes of diverse sizes more accurately, minimizing information loss. The MFEDB processes its inputs for filtration, fusion, and enhancement before undergoing sequential upsampling and decoding via UpBlocks in stage 2 and stage 1, culminating in the generation of the final segmented map.

3.2.1. MFEDB Design

The introduction of feature maps with disparate resolutions through a three-channel skip connection inadvertently introduces excess redundant information. Traditional decoding blocks, which primarily rely on direct upsampling without integrating features across various scales and depths, often falter in accurately segmenting pixels amid similar backgrounds, edges of defects, and minute lesions. As shown in Figure 4, the global feature fusion strategy of MFEDB can segment the above pixels more effectively. It employs an Efficient Multiscale Attention (EMA) [24] module for the downsampled scale branch feature map (f_1), producing an enhanced detail feature map (F_{ch1}). Concurrently, the differential branch feature map (F_3) undergoes channel splicing post-average pooling and 3 × 3 convolution, followed by maximum pooling and further 3 × 3 convolution, leading to the upsampled, enhanced semantic feature map (F_{ch3}), as delineated in Equation (3).

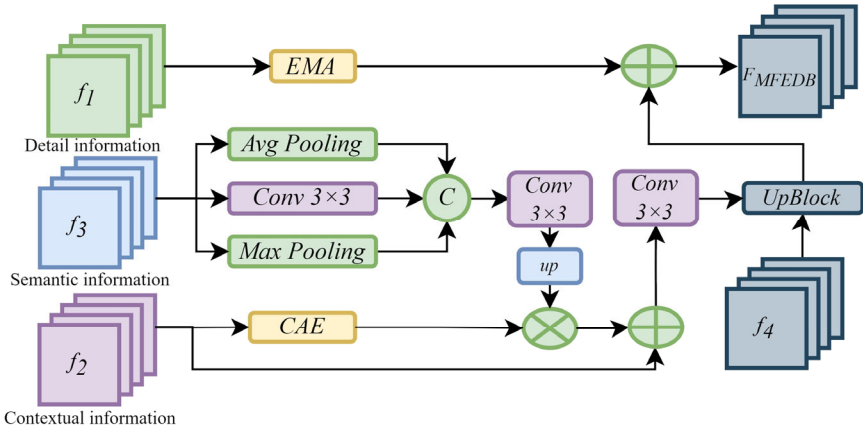

Figure 4. MFEDB module structure. The notation within the figure includes "⊕" for element-wise addition, "⊗" for element-wise multiplication, and "©" for feature map concatenation. EMA and CAE represent the corresponding modules.

Subsequent operations include a pixel-level dot product between F_{ch3} and the integral branch feature map (f_2), processed via the CAE module, with an addition of f_2 to fuse semantic with contextual information. A 3 × 3 convolution then extracts the amalgamated feature set to produce f_m (as per Equation (4)). The final enhanced feature map (F_{MFEDB}) emerges from upsampling, splicing, and decoding operations with f_4 via UpBlock, integrated with F_{ch1} to refine the detailed feature map with additional information, encapsulating contextual, semantic, and detail facets (Equation (5)). The structure of UpBlock is shown in Figure 5.

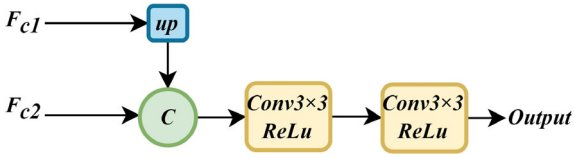

Figure 5. Upsampling decoder block structure. Within the figure, the notation "©" indicates feature map concatenation.

The EMA module within MFEDB diverges from conventional coordinate attention [25] by facilitating inter-channel information encoding and spatial structure retention within the channels, markedly bolstering detail fidelity. This enhancement is pivotal for segmenting elongated cracks and minuscule defects, reducing misdiagnoses and omissions.

$$F_{ch3} = \mathrm{Up}(\mathrm{Concat}(\mathrm{Avgpooling}(f_3), \mathrm{Maxpooling}(f_3), \mathrm{Conv3} \times 3(f_3))) \quad (3)$$

$$f_m = \mathrm{Conv3} \times 3(F_{ch3} \otimes \mathrm{CAE}(f_2) \oplus f_2) \quad (4)$$

$$F_{MFEDB} = \mathrm{UpBlock}(f_m, f_4) \oplus F_{ch1} \quad (5)$$

3.2.2. CAE Module Design

The CAE module introduces a focused strategy to accentuate critical feature information within the contextual feature map, leveraging global spatial context to amplify feature representation. Incorporated atop the foundational GCNet module [26], the CAE introduces an additional pathway (highlighted in the dashed box in Figure 6) for pixel informativeness learning, merging local and global features while clarifying informational overlap. The spatial context of each pixel is aggregated through Equation (6), with P_i and Q_i representing the input and output feature mappings for layer i, incorporating N_i pixels. The transformation matrices w_k and w_v project these mappings, while a_i—a reweighting matrix—modulates the aggregated spatial context intensity per pixel, utilizing softmax for

linear transformation of the Pi matrix (Equation (7)). Practical implementation involves 1×1 convolutions for matrix mapping, followed by pixel-level dot product operations and summation with original feature maps, facilitating a nuanced feature map amalgamation.

$$Q_i^j = P_i^j + a_i^j P_i^j \cdot \sum_{j=1}^{N_i} \left[\frac{\exp(w_k P_i^j)}{\sum_{m=1}^{N_i} \exp(w_k P_i^m)} \cdot w_v P_i^j \right] \tag{6}$$

$$a_i^j = \frac{\exp(w_a P_i^j)}{\sum_{n=1}^{N_i} \exp(w_a P_i^n)} \tag{7}$$

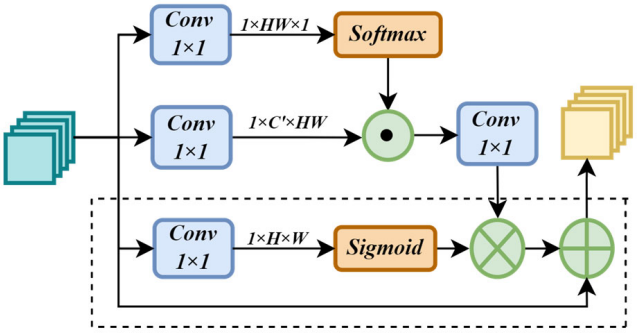

Figure 6. CAE module structure. The notation within the figure includes "⊕" for element-wise addition, "⊗" for element-wise multiplication, and "⊙" indicates the matrix multiplication.

4. Experimental Results and Analysis

4.1. Experimental Environment and Dataset

All the experiments in this paper were done based on a Linux-based server with the hardware configuration of 4 Intel(R) Xeon(R) Gold 5120 CPUs, 4 GPUs of TITAN RTX 24 GB. The software environment was Python 3.7, Pytorch 1.81. The model parameters were set as shown in Table 4.

Table 4. Hyperparameter configuration.

Parameter	Value	Parameter	Value
Learning rate	data	Epochs	100
Optimizer	SGD	Batch size	6
Weight-decay	0.0001	Momentum	0.9
Loss	CrossEntropyLoss	-	-

In this paper, a multi-defect semantic segmentation dataset of concrete bridges containing four types of apparent defects, namely crack, efflorescence, spallation, and exposedbar, is established. The original images for the dataset came from public datasets [27–29]. They were then annotated using the Pixel Annotation Tool, a segmentation dataset annotation software which ultimately yielded 2842 images with identified defects. The dataset was randomly split into training and testing sets in a 9:1 ratio with a fixed random seed to ensure consistency across different network experiments. The dataset pictures and truth labels are shown in Figure 7, in which the spallation defect, which is the result of concrete peeling off, may also form the exposedbar in the end, making a variety of defects in the same location. The efflorescence defect is shown as a white color, flaky distribution. The exposedbar defect is shown as the exposed steel reinforcement or corrosion, while the crack defect is mostly distributed in the form of irregular lines, with a thin and slender distribution accounting for the total area of the picture being relatively small.

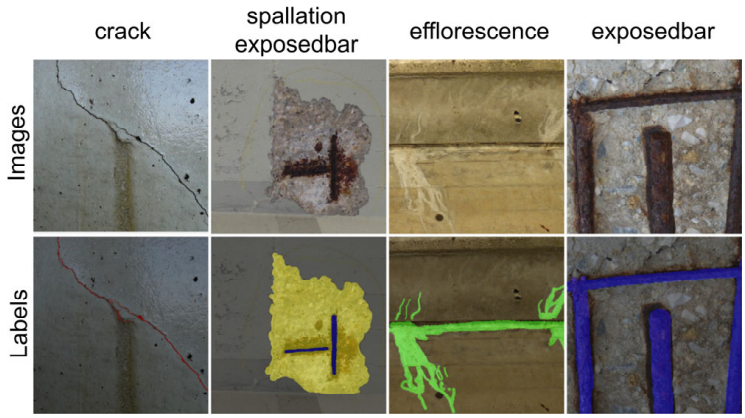

Figure 7. Defect pictures and labels.

4.2. Evaluation Metrics

This paper utilizes three metrics to evaluate the model's performance and prediction accuracy: mean Pixel accuracy (mPa), Intersection over Union (IoU), and mean Intersection over Union (mIoU). mPa is defined as the average accuracy across all categories, calculated by the ratio of correctly classified pixels to the total pixels in each category, as detailed in Equation (8). IoU measures the overlap between the predicted and true areas of a category, presented in Equation (9). mIoU, outlined in Equation (10), represents the average IoU across all categories. These metrics are essential for assessing the precision of the model's predictions.

$$\text{mPA} = \frac{1}{\text{C}}\sum_{i=1}^{\text{C}}\frac{TP_i}{TP_i + FN_i} \tag{8}$$

$$\text{IoU} = \frac{TP}{TP + FN + FP} \tag{9}$$

$$\text{mIoU} = \frac{1}{\text{C}}\sum_{i=1}^{\text{C}}\frac{TPi}{TP_i + FN_i + FP_i} \tag{10}$$

where C is the number of categories, TP is the number of true positives, TP_i is the number of true positives for the ith category, FN is the number of false negatives, FN_i is the number of false negatives for the ith category, FP is the number of false positives, and FP_i is the number of false positives for the ith category.

4.3. Ablation Experiments

In order to verify the effectiveness of each module of the PID-MHENet network, the following ablation experiments were done on the concrete bridge multi-defect semantic segmentation dataset.

4.3.1. Comparison of Different Encoders

The comparison test of different encoders on the multi-defect semantic segmentation dataset of concrete bridges is shown in Table 5, which shows that the PID encoder has the highest mIoU value and mPa value, and the feature mining ability is more powerful.

Table 5. Comparative encoder experiments results.

Method	Encoder	mIoU/%	mPa/%
UNet	VGG16	60.27%	67.63%
UNet	ResNet-50	53.39%	60.13%
UNet	PID Encodert	64.20%	71.44%

4.3.2. EMA and CAE Ablation Experiments

In order to verify the validity of EMA and CAE, ablation experiments were performed, the base network being UNet, the experimental network not containing the PID encoder, and the f_1 and f_3 feature maps in the decoder being all referenced to f_2. The experimental results are shown in Table 6.

Table 6. EMA and CAE ablation experiment results.

EMA	CAE	mIoU/%	mPa/%	↑mIoU/%
×	×	60.27%	67.63%	-
√	×	61.23%	68.13%	0.96%
×	√	61.64%	68.83%	1.37%

Note: where "√" indicates that this module is included in the model, "×" indicates that this module is not included in the model, and "↑mIoU" indicates the degree of improvement of the current network over the UNet network mIoU value.

As it can be seen from Table 6, the mPa values improved by 0.5% and 1.2% and the mIoU values improved by 0.96% and 1.37% in the corresponding positions of the decoder when the EMA and CAE modules were added to the decoder, respectively.

4.3.3. PID Encoder and MFEDB Ablation Experiments

In order to verify the effectiveness of the PID encoder and the MFEDB module, the ablation experiments were carried out by different combinations, the base network being UNet (Experiment 1). The results of the experiments are shown in Table 7 below.

Table 7. PID encoder and MFEDB ablation experiments results.

Serial Number	PID Encoder	MFEDB	IoU/%				mIoU/%	mPa/%	↑mIoU/%
			Efflorescence	Spallation	Exposedbar	Crack			
1	×	×	45.03%	62.39%	68.39%	29.36%	60.27%	67.63%	-
2	√	×	50.19%	66.74 %	71.49%	35.97%	64.20%	71.44%	↑3.93%
3	×	√	51.74%	62.57%	69.18%	33.24%	62.64%	70.23%	↑2.37%
4	√	√	55.27%	66.99%	72.48%	37.05%	65.73%	72.8%	↑5.46%

Note: where "√" indicates that this module is included in the model, "×" indicates that this module is not included in the model, and "↑mIoU" indicates the degree of improvement of the current network over the UNet network mIoU value.

As seen in Table 7, Experiments 2 and 3 replaced and added the PID encoder and MFEDB to the UNet network, respectively, the mPa values improved by 3.81% and 2.6%, and the mIoU values improved by 3.93% and 2.37%, respectively. Experiment 4 replaced the encoder of the original network with a PID encoder while adding the MFEDB module (which is the PID-MHENet network in this paper), with mPa and mIoU improving by 5.17% and 5.46%, respectively. The above proves the effectiveness of both PID encoder and MFEDB.

The following evaluates the performance of each module through a confusion matrix heatmap, as shown in Figure 8. In the heatmap matrix, each row represents the predicted proportion of pixels for various types of defects, and each column represents the actual proportion of various defects in the data. The larger the numbers on the diagonal, the more pixels are accurately segmented. It can be observed that Experiments 2 to 4 have larger numbers on the diagonal than Experiment 1, with Experiment 4 having the largest diagonal numbers.

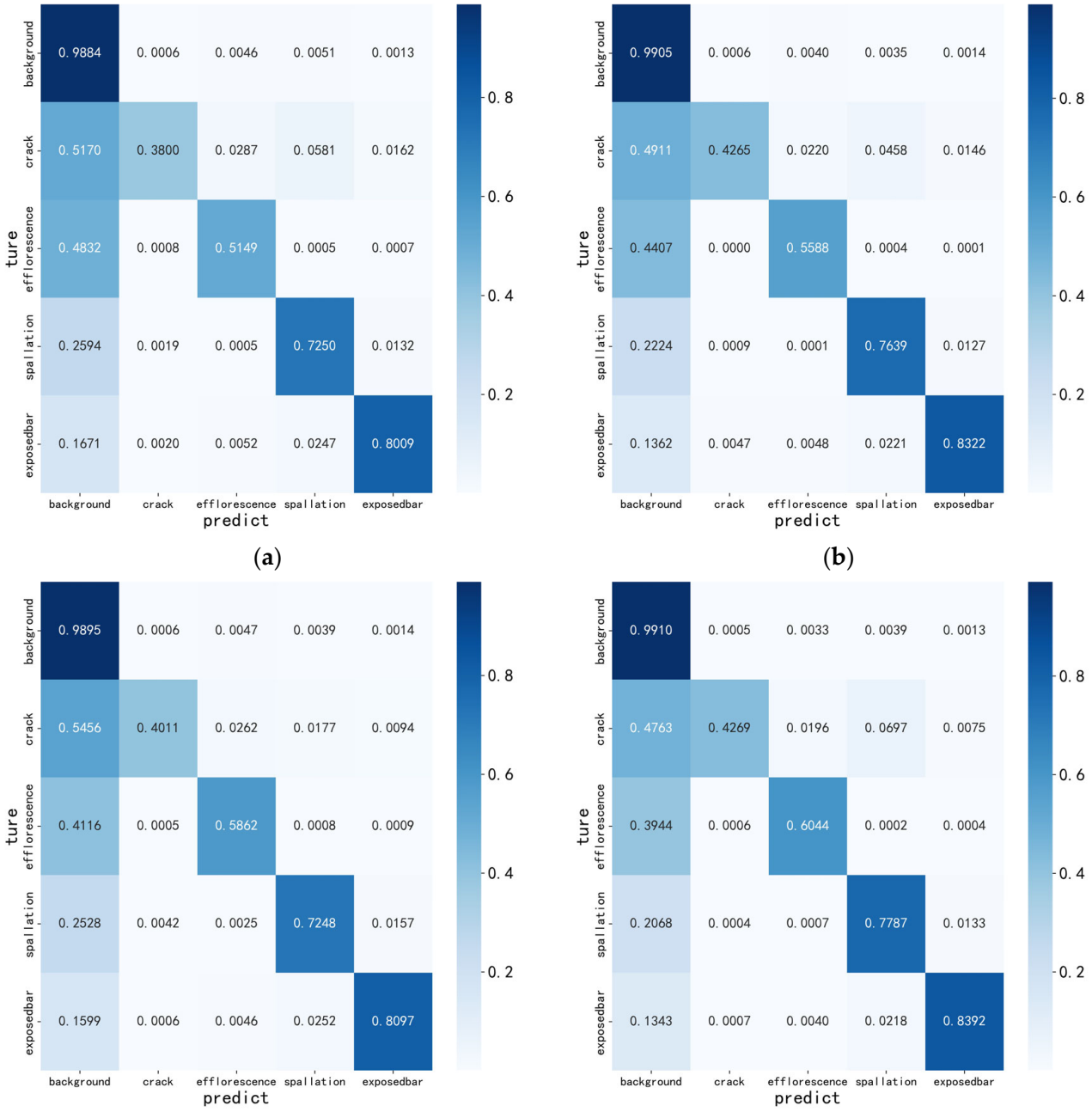

Figure 8. Comparison of confusion matrix visualizations. (**a**) Experiment 1 confusion matrix; (**b**) Experiment 2 confusion matrix; (**c**) Experiment 3 confusion matrix; (**d**) Experiment 4 confusion matrix.

On the other hand, the mIoU and loss values of each round of test set were recorded during model training, as shown in Figure 9. It can be seen that the mIoU value and loss value fluctuated up and down between epochs of 0–50, a phenomenon which may be caused by the fact that the network does not learn enough prior knowledge during the pre-training. As the training progressed, the advantages of the PID encoder and MHEDB module started to emerge after 50 epochs. After 80 epochs, the network mIoU value and loss value area stabilized, and the PID-MHENet network attained the highest mIoU value and the lowest loss value, further proving the effectiveness of each module of the network and the excellent segmentation performance of the network.

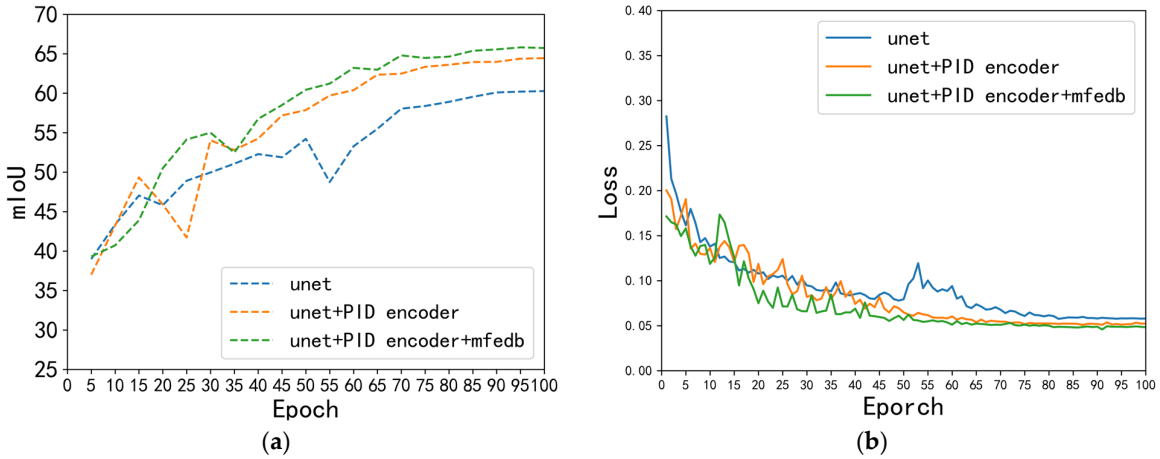

Figure 9. mIoU curve and loss curve. (**a**) mIoU comparison curve; (**b**) loss comparison curve.

4.4. Comparative Experiments

In order to verify the performance advantages of the PID-MHENet network proposed in this paper, the IoU, mIoU, and mPa values were used as evaluation indices, and the PID-MHENet was compared with, Segformer, HRNet, DeepLabv3+, UNet3+, PSPNet, and UNet networks for comparison experiments. The experimental results are shown in Table 8.

Table 8. Comparative experiments results.

Algorithm	IoU/%				mIoU/%	mPa/%
	Efflorescence	Spallation	Exposedbar	Crack		
Segformer	46.44%	54.41%	58.77%	17.48%	54.69%	60.16%
HRNet	39.2%	50.39%	48.27%	21.31%	50.93%	56.18%
DeepLabv3+	35.79%	49.1%	54.23%	24.94%	51.87%	58.18%
UNet3+	41.64%	58.91%	55.83%	25.78%	56.19%	61.57%
PSPNet	29.4%	42.59%	35.31%	13.9%	43.12%	49.06%
UNet	45.03%	62.39%	68.39%	29.36%	60.27%	67.63%
PID-MHENet	55.27%	66.99%	72.48%	37.05%	65.73%	72.8%

In Table 8, we observe significant enhancements in Intersection over Union (IoU) values for four distinct defects—efflorescence, spallation, exposedbar, and crack—when employing the PID-MHENet over the UNet network. Specifically, IoU improvements were 10.24%, 4.6%, 4.09%, and 7.69%, respectively, with efflorescence showing the most notable increase. Additionally, mean Pixel accuracy (mPa) and mean IoU (mIoU) values increased by 5.17% and 5.46%, respectively, against the UNet benchmark. Comparatively, against the Segformer network introduced in 2021, the defects' IoU values saw enhancements of 8.83%, 12.58%, 13.31%, and 19.57%, respectively, with the crack defect attaining the greatest improvement. The mPa and mIoU metrics also improved by 12.64% and 11.04%, respectively.

The visualization results in Figure 10 elucidate PID-MHENet's superiority in identifying less apparent segmented cracks, yielding more continuous detections. It showcases enhanced edge localization for spallation and fewer misidentifications within the exposedbar defect's central pixel points. Particularly for efflorescence, PID-MHENet achieved closer edge conformity to actual conditions and more precise localization. Furthermore, it reduced erroneous pixel detections as background for the exposedbar defect more efficiently than the comparator networks, underscoring its improved accuracy in distinguishing between defects and non-defects.

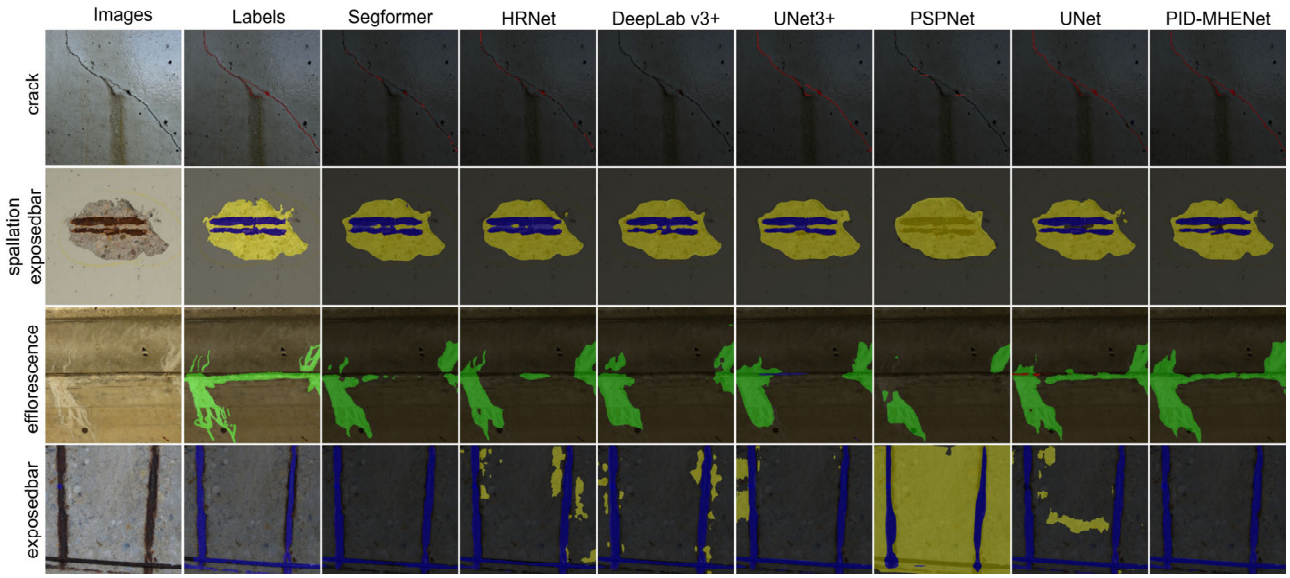

Figure 10. Comparison of experimental visualization results.

Combined with the analysis in Table 8 and Figure 10, the PID-MHENet network in the field of multi-defect detection segmentation of concrete bridges exhibited fewer pixel points of leakage, misdetection, and multi-detection for the detection segmentation of multi-defects, and the edge localization was more accurate and more closely fitted the actual defect. It is fully proved that the PID-MHENet network proposed in this paper is able to more fully explore the contextual information of defect features in the field of multi-defect semantic segmentation of concrete bridges, and, to a certain extent, it overcomes the effects of mutual interference among different defects, the diversity, complexity, and large background differences in the texture, shape, and scale of defects, and improves the segmentation accuracy and mIoU value.

4.5. GAPs384 Dataset Comparison Experiments

In order to further verify the robustness of the PID-MHENet network, the PID-MHENet network was tested for comparison by using the publicly available dataset GAPs384 [30], where GAPs384 is a single crack dataset for highways. As it can be seen from Table 9, PID-MHENet also achieved better results with respect to the segmentation of single cracks, with an IoU improvement of 2.05% compared to the UNet network with ResNet50 as the encoder.

Table 9. GAPs384 dataset comparison experiments results.

Method	Encoder	IoU%
SegFormer	Mit B0	41.67%
HRNet	HRNetv2-W18	41.62%
DeepLab V3+	ResNet50	40.92%
UNet	ResNet50	44.28%
UNet	VGG16	43.66%
PID-MHENet	PID Encoder	46.33%

4.6. Discussion

Advantages of PID-MHENet: this section of the discussion elaborates on the comparative performance advantages of the PID-MHENet which were initially presented in Section 4.4. PID-MHENet demonstrates significant improvements in IoU values across multiple types of defects in concrete bridges, notably outperforming established networks such as UNet and Seaformer. For instance, the enhancement in IoU for efflorescence and cracks are particularly notable, underscoring the model's superior capability in handling

complex defect delineations more accurately. Moreover, the advancements in mPa and mIoU not only highlight the enhanced accuracy of PID-MHENet, but also its reliability in multi-defect detection under varied conditions.

Theoretical contributions: this model's methodology leverages a novel approach to integrating deep learning techniques with PID principles, helping refine the detection and segmentation processes. The incorporation of multi-scale strategies enables the PID-MHENet to achieve more refined edge localization and a lower degree of misidentification, thereby enhancing the quality of the segmentation output. Such theoretical advancements contribute significantly to the field of semantic segmentation, particularly in the complex domain of infrastructure evaluation.

Practical implications: in practical terms, the PID-MHENet's ability to accurately segment multiple defects with high precision can lead to more reliable assessments of infrastructure integrity, thus aiding maintenance decisions and the prioritization of repairs. This could potentially result in cost savings and extended lifespan of concrete structures by addressing the most critical defects early on.

Future work: our initial plan is to deploy the model in an embedded device and later piggyback the embedded device on mobile devices such as drones and inspection carts for inspection to help the maintenance team plan and implement necessary repairs based on the detected defects. In addition, further research could explore the integration of PID-MHENet with other sensor data, such as that from drones or robots, to enhance detection capabilities and operational scalability. In practical deployments, we recommend that the angle and distance at which the mobile device takes pictures be kept at a constant value so that the data can be analyzed at a later stage.

Comparative analysis: compared to similar studies, PID-MHENet not only pushes the boundaries of defect detection accuracy, but also provides a new framework for handling diverse, complex scenarios in concrete bridge analysis. It surpasses traditional methods that often struggle with the intricacies of multi-defect segmentation and offers a more nuanced understanding of defect interactions and their impacts on the segmentation process.

5. Conclusions

In this paper, a PID-MHENet network applied to the segmentation of apparent multi-defect in concrete bridges is proposed, and a PID encoder (proportional branch, integral branch, differential branch) is constructed in which the PAE module further enhances the network's ability to extract contextual information and is able to better mine for information such as the boundary details, texture, luminance, and contextual differences of the defect. The MFEDB in the decoder performs global fusion and enhancement of the deep layers of the network for different layers of feature information at different resolutions, while retaining the contextual information at different scales. The experimental results show that PID-MHENet achieves better results with respect to the complex task of semantic segmentation of concrete bridges with multiple defects, providing a new way of thinking to minimize the effects of multiple defects such as interference.

Author Contributions: All of the authors extensively contributed to the work. Conceptualization, Y.L. and C.H.; methodology, Y.L. and C.H.; validation, Y.L., C.H. and Y.Y.; investigation, Y.L. and C.H.; writing–original draft preparation, C.H.; writing–review and editing, C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Research and Development Program of Shaanxi Province—International Science and Technology Co-operation Program Project (No. 2020KW-001), Contract for Xi'an Municipal Science and Technology Plan Project—Xi'an City Strong Foundation Innovation Plan (No. 21XJZZ0074), and the Key Project of Graduate Student Innovation Fund at Xi'an University of Posts and Telecom-munications (No. CXJJZL2023013).

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy.

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

References

1. Wu, Y.X.; Wang, X.C.; Fan, Y.; Shi, J.; Luo, C.; Wang, X. A Study on the Ultimate Span of a Concrete-Filled Steel Tube Arch Bridge. *Buildings* **2024**, *14*, 896. [CrossRef]
2. Wu, J.L.; Zhu, J.; Zhang, J.B.; Dang, P.; Li, W.L.; Guo, Y.K.; Fu, F.; Lai, J.B.; You, J.G.; Xie, Y.K.; et al. A dynamic holographic modelling method of digital twin scenes for bridge construction. *Int. J. Digit. Earth* **2023**, *16*, 2404–2425. [CrossRef]
3. Chepurnenko, A.; Turina, V.; Akopyan, V. Artificial Neural Network Models for Determining the Load-Bearing Capacity of Eccentrically Compressed Short Concrete-Filled Steel Tubular Columns. *CivilEng* **2024**, *5*, 150–168. [CrossRef]
4. Zhang, J.C.; Liu, S.C.; Tian, X.S. Review on disease detection technology for ballastless track concrete structure. *J. Beijing Jiaotong Univ.* **2022**, *46*, 80–92.
5. Wan, Q.; Huang, Z.L.; Lu, J.C.; Yu, G.; Zhang, L. Seaformer: Squeeze-enhanced axial transformer for mobile semantic segmentation. *arXiv* **2023**, arXiv:2301.13156.
6. Xie, E.Z.; Wang, W.H.; Yu, Z.D.; 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.
7. Tsai, T.H.; Tseng, Y.W. BiSeNet V3: Bilateral segmentation network with coordinate attention for real-time semantic segmentation. *Neurocomputing* **2023**, *532*, 33–42. [CrossRef]
8. Liu, F.; Wang, J.F.; Chen, Z.Y.; Xu, F. Parallel attention based UNet for crack detection. *J. Comput. Res. Dev.* **2021**, *58*, 1718–1726.
9. Zhao, Z.H.; He, P.; Hao, Z.Y. Variable-Scale VS-UNet Model for Road Crack Detection [EB/OL]. Journal of Hunan University (Natural Sciences). Available online: http://kns.cnki.net/kcms/detail/43.1061.N.20230905.0915.002.html (accessed on 12 April 2024).
10. Zhu, H.G.; Yuan, H.Q.; Long, L.Z.; Deng, L. A transformer-based cascade method for segmenting bridge cracks from high-resolution images. *China J. Highw. Transp.* **2024**, *37*, 65–76.
11. Liang, D.; Li, Y.J.; Zhang, S.J. Identification of cracks in concrete bridges through fusing improved ResNet-14 and RS-Unet models. *J. Beijing Jiaotong Univ.* **2023**, *47*, 10–18.
12. Deng, Z.L.; Luo, R.Z.; Fei, Y.; Li, H.F. Airport pavement crack detection based on FE-UNet. *J. Optoelectron. Laser.* **2023**, *34*, 34–42.
13. Zhang, W.G.; Zhong, J.T.; Hu, Y.J.; Ma, T.; Zhu, J.Q.; He, L. Extraction and quantification of pavement alligator crack morphology based on VGG16-UNet semantic segmentation model. *J. Traffic Transp. Eng.* **2023**, *23*, 166–182.
14. Peng, Y.N.; Liu, M.; Wan, Z.; Jiang, W.B.; He, W.X.; Wang, Y.N. A dual deep network based on the improved YOLO for fast bridge surface defect detection. *Acta Autom. Sin.* **2022**, *48*, 1018–1032.
15. Yu, C.Q.; Wang, J.B.; Peng, C.; Gao, C.X.; Yu, G.; Sang, N. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 325–341.
16. Yu, C.Q.; Gao, C.X.; Wang, J.B.; Yu, G.; Shen, C.H.; Sang, N. Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation. *Int. J. Comput. Vis.* **2021**, *129*, 3051–3068. [CrossRef]
17. Xu, J.C.; Xiog, Z.X.; Bhattacharyya, S.P. PIDNet: A real-time semantic segmentation network inspired by PID controllers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 19529–19539.
18. Alavandar, S.; Jain, T.; Nigam, M.J. Tuning of PID controller based on a multiobjective genetic algorithm applied to a robotic manipulator. *Expert Syst. Appl.* **2012**, *39*, 8968–8974.
19. Jayachitra, A.; Vinodha, R. Genetic algorithm based PID controller tuning approach for continuous stirred tank reactor. *Adv. Artif. Intell.* **2015**, *2014*, 791230. [CrossRef]
20. Khodabakhshian, A.; Hooshmand, R. A new PID controller design for automatic generation control of hydro power systems. *Int. J. Electr. Power Energy Syst.* **2010**, *32*, 375–382. [CrossRef]
21. Ma, R.J.; Li, S.Y.; Zhang, B.; Li, Z.M. Towards fast and robust real image denoising with attentive neural network and PID controller. *IEEE Trans. Multimed.* **2021**, *24*, 2366–2377. [CrossRef]
22. An, W.; Wang, H.; Sun, Q.; Xu, J.; Dai, Q.; Zhang, L. A PID controller approach for stochastic optimization of deep networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8522–8531.
23. Xu, J.; Bhattacharyya, S.P. A pid controller architecture inspired enhancement to the pso algorithm. In Proceedings of the Future of Information and Communication Conference, San Francisco, CA, USA, 3–4 March 2022; pp. 587–603.
24. Ouyang, D.L.; He, S.; Zhang, G.Z.; Luo, M.Z.; Guo, H.Y.; Zhan, J.; Huang, Z.J. Efficient mult-scale attention module with cross spatial learning. In Proceedings of the ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
25. Hou, Q.B.; Zhou, D.Q.; Feng, J.S. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 13708–13717.
26. Liu, Y.; Li, H.F.; Hu, C.; Luo, S.; Luo, Y.; Chen, C.H. Learning to aggregate Multi-Scale context for instance segmentation in remote sensing images. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, 1–15. [CrossRef] [PubMed]
27. Mundt, M.; Majumder, S.; Murali, S.; Panetsos, P.; Ramesh, V. Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11196–11205.

28. Hüthwohl, P.; Lu, R.; Brilakis, I. Multi-classifier for RC bridge defects. *Autom. Constr.* **2019**, *105*, 102824. [CrossRef]
29. Bianchi, E.; Hebdon, M. Concrete Crack Conglomerate Dataset. Available online: https://data.lib.vt.edu/articles/dataset/Concrete_Crack_Conglomerate_Dataset/16625056/1 (accessed on 4 May 2024).
30. Eisenbach, M.; Stricker, R.; Seichter, D.; Amende, K.; Debes, K.; Sesselmann, M.; Ebersbach, D.; Stoeckert, U.; Gross, H.-M. How to get pavement distress detection ready for deep learning? A systematic approach. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 2039–2047.