Article

Multi-Type Structural Damage Image Segmentation via Dual-Stage Optimization-Based Few-Shot Learning

Jiwei Zhong [1,2,†], Yunlei Fan [3,†], Xungang Zhao [1], Qiang Zhou [1] and Yang Xu [3,4,5,*]

1   National Key Laboratory of Bridge Intelligence and Green Construction, Wuhan 430034, China
2   School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430070, China
3   School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China
4   Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150090, China
5   Key Lab of Structures Dynamics Behavior and Control of the Ministry of Education, Harbin Institute of Technology, Harbin 150090, China
*   Correspondence: xyce@hit.edu.cn
†   These authors contributed equally to this work.

Abstract: The timely and accurate recognition of multi-type structural surface damage (e.g., cracks, spalling, corrosion, etc.) is vital for ensuring the structural safety and service performance of civil infrastructure and for accomplishing the intelligent maintenance of smart cities. Deep learning and computer vision have made profound impacts on automatic structural damage recognition using nondestructive test techniques, especially non-contact vision-based algorithms. However, the recognition accuracy highly depends on the training data volume and damage completeness in the conventional supervised learning pipeline, which significantly limits the model performance under actual application scenarios; the model performance and stability for multi-type structural damage categories are still challenging. To address the above issues, this study proposes a dual-stage optimization-based few-shot learning segmentation method using only a few images with supervised information for multi-type structural damage recognition. A dual-stage optimization paradigm is established encompassing an internal network optimization based on meta-task and an external meta-learning machine optimization based on meta-batch. The underlying image features pertinent to various structural damage types are learned as prior knowledge to expedite adaptability across diverse damage categories via only a few samples. Furthermore, a mathematical framework of optimization-based few-shot learning is formulated to intuitively express the perception mechanism. Comparative experiments are conducted to verify the effectiveness and necessity of the proposed method on a small-scale multi-type structural damage image set. The results show that the proposed method could achieve higher segmentation accuracies for various types of structural damage than directly training the original image segmentation network. In addition, the generalization ability for the unseen structural damage category is also validated. The proposed method provides an effective solution to achieve image-based structural damage recognition with high accuracy and robustness for bridges and buildings, which assists the unmanned intelligent inspection of civil infrastructure using drones and robotics in smart cities.

Keywords: structural health diagnosis; multi-type damage segmentation; few-shot learning; meta learning; limited annotated images

1. Introduction

The civil infrastructure is inevitably impacted by a variety of complicated factors throughout the life-long service period, leading to the initiation, propagation, and accumulation of multi-type structural damage. The timely and accurate recognition of multi-type structural surface damage (e.g., cracks, spalling, corrosion, etc.) is vital to ensure the

structural safety and service performance of civil infrastructure and accomplish intelligent maintenance of smart cities. Structural health monitoring and damage identification serve as pivotal techniques for condition assessment and maintenance decisions, thereby maintaining structural integrity and enhancing the structural reliability of civil infrastructure. Vision-based surface structural damage recognition, as one of the most intuitive and convenient approaches, is considered as a vital task in structural health diagnosis.

As the civil infrastructure continues to expand in scale and volume, current demands for intelligent and automated structural damage recognition have become increasingly urgent. This is attributed to the inherent limitations of conventional manual inspection methods, which are labor-intensive, time-consuming, lack stability and efficiency [1], and can often yield subjective results when attempting to obtain quantitative results. However, recent advances in computer vision and deep learning offer a non-contact and nondestructive solution for structural damage recognition using vision-based data. These data can be sourced from various platforms such as unmanned aerial vehicles, robotic inspections, and monitoring cameras [2,3]. By integrating these automated approaches with intelligent algorithms, it becomes possible to quantify damage, assist in monitoring, and establish thresholds to eliminate operational variability.

Initially, most studies of structural damage detection were conducted based on digital image processing techniques that required numerous pre-defined operations and parameters, including edge detection [4,5], threshold segmentation [6], and template matching [7]. Machine learning algorithms are investigated to train data-driven prediction models for new samples, which have been widely applied in areas of structural damage detection [8,9], condition assessment [10,11], and scene classification [12,13]. However, conventional machine learning typically relies on hand-crafted features and parameters that require prior knowledge and professional understanding of the source domain data. This would limit its generalization ability in target domains under new scenarios.

Deep learning has recently witnessed significant advances, particularly in the automatic extraction of abstract feature representations from large-scale datasets. This has propelled it to become a potent tool for structural health monitoring [14]. The application of computer vision in conjunction with novel deep learning algorithms facilitates the swift and precise identification and localization of structural surface damage by discerning intricate patterns from multi-source sensing data [15–18]. Sun et al. [19] conducted a comprehensive review of recent advances in artificial intelligence-enhanced bridge health monitoring. Bao and Li [20] established a unified framework of machine-learning-based structural health diagnosis to elucidate underlying mechanisms and structural dynamics within the multi-type monitoring data.

As one of the most frequently utilized deep learning models, convolutional neural networks (CNNs) are designed to extract multi-level image features and establish end-to-end connections with structural damage annotations. CNNs for computer vision have been utilized extensively in the field of structural damage recognition, including image classification [21,22], object detection [23,24], and semantic segmentation [25,26]. Crack recognition stands out as a key task in structural damage identification. Zhang et al. [27] introduced a novel gated recurrent unit that incorporated multi-layer nonlinear transformations into a recurrent neural network, thereby extracting embedding features of surface cracks. Xu et al. successively proposed high-performance identification frameworks for tiny steel fatigue cracks with complex handwriting marks [28] and further designed a lightweight segmentation model using real-world inspection images [29]. Pan et al. [30] integrated DeepLabV3+ and the dual attention module to explicitly model local feature representations of steel defects. In addition, a series of deep CNNs have been also established to tackle specific structural damage identification tasks, e.g., concrete wind-erosion [31], steel rust [32], and water leakage in tunnels [33].

The previously mentioned conventional deep-learning-based methods for structural damage identification predominantly utilize supervised learning, specifically designed for a particular type of damage. These methods lack the capacity to generalize across

small training datasets and new damage categories. Despite the consistent differences in morphological features among various types of structural damage, it is challenging to directly apply models trained on specific datasets under diverse real-world scenarios in engineering practices. This inevitably leads to the fragmented requirement of individualized models trained on different damage types. Additionally, the model performance and recognition accuracy are highly dependent on the quantity of training samples and class balance within the collected multi-type structural damage images. However, acquiring real-world images and the corresponding precise annotations of one particular damage type remains significantly challenging along with time-consuming and labor-intensive tedious labeling processes. Therefore, it is urgent to develop a stable training-testing approach of a high-performance, universal, and robust method for multi-type structural damage segmentation under limited supervised image-label pairs.

Few-shot learning has garnered significant attention in the deep learning field, serving as an efficacious cross-task learning paradigm, especially for computer vision recognition tasks of image classification, object detection, and semantic segmentation. In contrast to conventional deep learning methods that rely on a large-scale labeled dataset to ensure high training robustness and model accuracy, few-shot learning emphasizes the efficient acquisition of universal knowledge from limited datasets and rapid adaptation to new tasks. The primary objective of few-shot learning is to discover the underlying relationships and inherent similarities across various recognition tasks, which can be shared as transferable prior knowledge applicable to new object categories.

Based on intrinsic learning mechanisms and the utilization patterns of limited supervision information, few-shot learning can be categorized into four types: metric-learning-based methods, optimization-based methods, transfer-learning-based methods, and generative-model-based methods. Metric-learning-based methods utilize various prototypical networks to learn the optimal prototype approximations of each category, thereby classifying the unseen samples [34–36]. The model agnostic meta-learning (MAML) algorithm has garnered significant attention in the few-shot learning field due to its strong universality and effectiveness as an exemplary optimization-based method [37]. Nichol et al. [38] further proposed a simplified first-order optimization-based meta-learning algorithm to facilitate rapid fine-tuning on new tasks. Transfer-learning-based methods, for instance, the meta-transfer learning paradigm [39], simultaneously combine the benefits of both transfer learning and meta-learning to leverage prior knowledge from the meta-learner for fine-tuning on unseen tasks. In addition, recent large-scale pre-trained language models have demonstrated great potential in transferring generalized cross-task knowledge that can be adapted to specific tasks via only a few samples [40]. Generative-model-based methods focus on generating synthetic samples from the learned probability distributions of small samples for actual data, thereby augmenting the training dataset using mature generative models [41,42].

Recently, researchers have focused on few-shot learning applications in structural health monitoring and damage recognition. Guo et al. [43] developed a defect classifier based on meta-learning that iteratively adjusted the weight coefficients of loss function to alleviate the adverse effects of data imbalance. A metric-learning-based model was introduced for few-shot pavement defect classification, improving the distinguishability between various defect classes [44]. Xu et al. [45] proposed a meta-learning classification framework based on attribute representation vectors of multiple damage categories, where damage attributes act as the common inter-class transferable knowledge. A few-shot classification approach for previously unseen classes was presented by incorporating an extensible classifier with contrastive learning, addressing the challenges of data imbalance in small datasets [46].

The aforementioned methods offer viable solutions for certain few-shot learning tasks in structural health monitoring; however, they primarily concentrate on structural damage classification, neglecting to provide more detailed and nuanced predictions at the pixel level. In addition, although some recent, related studies have started to focus on the specific

task generation in meta learning, the task significance was represented with an interpretable task generation strategy by high-dimensional feature density clustering. The knowledge gap still clearly exists between theoretical meta-learning algorithms and the actual application of structural damage segmentation. To address this challenge, this study aims to establish an optimization-based few-shot learning approach for multi-type structural damage segmentation with inadequate supervised pixel-level annotations. Specifically, this study proposes a dual-stage optimization-based few-shot learning (DOFSL) framework for multi-type structural damage segmentation, mitigating the severe training instability and insufficient model robustness observed in conventional supervised learning when applied to small-scale annotated datasets. Furthermore, comparative studies and ablation experiments are systematically conducted to demonstrate the effectiveness, robustness, and generalization of the proposed method. The proposed method provides an effective solution to achieve image-based structural damage recognition with high accuracy and robustness for bridges and buildings, which assists the unmanned intelligent inspection of civil infrastructure using drones and robotics in smart cities.

The remainder of this article is organized as follows. Section 2 introduces the proposed DOFSL methodology for multi-type structural damage segmentation. Section 3 elaborates on the implementation details of the investigated multi-type damage image dataset, training hyperparameter configurations, loss function, and evaluation metrics. Section 4 shows the test results, comparative studies, and ablation experiments to validate the efficacy and necessity of the DOFSL method. Finally, Section 5 draws conclusions for this study.

2. Methodology

This study proposes a dual-stage optimization-based few-shot learning (DOFSL) algorithm inspired by the original MAML to enhance the model generalization ability for different structural damage categories. The methodology section is structured as follows. Section 2.1 provides the problem definition of few-shot learning for identifying multi-types of structural damage with only a limited set of images. Section 2.2 introduces the dual-stage optimization-based few-shot learning paradigm, which includes an internal semantic segmentation model optimization process based on an individual meta-task and an external meta-learning-machine optimization process based on various meta-batches. Section 2.3 presents the internal semantic segmentation model for the proposed DOFSL optimization algorithm in detail.

2.1. Problem Definition

The entire dataset D containing multi-type structural damage images is defined as follows:

$$D = \{D_i\} = \left\{ (I_{i,j}, Y_{i,j}) \right\}, i = 1, \ldots, N_D, j = 1, \ldots, N_i \tag{1}$$

where D_i denotes the image subsets for the ith structural damage category; N_D represents the overall quantity for considered damage categories; N_i represents the number of damage images included in the ith subsets D_i; and (I, Y) denotes a pair of the input image and associated annotation at the pixel level. Then, D is randomly divided into a training set D^{train} and a test set D^{test}, in which D^{train} is used for model training, while D^{test} is used to evaluate model performance as follows:

$$D^{train} = \left\{ (I_{i,j}^{train}, Y_{i,j}^{train}) \right\}, i = 1, \ldots, N_{train}, j = 1, \ldots, N_i^{train}$$
$$D^{test} = \left\{ (I_{i,j}^{test}, Y_{i,j}^{test}) \right\}, i = 1, \ldots, N_{test}, j = 1, \ldots, N_i^{test} \tag{2}$$

where N_{train} and N_{test} denote the quantity of damage categories including in training and test sets and N_i^{train} and N_i^{test} denote the image number for the ith damage category within them. It should be noted that there exist two representative scenarios in the model test stage: (1) prediction for the known categories in the training set, and (2) prediction for the completely new categories that have not been seen during the training stage.

Subsequently, the meta-task T is selected in N-way-K-shot form with K images from each of N categories from the training set by random sampling with replacement, which is randomly separated into a support set S^{train} and a query set Q^{train} as follows:

$$T_k = S_k^{train} \cup Q_k^{train} \subseteq D^{train}, k = 1, \dots, N_{task}$$
$$S_k^{train} = \left\{ (I_{k,i}^{train,S}, Y_{k,i}^{train,S}) \right\}_{i=1}^{N \times K^S}, Q_k = \left\{ (I_{k,i}^{train,Q}, Y_{k,i}^{train,Q}) \right\}_{i=1}^{N \times K^Q} \quad (3)$$

where the support set and the query set for a specific meta-task contain K^S and K^Q samples, respectively, i.e., $K^S + K^Q = K$; N denotes the included categories for structural damage in the meta-task T_k; and N_{task} denotes the number of meta-tasks within one meta-batch. The meta-batch Γ can be generated by iteratively absorbing the meta-task as follows:

$$\Gamma = \overset{N_{task}}{\underset{k=1}{\cup}} T_k \quad (4)$$

For the test stage, the N-way-K-shot test tasks should also be generated to maintain consistency with the training stage. The only difference is that the test support set is randomly sampled from the annotated samples, while the test query set is sequentially sampled from the predicted samples as follows:

$$T_k^{test} = S_k^{test} \cup Q_k^{test} = \left\{ (I_{k,i}^{test,S}, Y_{k,i}^{test,S}) \right\}_{i=1}^{N \times K^S} \cup \left\{ (I_{k,i}^{test,Q}, Y_{k,i}^{test,Q}) \right\}_{i=1}^{N \times K^Q} \quad (5)$$

where the superscript "test" indicates the test stage. Annotated test support images S_k^{test} are adopted for obtaining the fine-tuned model, which can be directly adapted to the unseen query samples in Q_k^{test} to calculate the model performance for prediction accuracy on the test query set.

2.2. Dual-Stage Optimization-Based Few-Shot Learning (DOFSL)

The overall schematic of the proposed DOFSL is illustrated in Figure 1. For the single-stage optimization method of conventional machine learning, the entire training data are generally fed into the network for gradient descent and parameter optimization. In contrast, the dual-stage optimization method separately utilizes the support and query sets for parameter updating in the training tasks of internal and external optimizers, respectively. The dual-stage training process can be basically summarized as follows:

(1) Multiple meta-tasks are generated by randomly sampling with replacement from the training set and disordered to form meta-batches for the latter model optimization;

(2) Each task inside a meta-batch is individually fed into the internal semantic segmentation network, in which the support set is utilized to update model parameters in the internal optimization stage, and the query set is adopted to compute the prediction loss by the updated model;

(3) The external optimization stage for the meta-learning machine is performed based on all the query losses inside a meta-batch, which is concretized as updating the initial network parameters. Following this manner, the internal semantic segmentation model learns universal prior knowledge among various damage categories from the training meta-tasks and transfers it to the test tasks.
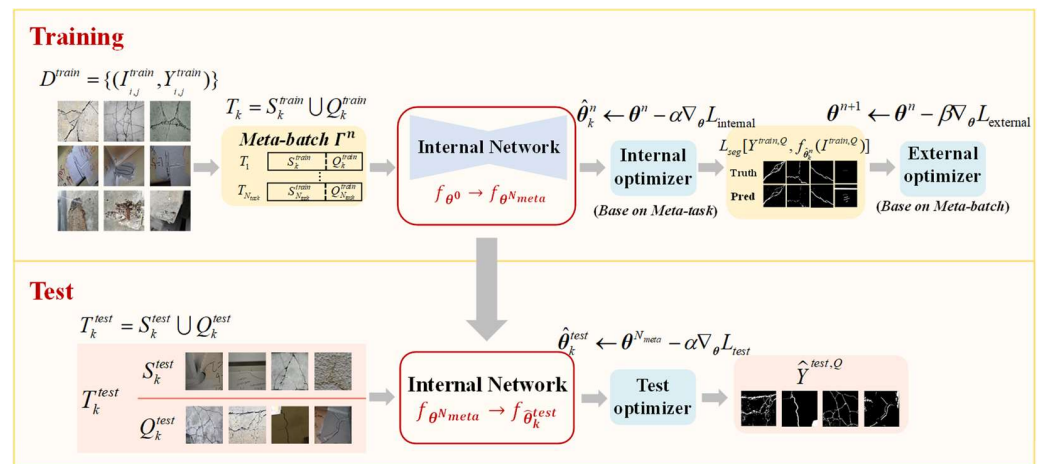
Figure 1. Overall schematic of dual-stage optimization-based few-shot learning for multi-type structural damage segmentation.

Suppose that the internal semantic segmentation model f_θ is defined with learnable network parameters θ. As shown in Figure 1, DOFSL comprises an internal optimization process based on an individual meta-task T_k^n and an external optimization process based on a meta-batch $\Gamma^n = \left\{ T_1^n, \ldots, T_k^n, \ldots, T_{N_{task}}^n \right\}, n = 1, \ldots, N_{meta}$, where N_{meta} is the quantity of meta-batches for the training process. The dual-stage optimization process can be formulated as shown below.

(1) The internal optimization stage is performed on the individual meta-task T_k^n, in which each support set is separately utilized to update the internal semantic segmentation network f_θ to obtain a series of updated network parameters $\hat{\theta}^n = \left\{ \hat{\theta}_1^n, \ldots, \hat{\theta}_k^n, \ldots, \hat{\theta}_{N_{task}}^n \right\}$ as follows:

$$\hat{\theta}_k^n \leftarrow \theta^n - \alpha \nabla_\theta L_{\text{internal}}, L_{\text{internal}} = \frac{1}{N^S} \sum_{i=1}^{N^S} L_{seg}[Y_i^{train,S}, f_{\theta^n}(I_i^{train,S})] \tag{6}$$

where θ^n represents the initial network parameters for the nth meta-batch; $\hat{\theta}_k^n$ denotes the optimized network parameters for the kth meta-task in the nth meta-batch; α denotes the internal learning rate; ∇_θ denotes the parameter gradient computation; and L_{seg} denotes the segmentation loss for model training.

(2) The external optimization stage is performed on a meta-batch Γ^n, in which all the query losses for meta-tasks are calculated based on the updated parameters $\hat{\theta}^n$ to optimize the initial parameters θ^n. Subsequently, the training process of across-task meta-learning machine is conducted in the external optimization stage via gradient descent considering all query losses in the meta-batch Γ^n as follows:

$$\theta^{n+1} \leftarrow \theta^n - \beta \nabla_\theta L_{\text{external}}, L_{\text{external}} = \frac{1}{N_{task}} \sum_{k=1}^{N_{task}} \frac{1}{N^Q} \sum_{i=1}^{N^Q} L_{seg}[Y_i^{train,Q}, f_{\hat{\theta}_k^n}(I_i^{train,Q})] \tag{7}$$

where β denotes the external learning rate. After completing all the external optimization processes for N_{meta} meta-batches, the updated model parameters $\theta^{N_{meta}}$ are obtained for the test tasks.

2.3. Internal Model Structure for Semantic Segmentation of Multi-Type Structural Damage

In this study, U-Net [47] is utilized as the internal model for the semantic segmentation of multi-type structural damage. The overall network structure of the internal semantic segmentation U-Net model is shown in Figure 2, adopting a classic encoder–decoder symmetric structure with a U-shape. It introduces multi-level feature concatenation by

short-cut skip connections between the same stage of encoder and decoder, leading to clearer edge detection and more refined segmentation granularity of multi-scale structural damage regions.
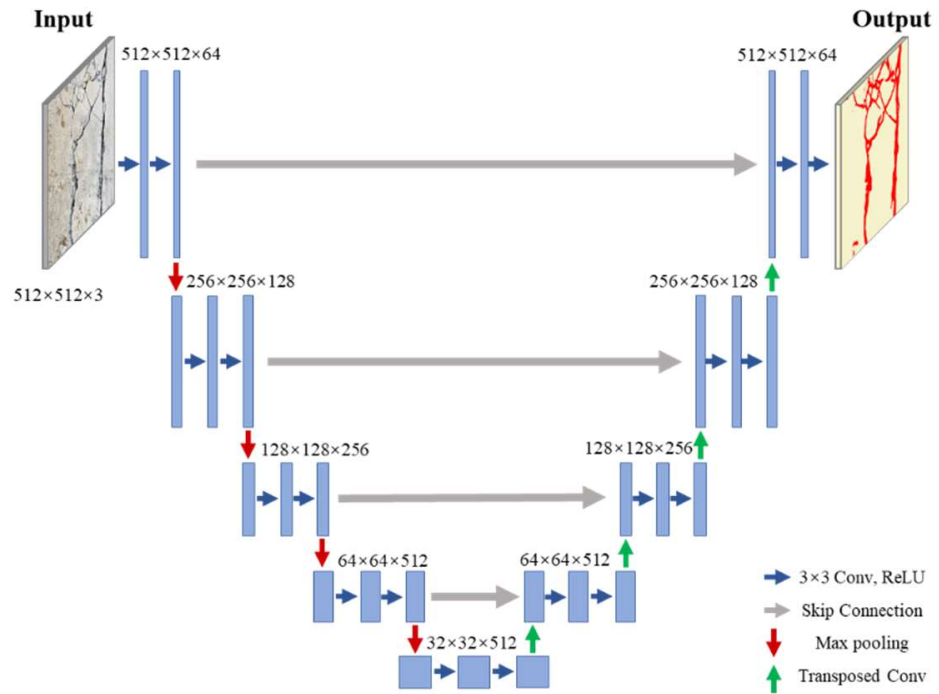

Figure 2. Network structure of internal semantic segmentation U-Net model.

The left-side encoder part serves as the feature extractor of structural damage, which performs image downsampling operations in succession via four convolutional stages. Each convolution module contains two 3×3 convolutional layers, a ReLU nonlinear activation layer, and a 2×2 max pooling operation perceiving the most significant sub-regions of convolutional feature maps. The horizontal and vertical size of feature maps is reduced to 1/2, and the number of channels is doubled through each encoder stage.

The right-side decoder part is utilized to accomplish image upsampling and restore the original resolution through four transposed convolutional stages. Compared with preset bidirectional interpolation, transposed convolution is regarded as a learnable upsampling operation with a richer complement of information, thereby leading to higher reliability and less loss of segmentation accuracy. The plane size transformation between the input and output feature maps for transposed convolution can be expressed as follows:

$$H^{l+1} = S \times (H^l - 1) - P_H^{l+} - P_H^{l-} + H' \tag{8}$$

where H represents the feature map size (width as the horizontal size or height as the vertical size); S represents the sliding stride; P represents the zero-padding size, + and $-$ represent positive and negative paddings in the horizontal or vertical directions; and H' represents the size of the transposed convolution kernel. Each transposed convolutional layer first employs a 2×2 transposed convolution with a stride of 2 to double the plane resolution of feature maps, followed by a 3×3 convolutional layer for feature fusion. The upsampling feature maps are then concatenated with the corresponding feature maps in the same stage of the encoder along the channel direction via skip connection. Then, the 1×1 point convolution is adopted to recover the original channel number. The schematic of the ship connection between the same stage of encoder and decoder is shown in Figure 3.
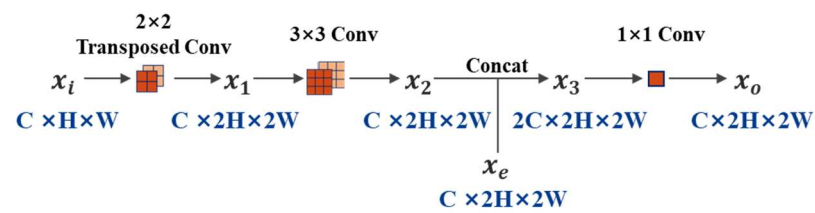
Figure 3. Schematic of ship connection between the same stage of encoder and decoder.

It should be noted that utilizing the pre-trained model based on large-scale datasets and transferring the learned knowledge to downstream tasks can achieve few-shot learning in another manner, which is an alternative manner and does not conflict with the proposed dual-stage optimization method. Furthermore, the proposed DOFSL algorithm is actually model-agnostic, and the internal network can be adjusted based on the task objective and data quality for an optimal match. When a large number of samples are available, the foundational meta-learner can be trained to employ internal networks with large-volume parameters for universal structural damage recognition of both known and unknown categories.

3. Implementation Details

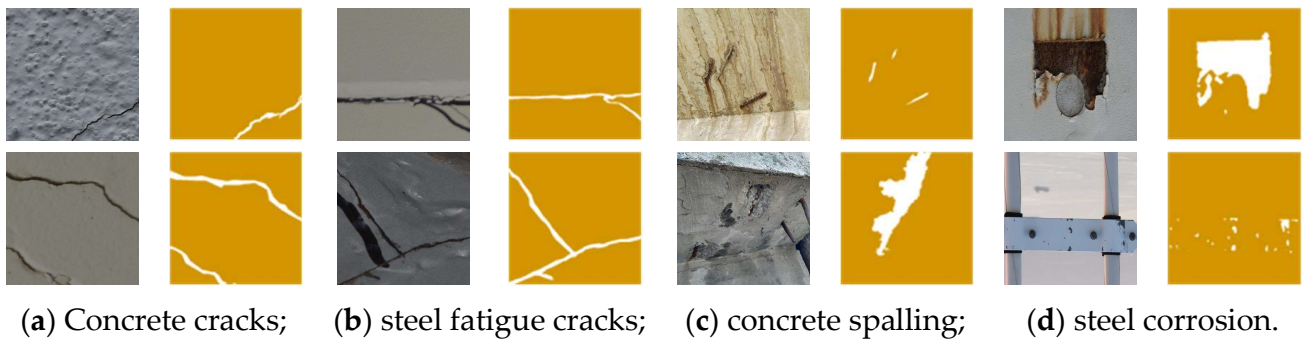### 3.1. Multi-Type Structural Damage Image Dataset

The investigated multi-type structural damage dataset includes four distinct categories: concrete cracks, steel fatigue cracks, concrete spalling, and steel corrosion. The collected images are derived from actual civil infrastructure inspections under real-world scenarios and cut into patches with a consistent resolution of 512×512 pixels. The investigated images were obtained through manual onsite inspection by various inspectors, and the shooting angles are unknown and different under diverse scenarios. However, a straightforward principle was utilized during the image capturing process to ensure that the shooting direction of the collected damage images is approximately perpendicular to the damage plane. This fundamental operation would thereby reduce image distortions and quantification errors as far as possible.

Figure 4 presents the partial representative image-label pairs of the investigated multi-type structural damage dataset. The images, featuring a yellow background and white foreground corresponding to structural damage, are labeled masks of the original left-sided images. These labels were manually obtained using the "labelme" tool to achieve pixel-level annotations. The training set is independently chosen from concrete cracks, steel fatigue cracks, and concrete spalling comprising 100 images in each category. Conversely, the test set is composed of four categories (including known categories of concrete cracks, steel fatigue cracks, and concrete spalling along with a new category of steel corrosion), each also containing 100 images. Half of them are randomly selected as annotated samples to form the test support set, while the remaining samples serve as the test query set.

Figure 4. *Cont.*

(**a**) Concrete cracks; (**b**) steel fatigue cracks; (**c**) concrete spalling; (**d**) steel corrosion.

Figure 4. Representative image-annotation pairs of investigated multi-type structural damage.

3.2. Training Hyperparameter Configurations

Hyperparameter settings are critical to the model performance of deep learning networks, thereby affecting the training stability of segmentation models for multiple damage categories, especially few-shot recognition under limited supervision. A series of trials were first conducted to obtain the preferable learning rate settings for model optimization as $\alpha = 0.01$, $\beta = 0.001$. Subsequently, the fairness criterion is emphasized for hyperparameter settings to ensure more reliable and convincing results in the comparative experiments. The main principle is to keep an approximate equivalence of the total utilization of damage images during the training process between the DOFSL algorithm and the original U-Net network.

The approximation for the total utilization of images of the proposed DOFSL can be computed by the following:

$$N_{DOFSL} = N_{step\_e} \times N_{meta} \times [N_{task} \times (N_{step\_i} \times K^S + K^Q)] \tag{9}$$

where N_{step_e} denotes the quantity of external optimization iterations for the meta-learning machine and N_{step_i} denotes the quantity of internal optimization iterations for the semantic segmentation network.

The approximation for the total utilization of images of the proposed DOFSL can be computed by the following:

$$N_O = N_{epoch\_O} \times N_i^{train} \times N_{train} \tag{10}$$

where N_{epoch_O} denotes the total training epochs using all labeled training images.

Following the fairness principle of $N_O \approx N_{DOFSL}$, the hyperparameter configurations are determined as follows: $N_{epoch\_O} = 20$, $N_i^{train} = 100$, $N_{train} = 3$, $N_{step\_e} = 1$, $N_{meta} = 100$, $N_{task} = 2$, $N_{step\_i} = 5$, $K^S = 5$, $K^Q = 2$.

3.3. Specifications of Training Loss Function and Test Evaluation Metrics

Due to the area imbalance problem existing between the damage regions and the background, the dice segmentation loss [48] is adopted in the training updating and test fine-tuning processes as follows:

$$L_{seg} = \frac{1}{B} \sum_{j=1}^{B} \frac{1 - 2 \sum_{i=1}^{H \times W} p_i \times y_i}{\sum_{i=1}^{H \times W} p_i + y_i} \tag{11}$$

where H and W represent the height and width of image; B represents the batch size; p_i represents the probability that the ith pixel is identified as a damage pixel; and y_i represents the ground-truth label of the ith pixel. According to Equation (11), the value of dice loss ranges from 0 to 1, thereby simplifying or reducing additional standardization techniques.

For practical application, a scaling parameter would be employed for observation quantification during the image capturing process to obtain the sizes of the damage from image pixel coordinates to physical parameters.

The Adam algorithm is adopted to update the internal U-Net segmentation model as follows:

$$
\begin{aligned}
&g_t = \frac{1}{N^s}\nabla_\theta Loss_{internal} \\
&m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t,\ v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2 \\
&\hat{m}_t = m_t/(1-\beta_1^t),\ \hat{v}_t = v_t/(1-\beta_2^t) \\
&\theta_t = \theta_{t-1} - \alpha\hat{m}_t/(\sqrt{\hat{m}_t} + \varepsilon)
\end{aligned}
\tag{12}
$$

where m_t and v_t denote the first-order and second-order moment estimates for the average loss gradient for the model parameters; $\hat{m}_t$ and $\hat{v}_t$ denote the gradient moment after deviation correction; β_1 and β_2 denote the exponential decay rates for the first-order and second-order moment estimates with the default values of $\beta_1 = 0.9$, $\beta_2 = 0.999$; and ε is a small value for numerical stability.

Typical evaluation metrics are utilized to evaluate the segmentation performance of structural damage regions, including mIoU (mean intersection-over-union) and mPA (mean pixel accuracy), as follows:

$$
\mathrm{mIoU} = \frac{1}{N+1}\sum_{i=1}^{N+1}\frac{p_{ii}}{\sum_{j\neq i}p_{ij} + \sum_{j\neq i}p_{ji} + p_{ii}},\ \mathrm{mPA} = \frac{1}{N+1}\sum_{i=1}^{N+1}\frac{p_{ii}}{\sum_{j\neq i}p_{ij}}
\tag{13}
$$

where N represents the number of pixel categories for foreground damage regions and p_{ij} represents the pixel number of the pixels belonging to the ith class identified as the jth class.

The model training and testing are conducted with the software environment of PyTorch 1.9.1 and Python 3.8 on a 24 G GPU of NVIDIA RTX A6000.

4. Results and Discussion

4.1. Test Results for Multi-Type Structural Damage Segmentation

This section compares and analyzes the performance of the proposed DOFSL method and the original semantic segmentation network (U-Net) for multi-type structural damage segmentation. Several representative segmentation results for concrete cracks, steel fatigue cracks, and concrete spalling are shown in Figure 5, and the comparative boxplots of average mIoU and mAP are shown in Figure 6. Compared to the original U-Net, the proposed DOFSL achieves higher recognition accuracies for multi-type structural damage, accommodating multi-scale, multi-morphology, diverse damage severities, and complex backgrounds under few-shot supervised scenarios. The proposed DOFSL method verifies its stability and robustness by the accurate identification of clear damage edges, resisting interference from complex noise and damage-like backgrounds, and exhibiting sensitivity to local, tiny damage regions. The quantitative evaluation results indicate that utilizing the proposed DOFSL over the original segmentation network U-Net leads to an average increase in mIoU and mPA of 5.5% and 10.0%, respectively. In summary, test results demonstrate the effectiveness of the proposed method for the cross-task recognition of multiple damage categories compared to the internal segmentation model itself.
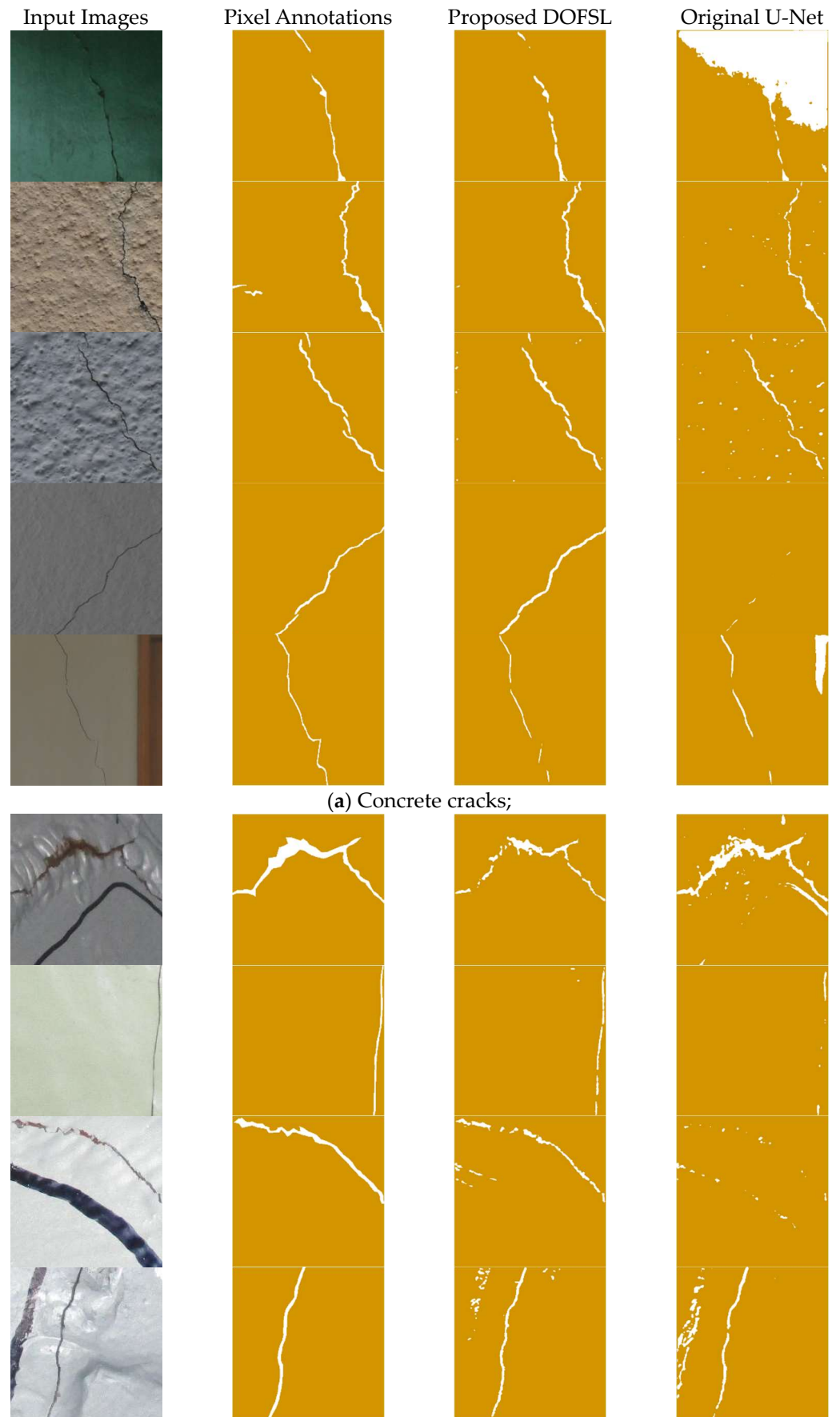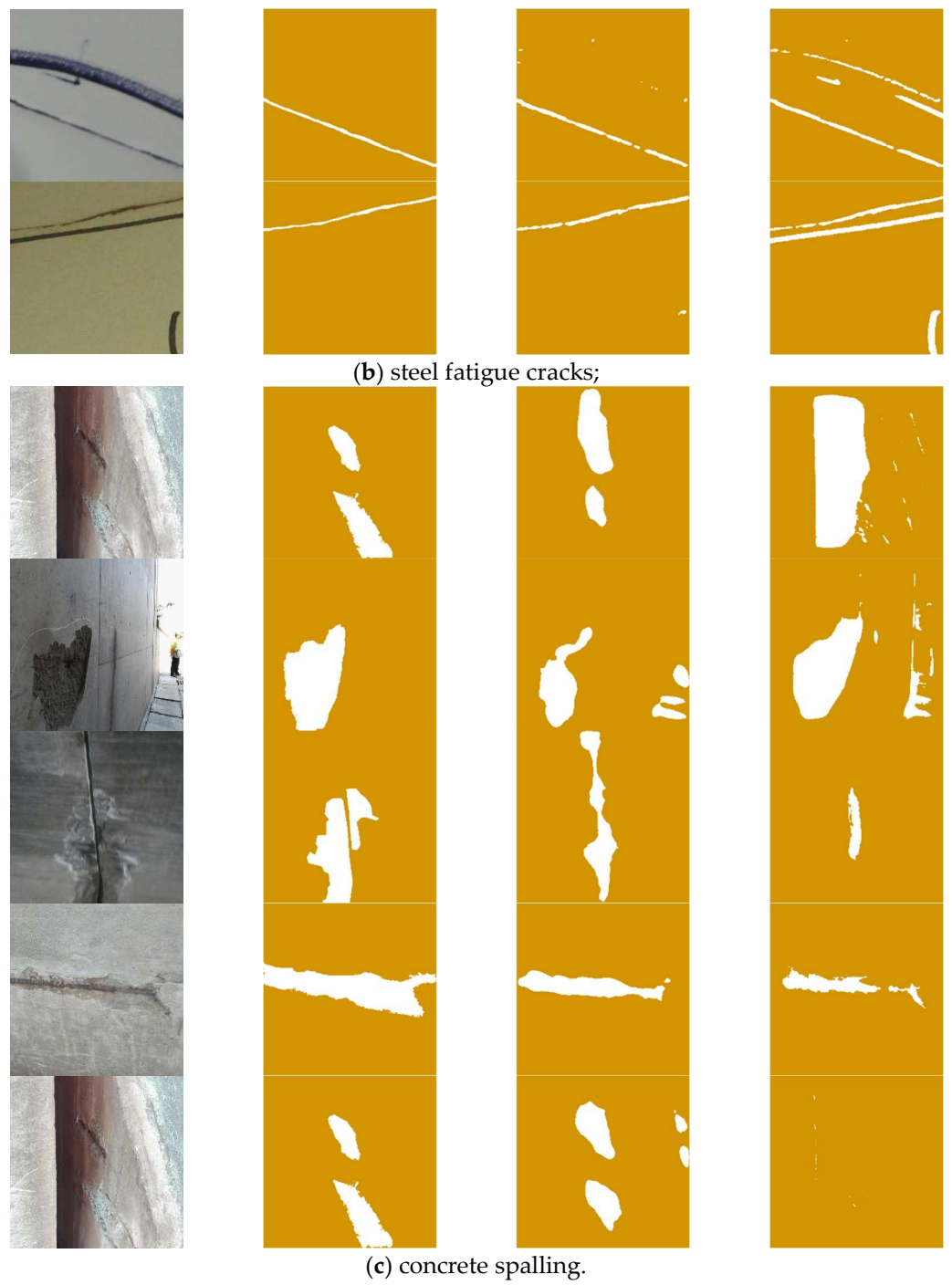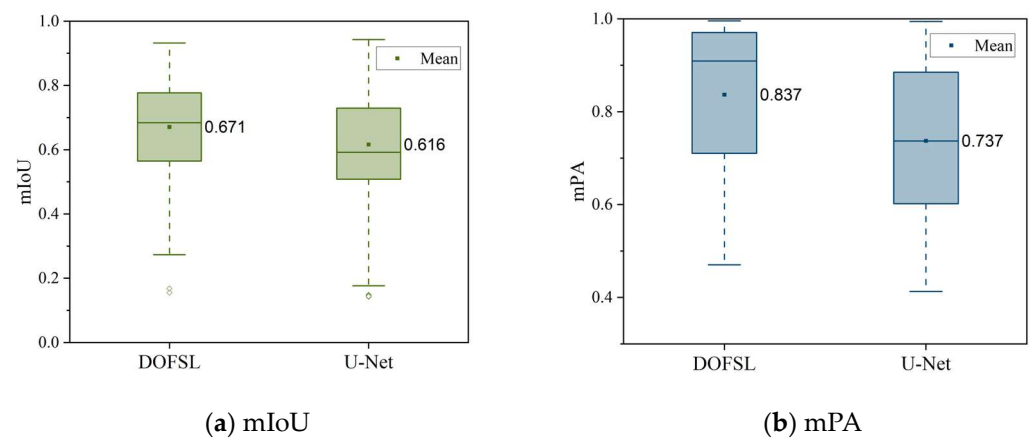
Input Images	Pixel Annotations	Proposed DOFSL	Original U-Net

(**a**) Concrete cracks;

Figure 5. *Cont.*

(**b**) steel fatigue cracks;

(**c**) concrete spalling.

Figure 5. Representative test results of image segmentation for multi-type structural damage.

(**a**) mIoU (**b**) mPA

Figure 6. Comparative boxplots of test evaluation metrics using DOFSL and original U-Net for multi-type structural damage. (Solid lines represent the min, Q1 quartile, Q2 quartile, Q3 quartile, and max values of the statistical evaluation metric, dashed lines indicate the value range, and circles denote the outliers).

*4.2. Validation of Generalization Ability for Unseen Structural Damage Category*

To demonstrate the generalization capacity for the newly emerging damage category, the known damage images with annotations (i.e., concrete cracks, steel fatigue cracks, and concrete spalling) are employed to train the meta-learning machine, which can be fine-tuned on the images of unseen steel corrosion in the test process. For the new steel corrosion damage, the test support set consists of 50 annotated images, and the test query set consists of the remaining 50 images. A few representative segmentation results and comparative boxplots of the test evaluation measures for the unseen damage of steel corrosion are illustrated in Figures 7 and 8, respectively. As shown in Figure 7, the comparison results clearly demonstrate that the proposed method can achieve reasonable perception and accurate recognition of steel corrosion regions. The quantitative analysis results reveal that the steel corrosion segmentation achieved an average increase in mIoU and mPA of 21.9% and 6.6%, which further verifies the robustness and generalization capacity of the DOFSL method for unknown and diverse damage categories. It should be noted that the corrosion damage is randomly selected as the unseen category to validate the generalization ability. The concrete cracks, steel fatigue cracks, and concrete spalling can also be arbitrarily considered as the unseen damage. In these instances, once an unseen category has been selected, the remaining three categories would be employed to train the model using the proposed method.
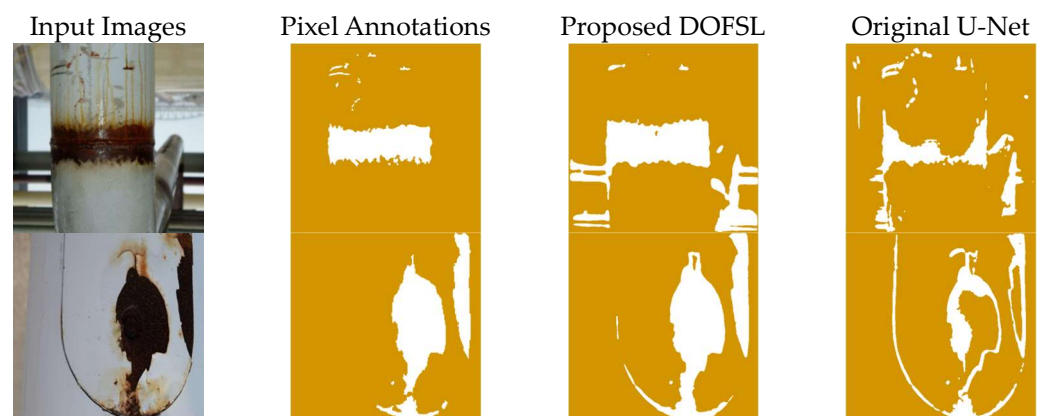
Figure 7. *Cont.*

Figure 7. Representative test segmentation results for unseen category of steel corrosion.
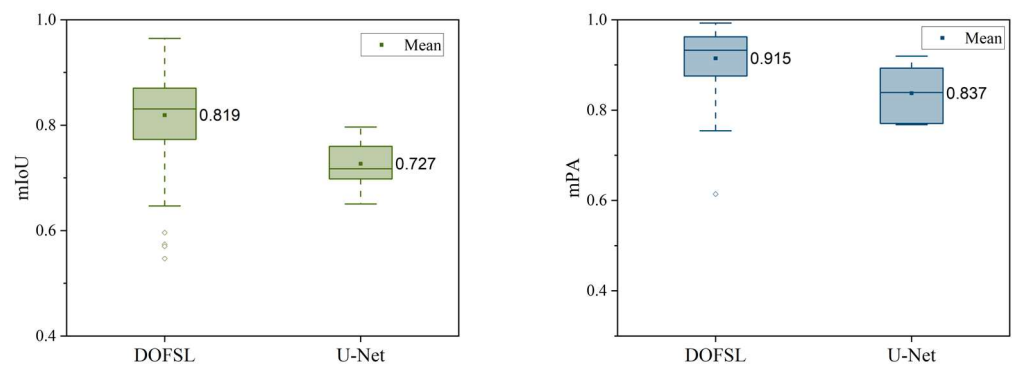

(**a**) mIoU

(**b**) mPA
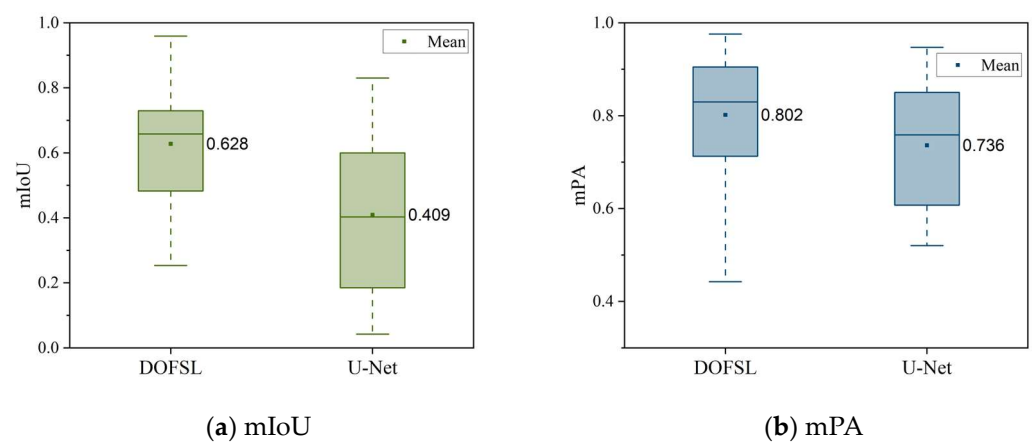
Figure 8. Comparative boxplots of test evaluation metrics using DOFSL and original U-Net for unseen category of steel corrosion. (Solid lines represent the min, Q1 quartile, Q2 quartile, Q3 quartile, and max values of the statistical evaluation metric, and dashed lines indicate the value range).

4.3. Ablation Studies for Individual-Type Damage Segmentation

The adaptability and universality of the proposed method are further considered when faced with various operational conditions encountered in actual engineering. This section demonstrates the model performance on individual-type damage as ablation studies (i.e., concrete cracks, steel fatigue cracks, and concrete spalling). The corresponding comparative analyses are conducted with the results obtained from the original segmentation model of U-Net. The comparative boxplots of test evaluation metrics using the proposed DOFSL method and original U-Net for distinct damage categories are shown in Figure 9. The comparative results indicate that the proposed DOFSL achieves the optimal performance on total datasets with noticeable improvements in both evaluation metrics, and can adapt to an arbitrary damage class. According to the test results, the proposed DOFSL method exhibits higher accuracies across all three types of structural damage with significant improvements in both average mIoU (9.2%, 7.3%, and 4.3%) and mPA (7.8%, 7.6%, and 6.6%) for concrete cracks, steel fatigue cracks, and concrete spalling, respectively. It also indicates that the proposed DOFSL can adapt and generalize to any structural damage category.

The variances in mIoU and mPA were larger for the proposed DOFSL compared to directly training the original U-Net. The main possible reasons are as follows: (1) the small-scale dataset and scattered data distribution within the same task for a specific damage type may lead to unstable test results; (2) samples of randomly selected damage images may exhibit significantly varied features, which result in intra-class diversity with a large variance; (3) the original U-Net tends to be generally insensitive to damage regions with only a few samples, in which the recognized damage regions are typically smaller with stable but much lower test accuracy.

(**a**) mIoU and mPA for concrete cracks;
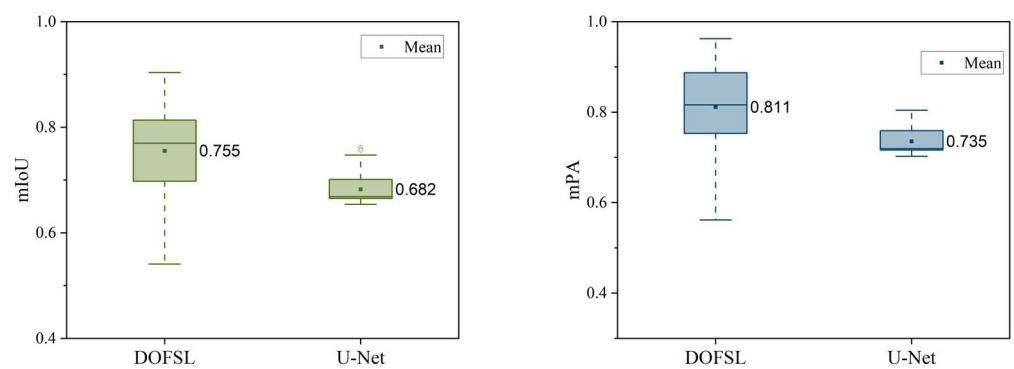
(**b**) mIoU and mPA for steel fatigue cracks;
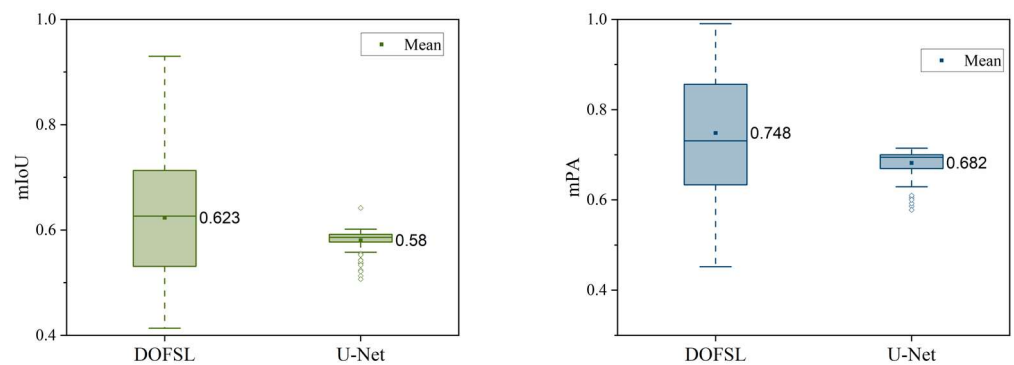
(**c**) mIoU and mPA for concrete spalling.

Figure 9. Comparative boxplots of test evaluation metrics using DOFSL and original U-Net. (Solid lines represent the min, Q1 quartile, Q2 quartile, Q3 quartile, and max values of the statistical evaluation metric, dashed lines indicate the value range, and circles denote the outliers).

Table 1 compares the average test evaluation metrics using different numbers of annotated training samples for individual damage categories. The proposed DOFSL method always yields optimal model performances with an average mIoU and mPA of 81.5% and 91.5% for concrete cracks, 75.5% and 81.1% for steel fatigue cracks, and 64.8% and 78.6% for concrete spalling. It is worth highlighting that the proposed DOFSL method exhibits insensitivity to variations in the number of training samples when compared to the original U-Net, which could maintain high recognition accuracy even with half of the training samples. This observation further underscores the effectiveness, stability, and robustness of the proposed DOFSL method for different damage categories segmentation with only a few supervised samples.

Table 1. Comparisons of average test metrics with different numbers of training samples.

Damage Category	Model	N_i^{train}	Average mIoU	Average mPA
Concrete cracks	Original U-Net	100	72.7%	83.7%
	Original U-Net	200	77.1%	84.2%
	Proposed DOFSL	100	81.4%	91.5%
	Proposed DOFSL	200	81.5%	89.7%
Steel fatigue cracks	Original U-Net	100	68.2%	73.5%
	Original U-Net	200	68.5%	81.3%
	Proposed DOFSL	100	75.5%	81.1%
	Proposed DOFSL	200	73.4%	80.9%
Concrete spalling	Original U-Net	100	58.0%	68.2%
	Original U-Net	200	59.9%	76.3%
	Proposed DOFSL	100	62.3%	74.8%
	Proposed DOFSL	200	64.8%	78.6%

5. Conclusions

This study proposes a dual-stage optimization-based few-shot learning approach for multi-type structural damage segmentation using limited pixel-annotated images. The primary conclusions of this study are provided as follows.

(1) The dual-stage optimization-based few-shot learning framework is established containing the internal network optimization stage based on meta-task and the external meta-learning-machine optimization based on meta-batch. The mathematical formulation of few-shot learning-based multi-type structural damage segmentation is formed exclusively relying on limited supervised images.

(2) Comparative experiments are conducted to verify the effectiveness and necessity of the proposed dual-stage optimization-based few-shot learning method using the multi-type structural damage image set including concrete cracks, steel fatigue cracks, concrete spalling, and steel corrosion. The results indicate that compared with the original image segmentation model, the proposed DOFSL achieves an average increase in mIoU and mPA of 5.5% and 10.0%, respectively.

(3) Furthermore, ablation studies for individual damage types and new damage categories are implemented to validate the model stability, generalization capacity, and universal applicability of the proposed DOFSL method for semantic segmentation of arbitrary structural damage categories. The quantitative analysis results achieve a significant improvement in average mIoU and mPA of 21.9% and 6.6% for unseen damage in the training dataset.

Future studies should be further performed on constructing a universal image segmentation model for any structural damage type via one-shot or zero-shot learning when facing completely new damage types that have rarely or never appeared in the training dataset.

6. Future Directions

Several possible directions could be further investigated to enhance the performance and applicability of vision-based structural damage recognition.

(1) Incorporating geometric constraints, such as curvature shapes and boundary condition features, can promote the robustness and accuracy of image segmentation [49,50]. These constraints can provide supplementary contextual information, facilitating model training to understand input images, predict structural damage, and avoid reliance on extensive labeled data.

(2) Network integration and modular design can be adopted to simplify network structures and reduce network complexity and training difficulty. Separate modules are designed and individually optimized for specific damage types, and they are subse-

quently integrated to the initial shared network to process and analyze multi-type structural damage using ensemble deep convolutional neural network models [51].

(3) Leveraging transfer learning and domain adaptation techniques can significantly improve the model performance for multi-type damage recognition, particularly when training samples are limited. The transferable knowledge is adaptively involved and optimized in specific fields, which enhances the generalization ability under different application scenes and damage scenarios [52].

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

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

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

References

1. Ali, A.; Sandhu, T.Y.; Usman, M. Ambient vibration testing of a pedestrian bridge using low-cost accelerometers for SHM applications. *Smart Cities* **2019**, *2*, 20–30. [CrossRef]
2. Liu, Y.; Cho, S.; Spencer, B.F., Jr.; Fan, J. Automated assessment of cracks on concrete surfaces using adaptive digital image processing. *Smart Struct. Syst.* **2014**, *14*, 719–741. [CrossRef]
3. Zakeri, H.; Nejad, F.M.; Fahimifar, A. Image based techniques for crack detection, classification and quantification in asphalt pavement: A review. *Arch. Comput. Methods Eng.* **2017**, *24*, 935–977. [CrossRef]
4. Adhikari, R.S.; Moselhi, O.; Bagchi, A. Image-based retrieval of concrete crack properties for bridge inspection. *Autom. Constr.* **2014**, *39*, 180–194. [CrossRef]
5. Luo, Q.; Ge, B.; Tian, Q. A fast adaptive crack detection algorithm based on a double-edge extraction operator of FSM. *Constr. Build. Mater.* **2019**, *204*, 244–254. [CrossRef]
6. German, S.; Brilakis, I.; DesRoches, R. Rapid entropy-based detection and properties measurement of concrete spalling with machine vision for post-earthquake safety assessments. *Adv. Eng. Inform.* **2012**, *26*, 846–858. [CrossRef]
7. Paal, S.G.; Jeon, J.S.; Brilakis, I.; DesRoches, R. Automated damage index estimation of reinforced concrete columns for post-earthquake evaluations. *J. Struct. Eng.* **2015**, *141*, 04014228. [CrossRef]
8. Figueiredo, E.; Park, G.; Farrar, C.R.; Worden, K.; Figueiras, J. Machine learning algorithms for damage detection under operational and environmental variability. *Struct. Health Monit.* **2011**, *10*, 559–572. [CrossRef]
9. Hsieh, Y.A.; Tsai, Y.J. Machine learning for crack detection: Review and model performance comparison. *J. Comput. Civ. Eng.* **2020**, *34*, 04020038. [CrossRef]
10. Morgenthal, G.; Hallermann, N.; Kersten, J.; Taraben, J.; Debus, P.; Helmrich, M.; Rodehorst, V. Framework for automated UAS-based structural condition assessment of bridges. *Autom. Constr.* **2019**, *97*, 77–95. [CrossRef]
11. Rafiei, M.H.; Adeli, H. A novel unsupervised deep learning model for global and local health condition assessment of structures. *Eng. Struct.* **2018**, *156*, 598–607. [CrossRef]
12. Xiao, Y.; Wu, J.; Yuan, J. mCENTRIST: A multi-channel feature generation mechanism for scene categorization. *IEEE Trans. Image Process.* **2013**, *23*, 823–836. [CrossRef]

13. Chen, C.; Zhang, B.; Su, H.; Li, W.; Wang, L. Land-use scene classification using multi-scale completed local binary patterns. *Signal Image Video Process.* **2016**, *10*, 745–752. [CrossRef]
14. Spencer, B.F., Jr.; Hoskere, V.; Narazaki, Y. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* **2019**, *5*, 199–222. [CrossRef]
15. Cha, Y.J.; Choi, W.; Suh, G.; Mahmoudkhani, S.; Büyüköztürk, O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 731–747. [CrossRef]
16. Kantsepolsky, B.; Aviv, I. Sensors in Civil Engineering: From Existing Gaps to Quantum Opportunities. *Smart Cities* **2024**, *7*, 277–301. [CrossRef]
17. Dong, C.Z.; Catbas, F.N. A review of computer vision-based structural health monitoring at local and global levels. *Struct. Health Monit.* **2021**, *20*, 692–743. [CrossRef]
18. Shirzad-Ghaleroudkhani, N.; Gül, M. An enhanced inverse filtering methodology for drive-by frequency identification of bridges using smartphones in real-life conditions. *Smart Cities* **2021**, *4*, 499–513. [CrossRef]
19. Sun, L.; Shang, Z.; Xia, Y.; Bhowmick, S.; Nagarajaiah, S. Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *J. Struct. Eng.* **2020**, *146*, 04020073. [CrossRef]
20. Bao, Y.; Li, H. Machine learning paradigm for structural health monitoring. *Struct. Health Monit.* **2021**, *20*, 1353–1372. [CrossRef]
21. Modarres, C.; Astorga, N.; Droguett, E.L.; Meruane, V. Convolutional neural networks for automated damage recognition and damage type identification. *Struct. Control. Health Monit.* **2018**, *25*, e2230. [CrossRef]
22. Gao, Y.; Mosalam, K.M. Deep transfer learning for image-based structural damage recognition. *Comput.-Aided Civ. Infrastruct. Eng.* **2018**, *33*, 748–768. [CrossRef]
23. Gulgec, N.S.; Takáč, M.; Pakzad, S.N. Convolutional neural network approach for robust structural damage detection and localization. *J. Comput. Civ. Eng.* **2019**, *33*, 04019005. [CrossRef]
24. Zhang, C.; Chang, C.C.; Jamshidi, M. Concrete bridge surface damage detection using a single-stage detector. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 389–409. [CrossRef]
25. Zhou, Q.; Ding, S.; Qing, G.; Hu, J. UAV vision detection method for crane surface cracks based on Faster R-CNN and image segmentation. *J. Civ. Struct. Health Monit.* **2022**, *12*, 845–855. [CrossRef]
26. Shokri, P.; Shahbazi, M.; Nielsen, J. Semantic Segmentation and 3D Reconstruction of Concrete Cracks. *Remote Sens.* **2022**, *14*, 5793. [CrossRef]
27. Zhang, A.; Wang, K.C.; Fei, Y.; Liu, Y.; Chen, C.; Yang, G.; Li, J.; Yang, E.; Qiu, S. Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **2019**, *34*, 213–229. [CrossRef]
28. Zhao, J.; Hu, F.; Qiao, W.; Zhai, W.; Xu, Y.; Bao, Y.; Li, H. A modified U-Net for crack segmentation by Self-Attention-Self-Adaption neuron and random elastic deformation. *Smart Struct. Syst.* **2022**, *29*, 1–16.
29. Xu, Y.; Fan, Y.; Li, H. Lightweight semantic segmentation of complex structural damage recognition for actual bridges. *Struct. Health Monit.* **2023**, *22*, 3250–3269. [CrossRef]
30. Pan, Y.; Zhang, L. Dual attention deep learning network for automatic steel surface defect segmentation. *Comput.-Aided Civ. Infrastruct. Eng.* **2021**, *37*, 1468–1487. [CrossRef]
31. Cui, X.; Wang, Q.; Li, S.; Dai, J.; Xie, C.; Duan, Y.; Wang, J. Deep learning for intelligent identification of concrete wind-erosion damage. *Autom. Constr.* **2022**, *141*, 104427. [CrossRef]
32. Xu, J.; Gui, C.; Han, Q. Recognition of rust grade and rust ratio of steel structures based on ensembled convolutional neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 1160–1174. [CrossRef]
33. Li, D.; Xie, Q.; Gong, X.; Yu, Z.; Xu, J.; Sun, Y.; Wang, J. Automatic defect detection of metro tunnel surfaces using a vision-based inspection system. *Adv. Eng. Inform.* **2021**, *47*, 101206. [CrossRef]
34. Snell, J.; Swersky, K.; Zemel, R. Prototypical networks for few-shot learning. *Adv. Neural Inf. Process. Syst.* **2017**, 4080–4090.
35. Fort, S. Gaussian prototypical networks for few-shot learning on omniglot. *arXiv* **2017**, arXiv:1708.02735.
36. Ji, Z.; Chai, X.; Yu, Y.; Pang, Y.; Zhang, Z. Improved prototypical networks for few-shot learning. *Pattern Recognit. Lett.* **2020**, *140*, 81–87. [CrossRef]
37. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135.
38. Nichol, A.; Schulman, J. Reptile: A scalable meta learning algorithm. *arXiv* **2018**, arXiv:1803.02999.
39. Sun, Q.; Liu, Y.; Chua, T.S.; Schiele, B. Meta-transfer learning for few-shot learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 403–412.
40. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
41. Mehrotra, A.; Dukkipati, A. Generative adversarial residual pair-wise networks for one shot learning. *arXiv* **2017**, arXiv:1703.08033.
42. Rezende, D.; Danihelka, I.; Gregor, K.; Wierstra, D. One-shot generalization in deep generative models. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 June 2016; pp. 1521–1529.
43. Guo, J.; Wang, Q.; Li, Y.; Liu, P. Façade defects classification from imbalanced dataset using meta learning-based convolutional neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **2020**, *35*, 1403–1418. [CrossRef]
44. Dong, H.; Song, K.; Wang, Q.; Yan, Y.; Jiang, P. Deep metric learning-based for multi-target few-shot pavement distress Classification. *IEEE Trans. Ind. Inform.* **2021**, *18*, 1801–1810. [CrossRef]

45. Xu, Y.; Bao, Y.; Zhang, Y.; Li, H. Attribute-based structural damage identification by few-shot meta learning with inter-class knowledge transfer. *Struct. Health Monit.* **2021**, *20*, 1494–1517. [CrossRef]
46. Cui, Z.; Wang, Q.; Guo, J.; Lu, N. Few-shot classification of façade defects based on extensible classifier and contrastive learning. *Autom. Constr.* **2022**, *141*, 104381. [CrossRef]
47. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015; Part III 18; pp. 234–241.
48. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Jorge Cardoso, M. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, Held in Conjunction with MICCAI, Québec City, QC, Canada, 14 September 2017; Proceedings 3; pp. 240–248.
49. Wang, Y.; Jing, X.; Xu, Y.; Cui, L.; Zhang, Q.; Li, H. Geometry-guided semantic segmentation for post-earthquake buildings using optical remote sensing images. *Earthq. Eng. Struct. Dyn.* **2023**, *52*, 3392–3413. [CrossRef]
50. Wang, Y.; Jing, X.; Cui, L.; Zhang, C.; Xu, Y.; Yuan, J.; Zhang, Q. Geometric consistency enhanced deep convolutional encoder-decoder for urban seismic damage assessment by UAV images. *Eng. Struct.* **2023**, *286*, 116132. [CrossRef]
51. Barkhordari, M.S.; Armaghani, D.J.; Asteris, P.G. Structural damage identification using ensemble deep convolutional neural network models. *Comput. Model. Eng. Sci.* **2023**, *134*, 835–855. [CrossRef]
52. Xu, Y.; Fan, Y.; Bao, Y.; Li, H. Few-shot learning for structural health diagnosis of civil infrastructure. *Adv. Eng. Inform.* **2024**, *62*, 102650. [CrossRef]