

Article

# A Deep-Learning-Based Bridge Damaged Object Automatic Detection Model Using a Bridge Member Model Combination Framework

Sung-Sam Hong<sup>1</sup>, Cheol-Hoon Hwang<sup>2</sup>, Su-Wan Chung<sup>3</sup> and Byung-Kon Kim<sup>3,\*</sup><sup>1</sup> Department of Multimedia Contents, Jangan University, Hwaseong-si 13557, Gyeonggi-do, Republic of Korea<sup>2</sup> Department of Research & Development, Rabahgroov Corporation, Yongin-si 16942, Gyeonggi-do, Republic of Korea<sup>3</sup> Department of Future and Smart Construction Research, Korea Institute of Civil Engineering and Building Technology (KICT), Goyang 10223, Gyeonggi-do, Republic of Korea

\* Correspondence: bkkim@kict.re.kr

**Featured Application:** Bridge Maintenance Management, Buildings and Structures Maintenance Management.

**Abstract:** More bridges today require maintenance with age, owing to increasing structural loads from traffic and natural disasters. Routine inspection for damages, including in the aftermath of special events, is conducted by experts. To address the limitations of human inspection, deep-learning-based analysis of bridge damage is being actively conducted. However, such models exhibit deteriorated performance in classifying multiple classes. Most existing algorithms do not use in situ images. Hence, the results of the model training do not accurately reflect the actual damage. This study utilizes an extant method and proposes a new model of combination training by bridge member. By integrating the two approaches, we propose a bridge damaged-object-detection deep-combination framework (BDODC-F). To ensure variety in the type of damaged objects and enhanced model performance, a deep-learning-based super-resolution module is employed. For performance improvement and optimization, a deep-learning combination model based on individual training by bridge member is proposed. The BDODC-F improved the mean average precision by 191.6% and 112.21% in the combination model. We expect the framework to aid engineers in the automated detection and identification of bridge damage.

**Keywords:** deep learning; aging bridge management; image analysis; image processing; classifier combination



**Citation:** Hong, S.-S.; Hwang, C.-H.; Chung, S.-W.; Kim, B.-K. A Deep-Learning-Based Bridge Damaged Object Automatic Detection Model Using a Bridge Member Model Combination Framework. *Appl. Sci.* **2022**, *12*, 12868. <https://doi.org/10.3390/app122412868>

Academic Editor: Giuseppe Lacidogna

Received: 1 November 2022

Accepted: 11 December 2022

Published: 14 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Civil infrastructure is regularly inspected to prevent accidents after structural damage. However, the structures requiring maintenance outnumber the available maintenance personnel, leading to difficulties in the systematic maintenance of these facilities [1–3]. Bridges are an integral part of the transportation system as they improve road connectivity over inaccessible terrain. These structures carry heavy vehicular loads throughout their service life, which may result in premature or end-of-life structural failure and a safety risk to the public. Therefore, regular and meticulous monitoring of bridges is imperative.

In most cases, including special events such as extreme weather and disasters, personnel routinely inspect bridges to detect and identify damage [4]. This type of manual inspection is expensive and requires professional input on bridge damage. Damage detection is also affected by the parameters of devices used to collect image data, such as the camera angle, and the location. These limitations hinder the achievement of a reliable level of detection and identification accuracy.

To address these problems, active research using various approaches is underway. The early stages of research were aimed at detecting cracks on the surface of bridges, because most of the other types of damage to aging bridges result from these cracks. Thus, algorithms such as canny and crack-forest and machine-learning techniques such as graph convolutional network (GCN) were used to detect crack edges [2]. More recently, intelligent bridge testing systems and inspection technologies exploiting unmanned aerial vehicles and image-processing techniques have been developed [1,4–6]. However, these are still fledgling technologies. Technologies for identifying different shapes and segmentations in captured images require further development. Currently, professional input is still required for manually identifying the damage type through visual inspection of captured images. Although studies have implemented deep learning (DL) algorithms [7–9], many of them were based on precise measurement at the laboratory scale, or their results were not commercially viable owing to the limited availability of data. Furthermore, as these studies did not use field-inspection images, the results of their model training did not accurately reflect the real-world bridge damage.

Recently, studies have investigated techniques to assess facility structures using DL models [1,2,5,8–11]. The convolutional neural network (CNN) has reportedly exhibited excellent performance in classifying and detecting objects using training image datasets. Hoskere et al. applied deep convolutional neural networks (DCNN) to detect six types of damage, including cracks, concrete scaling, and corrosion, in facility structures [11]. In the Republic of Korea, studies have utilized AlexNet for the detection of concrete spalling [8] and the analysis of railway bridge damage using faster regions with a convolutional neural network (Faster-RCNN) [1]. The limitations of DL-based algorithms are (1) the variation of detection performance with the training image, and (2) difficulty in classifying damage types when the classes have few features or when the features are unclear. Owing to the nature of DL technology, the greater the volume of data for training, the higher the classification accuracy. In terms of the dependence of classification performance on the clear representation of features, for the classification of damage types for bridges, into categories such as efflorescence, corrosion, cracks, concrete scaling, and spalling, the distinction between efflorescence, which appears white in color on the bridge, and cracks is highly accurate as they appear distinct, whereas the classification between scaling, spalling, and cracks is inaccurate as these are similar types of damage.

A previous study proposed a preprocessing technique for training images [3]. It allowed for the use of low-resolution bridge images from collected data for training. The method obtained images suitable for the learning model by employing super-resolution (SR) for normalization and data augmentation. The technique enhanced the resolution of the images, and through proper labeling, could detect small objects, achieving similar or superior performance to the existing bridge damage detection models.

In this study, we utilized that method and proposed a new model of combination training with bridge members. Subsequently, by integrating the two approaches, we proposed the bridge damaged-object-detection deep-combination framework (BDODC-F). In Step 1, to ensure diversity in the type of damaged objects and model performance improvement, a DL-based SR module was employed to improve the performance of individual models. In Step 2, for better performance and optimization of the integrated models, we proposed the bridge damage-detection DL combination model by individually training the models with bridge members. To accurately reflect the similarities and differences of shapes and features between the classes to be classified, the model was trained with bridge members such as abutment and slab for the damage type. Then, a combination model structure was designed, resulting in a single DL combination model. Consequently, to maximize the model detection performance, a framework was proposed to encompass the entire process of data input, dataset construction, training, model deployment, and detection. The proposed framework exhibited improved accuracy, i.e., the mean average precision (mAP), by 191.6% in image quality enhancement and 110.6% in the combined model.

The contributions of this study include:

- a framework capable of automated bridge damage safety diagnosis through image capture,
- an SR-based image pre-processing process suitable for identifying bridge damaged objects,
- a damaged-object-detection model framework optimized for each member of the bridge,
- research and development of a damaged-object-detection model with high accuracy.

The features of the proposed framework can be used to construct a system that can automatically identify damaged objects in situ in real time by collecting images through mobile devices based on the proposed framework.

The structure of the rest of the paper is organized as follows. In Section 2, the extant literature related to this work is reviewed. Section 3 presents the theoretical and experimental aspects of the proposed technology. Section 4 describes the results of the experiments conducted to evaluate the performance of the technology. Section 5 concludes the research with a summary of the study, limitations, and outlook.

## 2. Background

### 2.1. Existing Work on Bridge Damage

A previous study [8] used a web scraping technique to collect various concrete images (damaged or intact concrete images) for training an artificial neural network (ANN), and thus constructed a training image dataset. AlexNet—a DL-based image classifier—was utilized as the initial model to develop an ANN that enabled the automatic detection of spalling in an image using data learned by applying the concept of transfer learning. This model detected concrete scaling/spalling.

In a study on the automatic damage analysis of railway bridges based on a UAV and a DL model [1], damage types such as cracks, scaling, spalling, water leakage, and reinforced concrete (RC) exposure were detected using a deep-learning model. For training image data, approximately 26,476 images of a level C bridge structure or lower with several defects were collected using a camera. As conventional image-processing techniques, such as pattern recognition and edge detection, erroneously detected defects due to contamination of the damaged surfaces, the study applied a DL-based model. Tools such as sharpening, blurring, and hue/saturation were used to perform image preprocessing, and DL algorithms such as Light Head-RCNN, FPN, PSPNet, and U-Net were used for the investigation. Out of approximately 26,476 images of bridge damage, 80% were randomly selected for model training, and the remaining 20%, 5295 images, were used for testing. When U-Net was used, the detection recall was as follows: cracks 96%, concrete scaling/spalling 98%, rebar exposure 98%, and water leakage 92%. In the study, similar types of damage were grouped into a single type, and one model was trained with multiple types of damage.

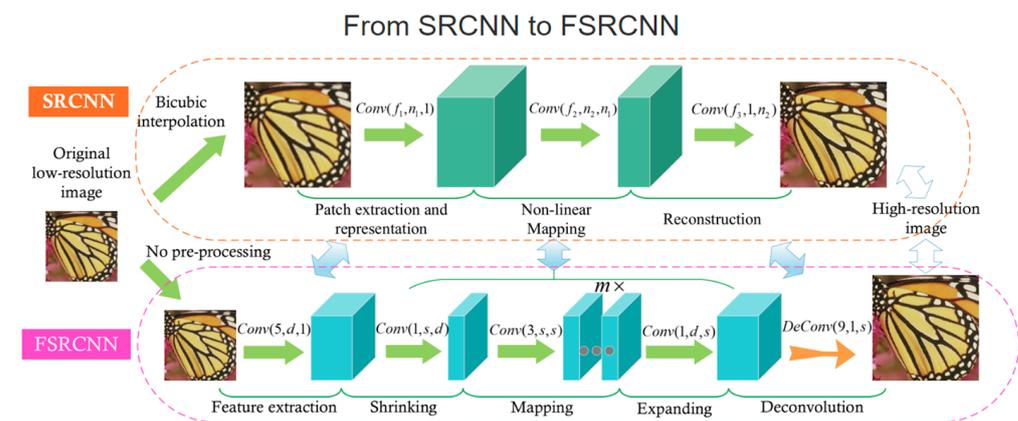
Another study that used DL to evaluate the condition of a bridge [2] highlighted its limitations in identifying multiple types of bridge damages, such as cracks, spalling, and corrosion. To overcome this, methods of training a model for a specific type of damage and applying techniques such as clustering or building a large amount of training data, including a wide range of damage types for classification, were considered as solutions. The study employed a data augmentation technique that increased the volume of training data by applying transformation methods such as image segmentation and rotation, along with transfer learning, to address the problem. As for the DL models used, Mask R-CNN and U-Net were compared in terms of model performance. For training image data, up to  $1024 \times 1024$ -pixel images were used, and the sliding window method was used to address the problems of damage exposure and quantification. Through experiments, Mask R-CNN was determined suitable for damage detection, as the input size was  $1000 \times 800$  pixels, which is relatively large. The model also allowed transfer learning with a limited volume of training data. A final total of 5140 images were used for training in the experiments, and a precision of 95.2% and recall of 93.8% were obtained. The study presented a solution to the limitations in classifying multiple types of damage by increasing the volume of training data, which exhibited robust performance. In contrast, our approach presents a method of model combination after classifying into each type of bridge member. Because

our approach creates a detection model optimized for each bridge member, high accuracy of detection can be expected.

## 2.2. Super-Resolution Convolution Neural Network (SRCNN)

The super-resolution convolution neural network (SRCNN) was the first model that applied DL in the field of super-resolution [12]. It used traditional bicubic interpolation was used for low-resolution images to increase their size to that of high-resolution images. The model involved passing through a convolution network thrice to enhance image quality. Traditional bicubic interpolation estimates the pixel value from the weighted average of the 16 closest neighboring pixels to expand the image pixels. The technique improved image quality while passing through a CNN during training [13].

SRCNN is methodologically simple, yet it performs well and is the most basic DL model in the field of SR. As it is based on a CNN model, it requires supervised learning; that is, it requires high-resolution images for training. The model is trained using low-resolution images and ground-truth (GT) images. When trained to minimize the difference between the unscaled image and the GT image, it outputs a high-resolution image. In a recent study, SRCNN was redesigned as a faster SRCNN (FSRCNN) [14] with accelerated speed and superior restoration quality. The new structure is illustrated in Figure 1.



**Figure 1.** Super-resolution convolution neural network (SRCNN) model structure.

## 2.3. Classifier Combination

Classifier combination is an effective method to improve the performance of pattern recognition [15]. It has been applied in various fields, including economic decision-making, forecasts of natural phenomena, and military decisions in national defense. The combinations can be categorized into mathematical and behavioral approaches. Mathematical approaches construct models and derive combination rules using logic and statistics to combine different classification models and outputs. Behavioral approaches assume discussions between experts in the applicable fields to combine multiple decisions into a single one.

Classifier combination has garnered great interest in the field of optical character recognition and biometric applications in the last decade. It has been widely applied in image classification for text recognition, voice recognition, facial image recognition, fingerprint recognition, and biometric applications. The combination function accepts N-dimensional vectors from M classifiers and outputs N final classification scores. Results for the problem can be finally redefined through combination rules or model definition. The combination allows for the determination of the optimal model, thus minimizing the misclassification cost. Several different types of classifier combinations exist, e.g., score combination function, ensembles, operating level of classifiers, classifier combination based on output types, and classifier combination based on complex types. A study proposed a feature combination network model [16] (Figure 2) in the field of DL to improve the model prediction accuracy.

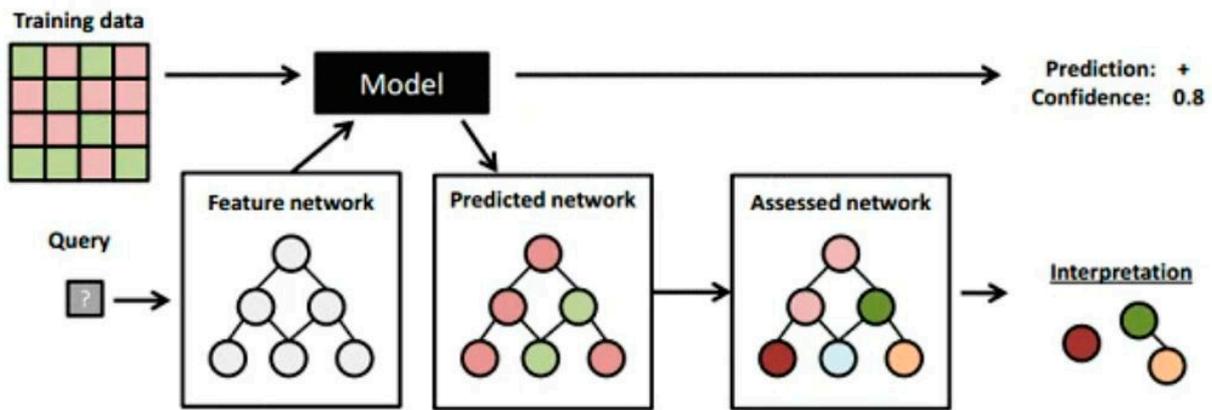


Figure 2. Deep-learning (DL) combination framework.

### 3. Materials and Methods

We proposed a DL framework to detect damages in bridges. In Step 1, the image quality was enhanced and normalized through SR to improve the detection performance by object, achieve more diversity in the dataset, and increase consistency. In Step 2, optimized detection member-specific models are developed through a bridge-damage-detection DL combination module based on individual learning by bridge members. These were aggregated into a single model, thus presenting an optimized bridge-damage-detection model. The proposed framework is a DL framework optimized for six types of bridge damage: efflorescence, concrete scaling, concrete spalling, crack, corrosion, and water leak. A schematic of the framework is depicted in Figure 3. We present a detailed description for each step in the following subsections.

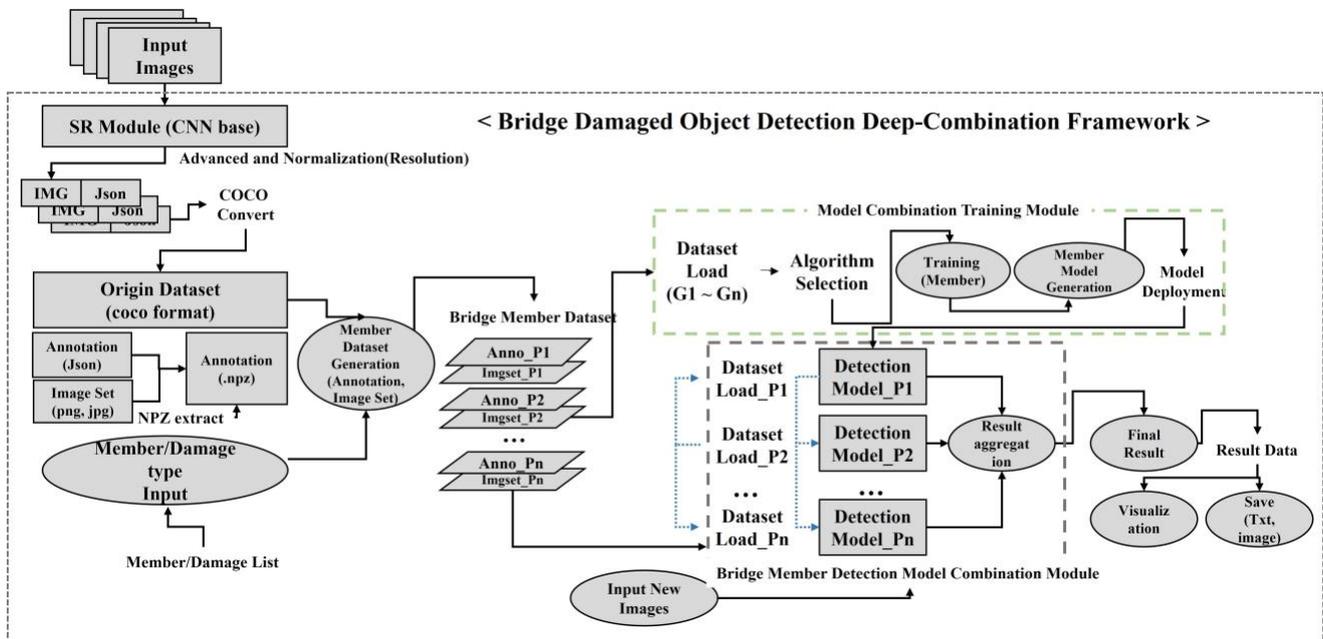


Figure 3. Proposed bridge damaged-object-detection deep-combination framework (BDODC-F).

#### 3.1. Step 1: Image Quality Enhancement Module through SR

In the preprocessing stage (Figure 4), an SR model trained using low- and high-resolution image training datasets was constructed to construct a model for image quality enhancement. Finally, the model was deployed. The deployed model received low-resolution images as input and converted them into high-resolution ones through the SR process (Figure 4). In the damaged-object-detection stage, the preprocessed images were

labeled using a labeler. The labeled images (training set) were used to train the detection model. The detection model was deployed, and when an image captured from field inspection was input to it, the high/low-resolution condition of input data was determined and appropriately preprocessed. The preprocessed image was input to the detection model, and a damaged object in the image was automatically detected. In the case of the CNN model, which was applied in the detection model, the detection performance improved when the image size was fixed and as the filter size increased. Thus, we considered that the proposed process could develop an optimized model to detect and identify damaged objects. The bridge damage was allotted to six classes: efflorescence, water leak, concrete scaling, concrete spalling, cracks, and RC corrosion. To construct the bridge-damage image net, a labeler was used to label damaged objects in the images. The detection accuracy can be improved by enhancing the image quality.

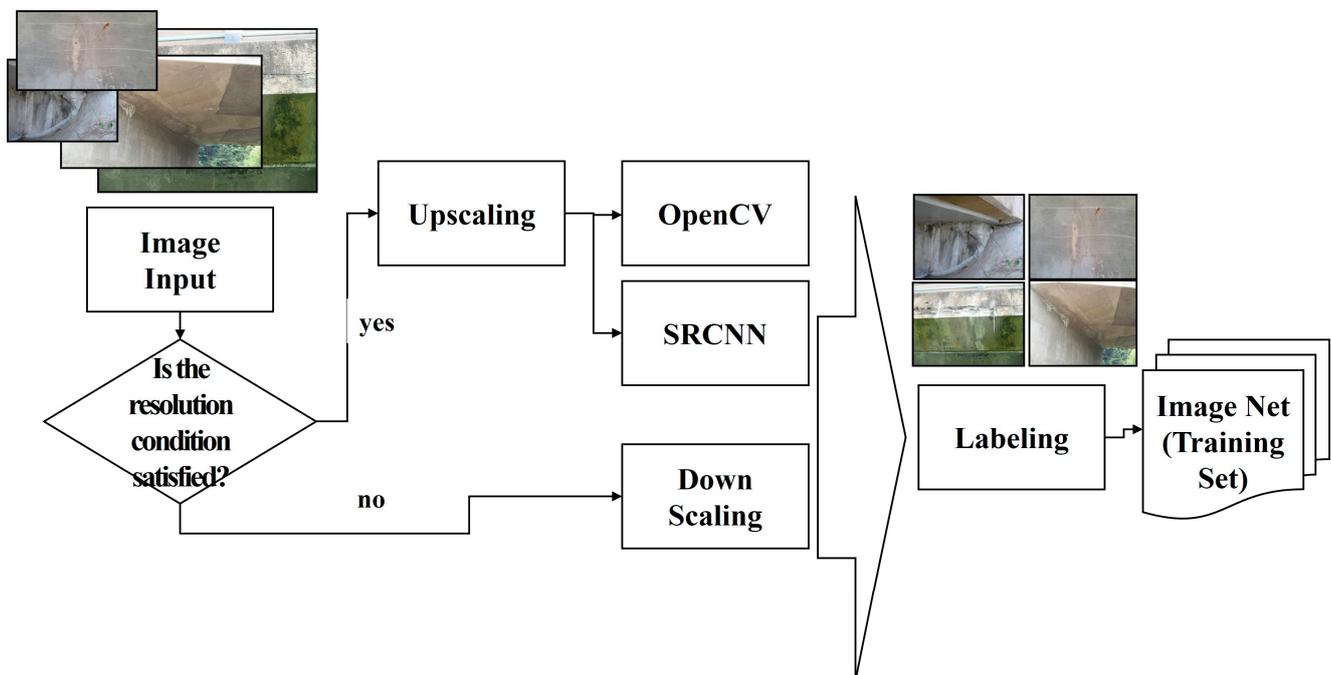


Figure 4. SR process.

SR improves the image quality. The first objective of the proposed framework was to normalize the various types of input images. This is because the DL model performs well when the input image is consistent. Second, pixel data were detailed owing to image expansion, and feature data for learning were precise. Detection performance might be improved using an image extended through SR, with small objects in an image that were difficult for the model to detect.

### 3.2. Step 2: DL Combination Model by Member

#### 3.2.1. Dataset Extraction by Member

In the step of dataset extraction by member (Figure 5), data to be used to train the DL model underwent optimization and preprocessing. The collected bridge images were labeled according to the type of damage and member, and based on the annotation, the labeled image data were extracted into the npz file format for transformation into a data type suitable for the graphic processor unit (GPU). Labeling was performed using Labelme [17], an open-source annotation tool. Thus, the final member dataset in the form of an image set with annotations was constructed.

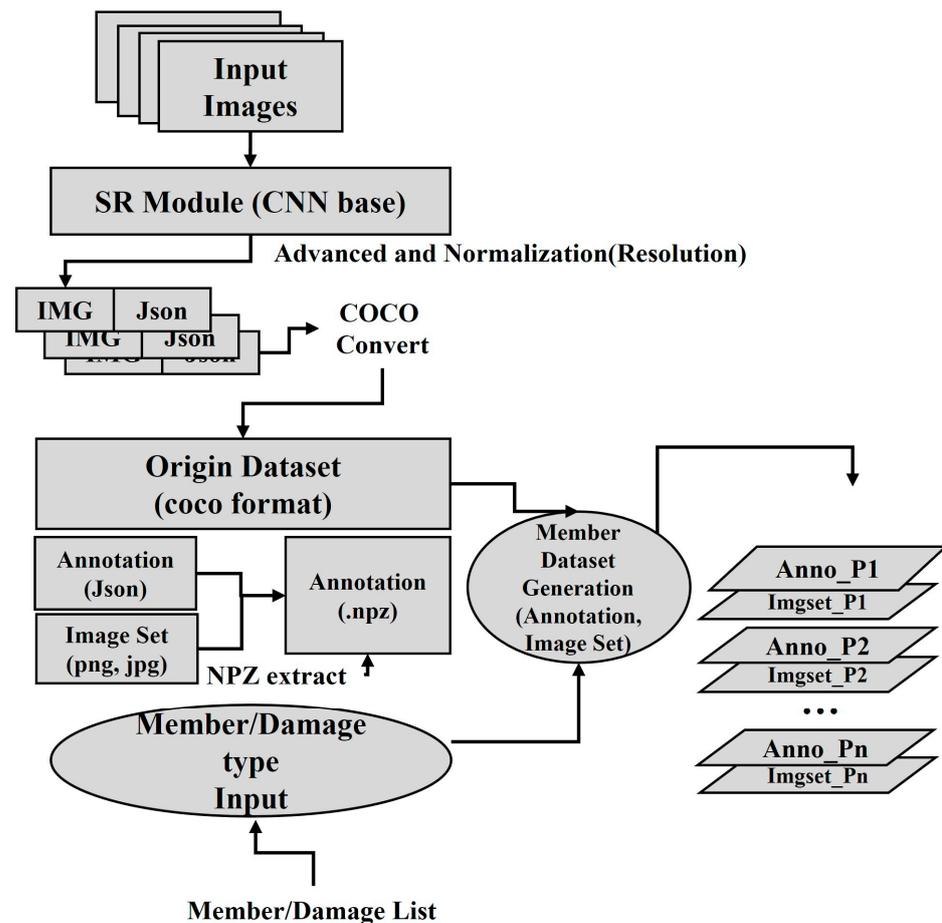


Figure 5. Dataset generation process.

### 3.2.2. Combination Training and Detection Module

In this step (Figure 6), a DL model optimized for detection was developed by analyzing the features of the bridge damage by member. Through effective training using the member-specific image networks constructed as above, the derived model was optimized to develop a model with excellent detection performance. Mask R-CNN [18,19], Yolo (R-CNN) [20,21], and Blendmask [22], which are DL models based on R-CNN, were used as models for the detection of damaged objects. This is because these models were reported to exhibit rapid detection and large input image size. We combined it with high-resolution image optimization and normalization through image-quality enhancement using the upscaling technique described in a previous study [3]. The detection model developed after completing the training was applied to detect damaged objects. As new bridge images were input, the model automatically detected the damaged objects. Based on the measured values of processing speed and accuracy, the optimal model was used as the detection model. DL models based on the R-CNN experience detection speed limitations due to a bottleneck in the process caused by separation between the object region proposal and detection processes, and the scale of network parameters increases. Therefore, a high-performance GPU was required to ensure maximum detection speed [3]. The combination method of the proposed model is not a fusion method between network layers or a combination method of features. Even with the same type of damage (e.g., white stain), the shapes that appeared in the images were distinct; therefore, we constructed optimized models for each member. Members with similar feature maps were grouped to build the optimal model. For one image input, the model for each member identified the damaged object in the image and derived the result through aggregation.

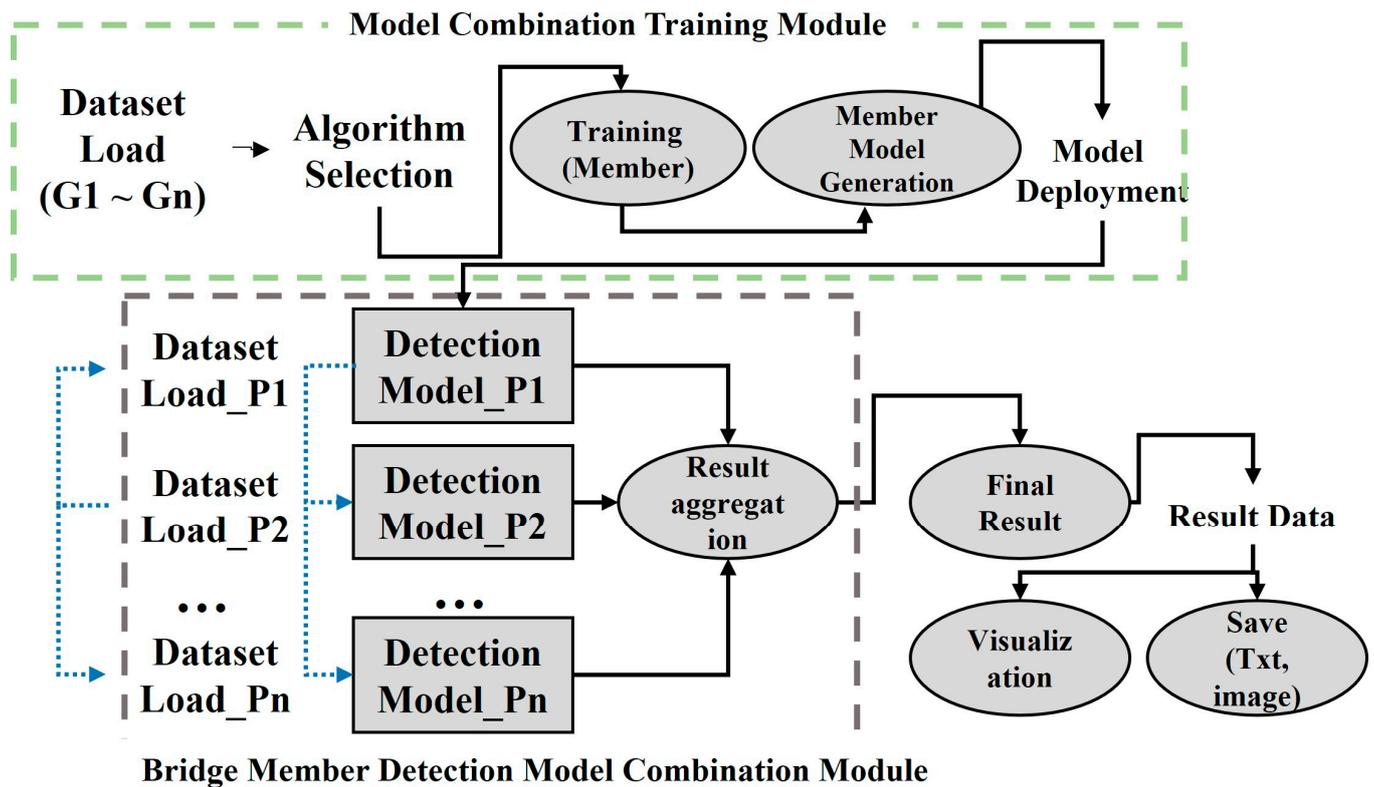


Figure 6. Concept of combined detection model framework by members.

The number of detection models trained with each bridge member was  $n$ , which is the number of members, and these models were aggregated through an aggregation module. The member-specific detection models were trained using member-specific datasets. When an image was input, it passed through the combination detection model framework by member, and the final combined detection result was output.

## 4. Experiments and Results

### 4.1. Experimental Setup

The system setup used in the experiment is as follows. To test the performance of the developed detection model, experiments were carried out on a system with a built-in GPU. Python was used as the programming language.

- CPU: Intel(R) Core(TM) i7-10900k CPU 2.90 GHz
- RAM: 96 GB
- GPU: NVIDIA GeForce RTX3090

### 4.2. Results of Image Quality Enhancement

The parameters of the model used in the experiment are summarized in Table 1. The experiments were performed using a Resnet-based model with a backbone depth of 101 layers. The detection model needs pixel-based detection for each damage type, so it must be detected with instance segmentation. We studied the optimal model based on mask-RCNN [18], which is a popular method in instance segmentation, and Blendmask capable of real-time detection.

**Table 1.** SRCNN, Blendmask, and Mask-RCNN main parameters.

SRCNN	Blendmask	Mask-RCNN
epoch: 300 batch size: 100 loss rate: $1 \times 10^{-4}$	BACKBONE: NAME: "build_fcos_resnet_fpn_backbone" DEPTH: 101 ROI_HEADS: BATCH_SIZE_PER_IMAGE: 512 SOLVER: BASE_LR: 0.01 BIAS_LR_FACTOR: 1.0 MOMENTUM: 0.9 WARMUP_METHOD: linear WEIGHT_DECAY: 0.0001	MODEL: BACKBONE: NAME: "build_resnet_fpn_backbone" DEPTH: 101 ROI_HEADS: NAME: "StandardROIHeads" IN_FEATURES: ["p2", "p3", "p4", "p5"] ROI_BOX_HEAD: NAME: "FastRCNNConvFCHead" NUM_FC: 2 POOLER_RESOLUTION: 7 ROI_MASK_HEAD: NAME: "MaskRCNNConvUpsampleHead" NUM_CONV: 4 POOLER_RESOLUTION: 14 SOLVER: IMS_PER_BATCH: 4 BASE_LR: 0.01

#### 4.2.1. Measurement Method

- Measurement of SR performance

In the blind/referenceless image spatial quality evaluator (BRISQUE) [23], when mean subtraction and contrast normalization (MSCN) is applied, the histogram of pixels follows a Gaussian distribution, the generalized Gaussian distribution is mapped to the image histograms that undergo MSCN processing, and the shape information is used as features to assess the image quality.

The structural similarity index (SSIM) [24] is a representative full-reference image-quality assessment method designed to evaluate the difference and similarity between the image quality of human visual perception, and not the numerical errors. For comparison of an original image A and a distorted image B, the SSIM compares the luminance, contrast, and structure of the two images. A value close to  $-1$  indicates high image quality, and a value close to 0 indicates low image quality.

- Evaluation of model accuracy

To evaluate the model accuracy, we employed general accuracy and average precision (AP), which is a representative measure of the prediction accuracy of DL detection models. By measuring the intersection over union between the ground truth object area and the area predicted by the model, we determined the classification performance according to the ratio of agreement between the prediction and the ground truth. AP 50 indicates that the classification was considered successful when the ground truth and prediction concurred by 50% or more. For AP 75, the ratio of agreement should be 75% or more. The notations APs/APm/APl represent the value of AP measurements for s (small)/m (medium)/l (large) objects, respectively, based on the size of the object to be detected, and success in the classification of classes is determined for all results with 50% to 95% agreement between the ground truth area and the predicted area. mAP represents the average performance of the model based on the mean of all AP values.

#### 4.2.2. Measurement of SR Performance and Experiments and Evaluation for Improved Detection Model Performance

In the first experiment, the SR performance of each model was tested and verified using image sets typically used in SR research. Figure 7 illustrates the five images used in this study. The resolution of each image was  $640 \times 480$ , which was the result of upscaling the resolution by four times through SR and enhancing the image quality to  $2560 \times 1920$ .

As indicated in Table 2, the result of SRCNN exhibits a high SSIM value, while that of the super-resolution generative adversarial network (SRGAN) demonstrated a high BRISQUE value. This result confirmed that an image of quality close to that of visual perception or high-resolution original image was derived.



Figure 7. Typical set of images used in experiment.

Table 2. BRISQUE, SSIM measuring result: typical image set.

Image	With Conversion into High-Resolution Image		With Conversion into High-Resolution Image	
	BRISQUE		SSIM	
	SRCNN	Fast-SRGAN	SRCNN	Fast-SRGAN
0001x4.png	45.537	30.103	0.7779	0.7211
0002x4.png	53.909	30.786	0.6475	0.5849
0003x4.png	56.625	36.472	0.6792	0.6280
0004x4.png	57.018	32.233	0.8819	0.8519
0005x4.png	63.356	50.307	0.7778	0.7326

Next, as illustrated in Figure 8, the SR values of five images selected from the bridge image set collected during field inspection are measured. For this image set, unlike other general image sets, no original high-resolution image exists for comparison. Therefore, in this case, the SSIM could not be calculated. Nevertheless, the image quality was assessed based on the measurement of BRISQUE value alone, which is a method of measuring the image quality without comparison. Figure 9 displays the results of image quality enhancement. Table 3 presents the values of BRISQUE measurement.

Table 3. BRISQUE measuring result: real bridge image.

Image	BRISQUE	
	SRCNN	Fast-SRGAN
001_2018.png	62.016	44.832
002_2018.png	58.563	29.667
003_2018.png	59.600	37.275
004_2018.jpg	51.470	27.746
005_2018.jpg	60.091	39.898

Next, to verify the improvement of the damaged-object-identification performance due to image quality enhancement, the Mask R-CNN model was constructed to perform experiments to detect damaged objects. In this experiment, low- and high-resolution images with image quality enhanced by the SRCNN were used as the input to the damaged-object-identification model. The experiment to identify efflorescent objects was performed among various damaged objects in the image. The parameter values used in the experiment are

summarized in Table 4. The images used in the experiment and experimental results are presented in Table 5. There are a total of five images.



Figure 8. Real bridge image set.

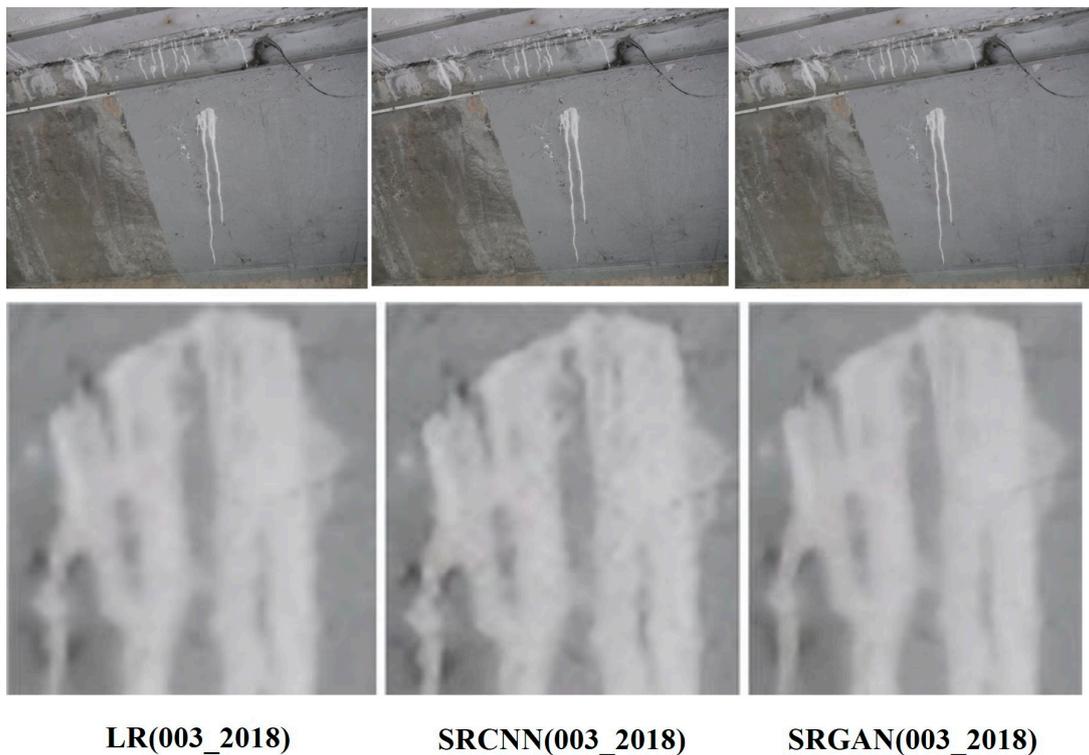


Figure 9. Experimental image result: real bridge image set.

Table 4. Mask R-CNN parameter setting.

Mask R CNN
Epoch: 200, batch size: 100, train: training with images from actual inspection site (low-resolution + high-resolution) [type: coco weight]
Test: low-resolution, SRCNN high-resolution [type: coco weight], loss rate: $1 \times 10^{-4}$

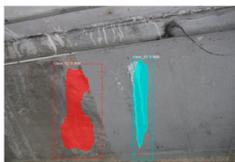
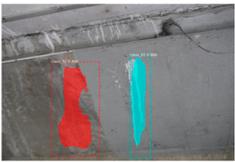
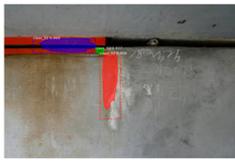
**Table 5.** Bridge damaged-object-detection result.

Image	Image Detection Result [Accuracy]							
	LR				SRCNN			
	DIC	DIA	FDI	TDIA	DIC	DIA	FDI	TDIA
001.jpg	0	0	0	0%	1	[0.9012]	0	100%
002.jpg	2	[0.9230 0.9108]	1	50%	2	[0.9662 0.9060]	1	50%
003.jpg	1	[0.9067]	1	0%	3	[0.9688 0.9558 0.9371]	2	33.3%
004.jpg	2	[0.9230 0.9175]	1	50%	1	[0.9538]	0	100%
005.jpg	1	[0.9464]	0	100%	1	[0.9658]	0	100%

DIC: detected image count, DIA: detected image accuracy, FDI: falsely detected image, TDIA: truly detected image accuracy.

The results of efflorescent object identification in the image are summarized in Table 6. Comparing the accuracy of the experimental results between high- and low-resolution images, the specific type of object identification performance was improved in the image converted to high resolution over the low-resolution image. In terms of the true image detection accuracy for the detection of efflorescence, the high-resolution image obtained using the SRCNN performed vastly better than the LR image without requiring image quality enhancement. Compared with the SRGAN, the result of the SRCNN outputting an image similar in quality to the original is appreciable. The SRGAN is more usable as a data-augmentation model that generates similar images when no original is available. The experiment revealed that the model using the image output by the SRCNN had a high detection accuracy. The size of the image was normalized through quality improvement, and pixel data were refined through pixel expansion, which appeared to improve the detection accuracy of the model.

**Table 6.** Efflorescence object identification result.

Image	Original Image	Image Detection Result		
		LR	SRCNN	Fast-SRGAN
002.jpg				
003.jpg				

Consequently, through the experiment, the results demonstrated that the proposed technology improved the damaged-object-identification performance. In the case of images containing damaged objects with complex structures, the experimental results verified that the identification performance was further improved.

### 4.2.3. Experiment and Evaluation of the Model Performance Optimized by Member

This study focused on six types of damage: efflorescence, concrete scaling, concrete spalling, cracks, water leak, and RC corrosion. However, excluding the objects that were not observed in all members, and those objects that were too few, we analyzed individual models by member for efflorescence, concrete scaling, and concrete spalling. However, the detection performance value (mAP, etc.) represents performance for all types. Models for the road pavement/abutment and pier/slab/rail were constructed to conduct performance testing. Fifty images of each member were used for training, 80 images for the aggregated model, 20 images for validation. For each member, 75,000 epochs were completed, and the performance was evaluated with the derived model for each member. Figure 10 shows the types of bridge members.



Figure 10. Type of bridge members.

To evaluate the improvement in performance after optimization of the respective member-specific models, an existing model trained with the damaged-object images of all members without differentiation by member was used for comparative performance assessment. First, the experimental results of the existing model are presented as follows. Figure 11 shows the result of object detection using the model trained for each type of bridge damage.

Table 7 shows the performance results of the existing model. The existing model exhibited an accuracy of 92.675 (Blendmask), 98.679 (Mask-RCNN) based on AP 50, and 94.013 (Blendmask), 91.372 (Mask-RCNN) for a large object (Apl) and 76.414 (Blendmask), 88.051 (Mask-RCNN) for a small object (APs). The existing model also demonstrated a fair level of overall performance. The overall performance is slightly better with Blendmask.

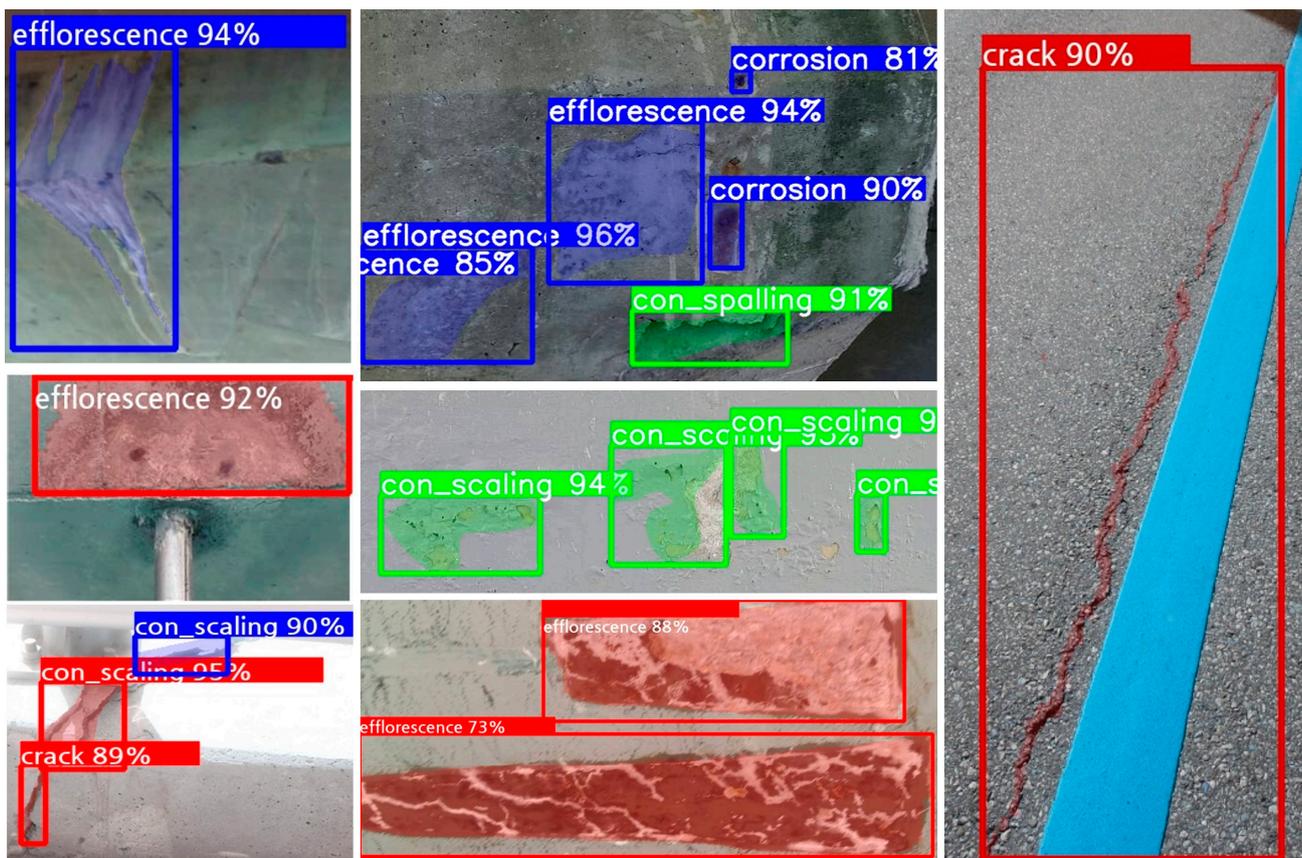


Figure 11. Result image of bridge damaged-object type and detection.

Table 7. Damaged-object-detection performance results for existing models (all members integrated).

Measure	Existing Model (Model without Detection Training by Member)	
	Blendmask	Mask-RCNN
AP50	92.675	98.679
AP75	92.121	96.341
eps	76.414	79.006
Apm	92.508	88.051
Apl	94.013	91.372
mAP	83.965	83.977

Table 8 presents the results of testing the detection performance of the member-specific models. The overall performance indices improved for each model, compared to the existing integrated model. As each member yielded different structures or shapes according to the type of damage, the results confirmed that the proposed model, which was trained separately on individual members followed by model optimization, displayed markedly better performance than the existing model. However, because of the small volume of collected image data, we observed overfitting of the results in a few types of damage. This demerit must be addressed by acquiring sufficient data. In terms of consideration of the performance indices for member-specific models, the value of mAP improved by 11.31 (Blendmask), 8.893 (Mask-RCNN) percent points over the existing model (all members integrated).

**Table 8.** Damaged-object-detection performance results for member specific models.

Measure	Abutment/ Pier Model		Road Pavement		Slab Model		Rail Model		Model Average by Member	
	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN
AP50	92.871	100.00	98.020	91.412	100	98.613	94.257	100.00	96.96	97.506
AP75	92.871	96.923	98.020	100.00	100	93.069	94.257	100.00	96.96	97.498
Aps	70.677	80.387	89.318	100.00	94.718	81.638	80.889	92.885	84.90	88.727
Apm	94.326	88.861	94.685	90.915	97.949	86.380	86.888	97.764	95.65	90.980
Apl	98.389	91.088	97.789	92.429	100	93.676	98.195	98.985	98.73	94.044
mAP	90.470	87.426	91.703	92.446	96.426	86.347	86.409	95.277	92.87	90.374

Table 9 illustrates the detection performance for member-specific models categorized into the four damage types. Overall, the model detected concrete scaling/spalling accurately. For the slab model, the mAP value was higher for the detection of efflorescence than for the detection of abutment/pier or rail, because the shape of efflorescence is more consistent across the images than other damage types. In the case of road pavement, cracks were the only type of damage, and this model cannot detect multiple damaged objects. The shape of the damaged object for each member is distinct. Application of the same model to all members is not an optimal detection approach. As the optimized detection model for each member was constructed through learning, precise detection was possible, and the accuracy performance demonstrably improved by 11.31% (Blendmask), and 8.893% (Mask-RCNN). In experimental model for each member, the overall performance of the Blendmask (mAP 92.87) is better than that of the mask-RCNN (mAP 90.374). In the case of rail and road pavement models, the mask-RCNN performance is higher, so mask-RCNN can be used as the optimal model for this member. In conclusion, the optimal model for each member could be verified through experiments.

**Table 9.** Damaged-object-detection performance results of member specific models by damage type.

Measure (mAP)	Abutment/Pier Model		Road Pavement Model		Slab Model		Rail Model	
	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN	Blendmask	Mask- RCNN
Efflorescence	90.614	88.875	-	-	92.525	91.507	84.050	95.050
Scaling	98.064	88.239	-	-	94.863	91.740	89.894	100.00
Spalling	98.069	92.003	-	-	100	91.740	90.470	91.854
Cracks	100	90.000	91.703	91.412	98.317	71.386	69.802	94.208

The experimental results confirmed that the proposed model exhibited better overall performance in detecting various types of damages. As, in some cases, the occurrence of the type of damage differs with the bridge member, constructing an optimized model for each member could serve as an approach for improving the model performance for the automatic detection of different types of damage.

## 5. Conclusions

The limitations of existing research on automatic detection of bridge damage using DL technology are (1) the training image and (2) the shape of the target object. To address these limitations, a modeling framework based on a DL model with a two-step process was proposed as follows: (1) SR was used to enhance the image quality, thus ensuring diversity in the images of bridge damage and constructing a detection model that enabled the identification of damaged objects with complex structures. (2) Rather than performing detection as a single integrated model, DL models optimized for each member and each type

of damage were derived, which were finally aggregated. Thus, we proposed the BDODC-F, which encompassed all processes, namely data input, dataset construction, model training, detection, and output of results. As the shapes considerably differ with the type of damage for each member, the use of a single model for all members to perform training and detection would limit model performance. Therefore, in this model, we constructed individual models optimized for each member. Further, through these member-specific models, each type of damage was detected, and the results were aggregated. The experimental results demonstrated an improvement in detection performance with an increase in the mAP value of 110.6% over the existing model. However, as the model used an insufficient number of images, more images must be collected to reduce overfitting and improve the model's generalizability. Nevertheless, the performance of the combination model based on member-specific optimization is satisfactory. Thus, the proposed framework might emerge as an intelligent technology for application in the field of damage inspection.

The framework was studied for small-scale bridges with concrete structures that can be photographed by humans. Nonetheless, it can even be applied for suspension and cable-stayed bridges, or large bridges, if the member is a concrete structure.

As the inference time of the proposed model is approximately 0.2 s, real-time detection is possible. However, because it uses approximately 2.5 G of GPU memory (VRAM) per model and 300~350 w power (case of RTX 3090), the model cannot be deployed on a mobile device. However, using an edge computing device (e.g., jetson nano, etc.), it is possible to immediately detect and analyze images input from the image sensor by loading the model (requires 5~10w power). As another approach, the images collected by imaging equipment (smartphone, UAV, etc.) can be transmitted to the network and detected at the DL server. However, network latency must be considered.

Future research can investigate and develop a model capable of improving the instance segmentation performance based on an ensemble of multiple models, as well as a model to automatically detect damaged objects in a bridge optimized through parameter tuning for each member/model. Furthermore, a quantification model that enables automatic estimation of the size of damaged objects can be developed.

**Author Contributions:** Conceptualization, S.-S.H.; Software, S.-S.H. and C.-H.H.; Validation, S.-S.H. and S.-W.C.; Resources, S.-W.C.; Data curation, C.-H.H.; Writing—original draft, S.-S.H. and C.-H.H.; Writing—review & editing, S.-S.H., S.-W.C. and B.-K.K.; Supervision, B.-K.K.; Project administration, B.-K.K.; Funding acquisition, B.-K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** Research for this paper was carried out under the KICT Research Program (project no. 20220217-001, Development of DNA-based Smart Maintenance Platform and Application Technologies for Aging Bridges) funded by the Ministry of Science and ICT.

**Data Availability Statement:** DIV2K dataset—Super Resolution Benchmark Dataset (link: <https://data.vision.ee.ethz.ch/cvl/DIV2K/>, accessed on 30 October 2022).

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

## References

1. Na, Y.H.; Park, M.Y. A study of railway bridge automatic damage analysis method using unmanned aerial vehicle and deep learning-based image analysis technology. *J. Soc. Disaster Inf.* **2021**, *17*, 556–567.
2. Woo-Suk, N.; Jung, H.; Park, K.-H.; Kim, C.-M.; Kim, G.-S. Development of deep learning-based damage detection prototype for concrete bridge condition evaluation. *J. Civ. Environ. Eng Res.* **2022**, *42*, 107–116.
3. Hong, S.-S.; Hwang, C.; Kim, H.-K.; Kim, B.-K. Deep learning-based bridge image pretreatment and damaged objects automatic detection model for bridge damage management. *J. Next-Gener. Converg. Inf. Serv. Technol.* **2021**, *10*, 497–511. [[CrossRef](#)]
4. Zhang, C.; Chang, C.-C.; Jamshidi, M. Bridge damage detection using a single-stage detector and field inspection images. *arXiv* **2018**, arXiv:1812.10590.
5. Young, C.D.; Hyun, P.S.; Kim, Y.K.; Jung, S.W.; Kim, D.-N. Deep-learning crack analysis for visual-safety inspection of bridge by drones. *J. Korean Inst. Inf. Technol.* **2021**, *19*, 115–121.

6. Ellenberg, A.; Kontsos, A.; Bartoli, I.; Pradhan, A. Masonry crack detection application of an unmanned aerial vehicle. In Proceedings of the 2014 International Conference on Computing in Civil and Building Engineering, Orlando, FL, USA, 23–25 June 2014; pp. 1788–1795.
7. Kim, H.; Sim, S.-H.; Cho, S. Unmanned aerial vehicle (UAV)-powered concrete crack detection based on digital image processing. In Proceedings of the 6th International Conference on Advances in Experimental Structural Engineering 11th International Workshop on Advanced Smart Materials and Smart Structures Technology, Urbana, IL, USA, 1–2 August 2015.
8. Lee, Y.-I.; Kim, B.; Cho, S. Image-based spalling detection of concrete structures using deep learning. *J. Korea Concr. Inst.* **2018**, *30*, 91–99. [[CrossRef](#)]
9. Jung, S.; Lee, S.K.; Park, C.; Cho, S.; Yu, J. A method for detecting concrete cracks using deep-learning and image processing. *J. Archit. Inst. Korea Struct. Constr.* **2019**, *35*, 163–170.
10. Kim, A.; Kim, D.; Byun, Y.; Lee, S.W. Crack detection of concrete structure using deep learning and image processing method in geotechnical engineering cracks. *J. Kor. Geotech. Soc.* **2018**, *34*, 145–154.
11. Hoskere, V.; Narazaki, Y.; Hoang, T.; Spencer, B., Jr. Vision-based structural inspection using multiscale deep convolutional neural networks. *arXiv* **2018**, arXiv:1805.01055v.
12. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [[CrossRef](#)]
13. Kim, T.W. CNN, Summary of Convolutional Neural Network. Available online: <http://taewan.kim/post/cnn/> (accessed on 4 January 2018).
14. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016.
15. Tulyakov, S.; Jaeger, S.; Govindaraju, V.; Doermann, D. Review of classifier combination methods. In *Machine Learning in Document Analysis and Recognition*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 90, pp. 361–386.
16. Webb, S.J.; Hanser, T.; Howlin, B.; Krause, P.; Vessey, J.D. Feature combination networks for the interpretation of statistical machine learning models: Application to Ames mutagenicity. *J. Cheminform.* **2014**, *6*, 8. [[CrossRef](#)]
17. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A database and web-based tool for image annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [[CrossRef](#)]
18. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
19. Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A survey of deep learning-based object detection. *IEEE Access* **2019**, *7*, 128837–128868. [[CrossRef](#)]
20. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
21. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
22. Chen, H.; Sun, K.; Tian, Z.; Shen, C.; Huang, Y.; Yan, Y. Blendmask: Top-down meets bottom-up for instance segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 8573–8581.
23. Mittal, A.; Moorthy, A.K.; Bovik, A.C. No-reference image quality assessment in the spatial domain. *IEEE Trans. Imag. Process.* **2012**, *21*, 4695–4708. [[CrossRef](#)] [[PubMed](#)]
24. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Imag. Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]