



# Article Distress Detection in Subway Tunnel Images via Data Augmentation Based on Selective Image Cropping and Patching

Keisuke Maeda <sup>1</sup>, Saya Takada <sup>2</sup>, Tomoki Haruyama <sup>2</sup>, Ren Togo <sup>1</sup>, Takahiro Ogawa <sup>1</sup>, and Miki Haseyama <sup>1,\*</sup>

- <sup>1</sup> Faculty of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-ku, Sapporo 060-0814, Hokkaido, Japan
- <sup>2</sup> Graduate School of Information Science and Technology, Hokkaido University, N-14, W-9, Kita-ku, Sapporo 060-0814, Hokkaido, Japan
- \* Correspondence: mhaseyama@lmd.ist.hokudai.ac.jp

Abstract: Distresses, such as cracks, directly reflect the structural integrity of subway tunnels. Therefore, the detection of subway tunnel distress is an essential task in tunnel structure maintenance. This paper presents the performance improvement of deep learning-based distress detection to support the maintenance of subway tunnels through a new data augmentation method, selective image cropping and patching (SICAP). Specifically, we generate effective data for training the distress detection model by focusing on the distressed regions via SICAP. After the data augmentation, we train a distress detection model using the expanded training data. The new image generated based on SICAP does not change the pixel values of the original image. Thus, there is little loss of information, and the generated images are effective in constructing a robust model for various subway tunnel lines. We conducted experiments with some comparative methods. The experimental results show that the detection performance can be improved by our data augmentation.



# 1. Introduction

The subway networks play an essential role in the current urban infrastructure in transporting millions of commuters in major cities [1–3]. However, it has been reported in [4] that many concrete structures show significant signs of degradation after only 20 to 30 years due to the joint action of mechanical and environmental effects. Many facilities consisting of concrete structures were intensively built during rapid economic growth, and the number of aging facilities has dramatically increased [5,6]. For instance, cracks, as an important type of distress, can influence the stability of the tunnel structure and require periodic inspection to assess subway line structural integrity and ensure traffic safety. Currently, many countries primarily conduct crack detection in subway tunnels by manual inspection [2]. It has been reported that the annual maintenance costs on a global scale have increased to more than 3% of the world's gross domestic product [7]. Furthermore, the inspection time outside operating hours of subways is short, from the end of the working day to the beginning of the working day, and there is the possibility of human errors since engineers have been performing their operations based on their knowledge and experience [3,8,9].

Under these circumstances, advanced infrastructure maintenance techniques that can reduce costs and the burden on engineers are needed [10–13]. Some infrastructure maintenance methods have been proposed based on computer vision with images of infrastructure [14–16]. These methods can be roughly divided into the following three categories: image processing-based, machine learning-based, and deep learning-based methods. According to the characteristics of distress, which include color [17], texture [18], and shape [19], image processing-based methods [20], edge detection [21], minimum path [22], and morphology [23] are useful. Although these methods have the advantage



Citation: Maeda, K.; Takada, S.; Haruyama, T.; Togo, R.; Ogawa, T.; Haseyama, M. Distress Detection in Subway Tunnel Images via Data Augmentation Based on Selective Image Cropping and Patching. *Sensors* **2022**, *22*, 8932. https:// doi.org/10.3390/s22228932

Academic Editor: Gemine Vivone

Received: 22 October 2022 Accepted: 14 November 2022 Published: 18 November 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/). of simplicity and practicability because of the strong correlation with specific tasks, the detection accuracy is low for images of distress due to their complex backgrounds. Combing with image processing technology to extract distressed features, machine learning-based methods use support-vector machines [24,25], neural networks [26–31], clustering [32], and random forests [27,33,34] to detect distresses. These methods can automatically learn the relationships between distressed features and improve detection performance. However, since the features used by these methods are artificially designed and incomplete, the detection performance is limited by the feature extraction ability. Benefiting from the advances of deep learning technologies in recent years [35–37], some methods for detecting distressed regions based on deep learning with images of subway tunnels [38–45] have been proposed. The study [41] proposed a crack detection method based on a deep fully convolutional network (FCN) for the semantic segmentation of concrete crack images. They used concrete crack images collected at various campus buildings of Middle East Technical University [46], but these images were maintained for crack detection, which differs from the characteristics of the data used in actual tunnels. In the method [44], pixel-level crack detection using FCN has also been used to detect cracks in pavement and concrete walls. These studies were conducted on images taken in environments completely different from those of subway tunnels. On the other hand, some methods [45,47] have performed crack detection in subway tunnels, but they focus only on specific lines of tunnels. In the actual situation, there are many subway lines, and the years of operation and operating conditions of each line vary widely. Depending on the subway line to be analyzed, insufficient training data can be collected, or the subway tunnel structure is unsuitable for training the model since the conventional methods are constructed to be specific to a certain subway line. Therefore, it is necessary to construct a model with high generalizability that can consider the diversity of subway lines.

Although several data augmentation methods [48–50] have been recently proposed to improve the model's generalizability, it is essential to select an appropriate data extension method for the training data [51]. Specifically, SamplePairling [49] is a data augmentation method that randomly extracts two images from a set of images and combines them with the same transmittance. Mixup [48] is a data augmentation method that randomly extracts two images from a set of images and combines them with the transmittances determined according to the beta distribution. These methods [48,49] change the original pixel values to augment data. The changes in original pixel values can affect the performance of distress detection. However, random image cropping and patching (RICAP) [50] is a data augmentation method that randomly selects four images from the dataset and combines them to generate a new image. RICAP does not change the original pixel values to augment data. Additionally, RICAP can generate new images containing images of multiple subway lines. In our previous study, we validated the performance of distress detection using each data augmentation method [48–50]. The results showed that there is little loss of information using the data augmentation method that does not change the pixel values of the original images, and the generated data are effective for distress detection. In the study [52], data augmentation was conducted by generating new images based on RICAP [50] to include the regions of multiple lines. Since there is little loss of information in the new images generated based on RICAP, the generated images are effective training images for building a robust model for various subway tunnel lines. The detection performance of the distressed regions has been improved using these images as additional training data. However, most images of distress contain regions where no distressed region exists. In this case, the data augmentation methods, such as RICAP, are likely to generate images that contain many regions where no distressed region exists, limiting the improvement of the distress detection performance. Therefore, there is a possibility that we can generate further effective images for building a robust model for various subway tunnel lines by focusing only on the regions where the distressed regions exist.

In this paper, we propose a new data augmentation method, selective image cropping and patching (SICAP), to accurately detect subway tunnel regions of distress by extracting only the distressed regions from the images of distress from multiple lines and combining them to expand the data. The proposed method can improve the performance of distress detection by expanding the data and focusing only on the distressed regions. We can generate a lot of effective data for training the model for distress detection by focusing on the distressed regions via SICAP. After the data augmentation, we construct a distress detection model based on the deep convolutional neural networks from the expanded training data. The new image generated based on SICAP does not change the pixel values of the original image, and there is little loss of information. Additionally, it is an effective training image for building a robust model for various subway tunnel lines since the expanded data invariably contain distressed regions. Therefore, data augmentation based on SICAP can improve the detection of regions of distress in subway tunnels.

For obtaining robustness, recent studies have focused on the pre-training of models based on self-supervised learning (SSL). SSL is an approach for acquiring knowledge from unsupervised data as model parameters and applying them to learning various downstream tasks. While this approach is basically effective, it is necessary to optimize both the training of the SSL model and the model for the downstream task. In research aimed at real-world applications of AI, such as this research, it is undesirable to train multiple models, since the model must be re-trained each time new data are obtained. From the standpoint of such real-world use, data extensions that directly improve the target task are suitable.

The remainder of the paper is organized as follows. In Section 2, we describe our data augmentation method for detecting distressed regions of subway tunnels. Section 3 presents the experimental results to verify the effectiveness of the proposed method. Finally, Section 4 presents the conclusion. For a smooth explanation of the proposed method, the mathematical variables and symbols used in this paper are shown in Table 1.

Variables	Meanings
n	Index of distress images
m	Index of patches
$I_{\rm orig}^{(n)}$	<i>n</i> -th distress image
$I_{\text{orig}}^{(n,m)}$	<i>m</i> -th patch extracted from $I_{\text{orig}}^{(n)}$
$I_{\mathrm{aug}}^{(s)}$	s-th augmented patches
(w, h)	Boundary position of the augmented patch $I_{aug}^{(s)}$
$w_k, h_k$	Width and height of each region extracted from patch $I_k$
$I_x$ , $I_y$	Width and height of patch $I_{\text{orig}}^{(n,m)}$
R <sup>(n)</sup> <sub>orig</sub>	Ground truth of image $I_{\text{orig}}^{(n)}$

Table 1. Explanation of the mathematical variables and symbols.

### 2. Distress Detection Based on Selective Image Cropping and Patching

# 2.1. Selective Image Cropping and Patching

In this section, we explain the proposed data augmentation method, SICAP. Figure 1 shows the overview of the proposed method. The proposed SICAP selects four images from the dataset, performs region extraction for each image, and combines the extracted images to generate a new image. First, we perform patch segmentation on the image of distress  $I_{\text{orig}}^{(n)}$  ( $n = 1, 2, ..., |\mathcal{I}|$ , where  $\mathcal{I}$  is the set of images of distress  $I_{\text{orig}}^{(n)}$ ), to obtain the set of patches  $I_{\text{orig}}^{(n,m)}$  ( $m = 1, 2, ..., |\mathcal{I}|$ , with  $|\mathcal{I}_{\text{orig}}^{(n,m)}|$  being the set of patches to be partitioned from  $I_{\text{orig}}^{(n)}$ ). Then, we perform data augmentation on patches  $I_{\text{orig}}^{(n,m)}$  by SICAP. Specifically, the proposed method selects four patches  $I_k$  ( $k \in \{1, 2, 3, 4\}$ ) from the set of patches  $|\mathcal{I}^{(n)}|$  that contain distress regions and include multiple types of lines. Next, the boundary points

(w, h) for generating one image from these regions are calculated according to the beta distribution from the following equations:

$$w = \operatorname{round}(w'I_x), \ w' \sim \mathcal{B}(\alpha, \beta), \tag{1}$$

$$h = \operatorname{round}(h'I_y), h' \sim \mathcal{B}(\alpha, \beta),$$
 (2)

where  $I_x$  and  $I_y$  are the width and height of the patch  $I_{\text{orig}}^{(n,m)}$ .  $\alpha \in (0,\infty)$  and  $\beta \in (0,\infty)$  are hyperparameters, and round(·) is a function that performs fractional processing. The width  $w_k$  and height  $h_k$  of each region extracted from the four patches  $I_k$  are determined using w and h as follows:

w

h2

$$w_1 = w_3 \tag{3}$$

$$= w,$$
 (4)

$$_{2} = w_{4}$$
 (5)

$$= I_x - w, (6)$$

$$h_1 = h_3 \tag{7}$$

$$= h,$$
 (8)

$$= h_4 \tag{9}$$

$$= I_y - h. \tag{10}$$

Using  $w_k$  and  $h_k$ , we determine the coordinates  $x_k$  and  $y_k$  for region extraction according to the uniform distribution as follows:

$$x_k \sim \mathcal{U}(0, I_x - w_k),\tag{11}$$

$$y_k \sim \mathcal{U}(0, I_y - h_k). \tag{12}$$

The determination of  $x_k$  and  $y_k$  is repeated until all four regions contain distressed regions. Finally, the four regions extracted based on  $(x_k, y_k)$  and  $(w_k, h_k)$  are joined so that they touch at the boundary point (w, h) to generate a new patch  $I_{\text{aug}}^{(s)}$  (s = 1, 2, ..., |S|, where S is the set of augmented patches). We also apply the same process into the label  $R_{\text{orig}}^{(n)}$  ( $n = 1, 2, ..., |\mathcal{R}|$ , where  $\mathcal{R}$  is the set of the distress label  $R_{\text{orig}}^{(n)}$ ) to obtain a patch  $R_{\text{aug}}^{(s)}$  corresponding to the new patch  $I_{\text{aug}}^{(s)}$ . Consequently, the data augmentation using SICAP for the distressed images of subway tunnels can generate the training data for constructing robust models for various subway tunnel lines.

The previous study [50] claims that RICAP replaces the classification task with the occupancy estimation task by mixing the four class labels with ratios proportional to the areas of the four cropped images. They also claim that RICAP forces the CNN to classify each pixel in a weakly supervised manner, and thereby, the CNN can focus on minor features, partial features, backgrounds, and any other information that is often ignored. Therefore, although there may not be a significant change in the diversity of the data, the data augmentation with RICAP and SICAP will allow a variety of objects to be recognized more efficiently than with the original images alone. Furthermore, pixel-wise blending methods such as mixup [48] superimpose images, which can lead to over-focusing on salient regions in the image. On the other hand, RICAP and our SICAP are spatial blending approaches that can effectively learn the original local features and newly extract global features from the combined images. Thus, this approach avoids overfitting compared to conventional data augmentation methods.



Figure 1. Overview of the proposed method for detecting regions of distress in subway tunnels.

#### 2.2. Distress Detection with DeepLabv2

In this section, we perform the training of DeepLabv2 [53], which detects the distressed region for each input image pixel, using the patches acquired from the images of distress and the corresponding pixel-level distress labels. DeepLabv2 is a model that introduces an Atrous convolution layer into ResNet-101 [36]. The Atrous convolution layer performs wide-area convolution compared to the normal convolution layer to achieve robust features. We replace the classification layer in ResNet-101 with Atrous spatial pyramid pooling (ASPP). ASPP applies the Atrous convolution layer on multiple scales and integrates and pools features spatially. This approach is inspired by the success of the spatial pyramid pooling method [54], which shows that regions of an arbitrary scale can be accurately and efficiently classified by resampling convolutional features extracted at a single scale. ASPP has been implemented as a variant of the scheme that uses multiple parallel Atrous convolutional layers with different sampling rates. The features extracted for each sampling rate are further processed in separate branches and fused to generate the final result. Therefore, DeepLabv2 can calculate the feature map with an Atrous convolution feature layer while maintaining the resolution.

In the training phase, we take a patch-based training strategy since subway tunnel images have high resolution. In the proposed method, we divide subway tunnel images into multiple patches and perform data augmentation based on SICAP. Then, we use the augmented patches  $I_{aug}^{(s)}$  to train our network. The network is initialized with parameters trained on the ImageNet dataset [55]. We perform a multi-class classification, such as

"cracks", "all the others", and "no distress", to detect the distressed region for each pixel of the input image. Since the pixels of the background class that do not contain a distressed region are more than those of the distress classes, the training may be relatively dominated by the background class. The final classification layer of the DeepLabv2 network is a spatial pyramid pooling layer consisting of several classification layers with different dilation rates. Layers with different dilation rates have different receptive field sizes, thus facilitating multiple-scale detection. At the end of the training, we use validation data to observe the training process and fix the parameters of our model. Note that the data are class-unbalanced due to the overwhelming number of background areas. To consider this situation, we added weights to the losses related to crack segmentation during the DeepLabv2 training.

In the test phase, we divide subway tunnel images into multiple patches, input the test patches into the trained model, and estimate the distressed regions. The results of each test patch are integrated into a single image. In this way, we successfully perform distress detection. In the proposed method, we train the model with patches generated based on SICAP so that the patches invariably contain distressed regions. Thus, we can build a robust model for various subway tunnel lines.

#### 3. Experimental Results

## 3.1. Dataset

In this section, we explain the dataset used in our experiments. The dataset used in our experiment consists of distressed images of two lines (lines A and B), and their corresponding distress labels were provided by Tokyo Metro Company Limited. Figure 2 shows some examples of images in the dataset. Since the years of operation and the operating conditions of each line vary, the condition of the tunnels and the color of the images vary widely. Specifically, 44 images of distress were taken for each line (lines A and B). The images were taken as RGB images of approximately  $5000 \times 6400$  pixels. We adopted "crack", "desquamation chipping", "cold joint", "deposition", "masonry joint", and "peeling", as shown in Figure 3. In the images of distress, we have pixel-level labels of distress assigned by technical experts, as shown in Figure 4. Each label in Figure 4 corresponds to the kind of distress in Figure 3.



Figure 2. Examples of image of two lines (lines A and B) in the dataset.





Figure 4. Examples of labels in the dataset, corresponding to Figure 3.

#### 3.2. Settings

In this section, we describe our experimental settings. We conducted the experiment and ablation study to validate the effectiveness of the proposed method. In the experiment, we perform distress detection using SICAP for all kinds of images of distress using two subway lines to verify the robustness of the proposed method for various lines. In the ablation study, we perform distress detection using SICAP for each kind of distress using one subway line to verify the effectiveness of selecting the target distress.

We first explain the experimental settings. We selected 44 images from each of lines A and B. We used five images for validation and five images for the test from the dataset. Since the images of distress were taken with high resolutions, we divided them into patches with a patch size of  $256 \times 256$  pixels and a slide width of 100 pixels. We generated 72,238 and 75,240 patches from the images of distress of lines A and B, respectively. The distress labels used in this experiment were "crack" and "all the others", according to the literature [47]. Note that the "all the other" label includes "desquamation chipping", "cold joint", "deposition", "masonry joint", and "peeling". The DeepLabv2 network was trained to classify each pixel in the input image into three classes: "cracks", "all the others", and "no distress". Note that the hyperparameters  $\alpha$  and  $\beta$  were set to 5.0. In the training phase, we used stochastic gradient descent as the optimizer with a batch size of 16, momentum of 0.9, and weight decay of  $5 \times 10^{-4}$ , and the number of training epochs was four.

In this experiment, we used intersection over union (IoU) as the evaluation index, which is defined by the following equation:

$$IoU = \frac{TP}{TP + FN + FP},$$
(13)

where TP, FN, and FP are the number of true-positive, false-negative, and false-positive samples, respectively.

We used the following comparative methods (CMs1–6) in this experiment to verify the effectiveness of the proposed method.

- CM1: A method using patches obtained from images of distress of lines A and B together without data augmentation.
- CM2: A data augmentation method based on SamplePairing [49] for CM1. Two images are randomly extracted from the set of images and combined with the same transmittance.
- CM3: A data augmentation method based on SamplePairing [49] using only patches that contain regions of distress for CM1 (Selective SamplePairing).
- CM4: A data augmentation method based on Mixup [48] for CM1. Two images are randomly extracted from the set of images to be augmented, and their transmittances are determined according to the beta distribution.
- CM5: A data augmentation method based on Mixup [48] using only patches that contain regions of distress for CM1 (Selective Mixup).
- CM6: A data augmentation method based on RICAP [50] for CM1. This corresponds to the conventional method [52] in our research.

Note that CMs2–5 change the original pixel values when performing data augmentation. The reasons for adopting each comparison method are described below. We verified the effectiveness of introducing data augmentation by comparing ours with CM1. CMs 2, 4, and 6 are common data augmentation methods, as described in the introduction, and we verified their effectiveness as data augmentation methods by comparing ours with these methods. In addition, CMs 3 and 5 are data augmentation methods that introduce the supervised information into CMs 2 and 4, respectively. By comparing the proposed method with these methods, we verified the effectiveness of the proposed method as the data augmentation method with the supervised information.

Next, we explain the settings of the ablation study. The number of pixels in the region of distress in the training data varies significantly for each distress. Therefore, it is expected to improve the final detection performance by reducing the influence of the distressed region other than the detection target by performing SICAP for each arbitrary distress. Therefore, in this experiment, we applied SICAP to each arbitrary distress in the detection of subway tunnel images to verify the detection performance. The distress labels used in this experiment were "crack", "peeling", and "all the others". The DeepLabv2 network was trained to classify each pixel in the input image into four classes: "cracks", "peeling", "all the others", and "no distress". Note that the hyperparameters  $\alpha$  and  $\beta$  were set to 5.0; the number of training epochs was 4, and the batch size was 16. In this experiment, we used IoU as the evaluation index, as in the first experiment. We calculated the IoU of each distress and used the average of the IoU when detecting "crack" and "others" as the evaluation index. We used the following comparative methods (CM1–2) in this experiment to verify the effectiveness of the proposed method.

- CM1: A method using patches obtained from images of distress of line A without data augmentation.
- CM2: A data augmentation method based on RICAP [50] for CM1. This corresponds to the conventional method [52] in our research.

## 3.3. Experiment: Performing SICAP for All Kinds of Distress

In this section, we describe the experimental result and quantitative and qualitative evaluation of the proposed method. Table 2 presents the experimental results of IoU for each data augmentation method when the test data are the images of distress of lines A and B. The number of augmented patches of each line is 1.5 times the original patches. As presented in Table 2, the averages of the IoU of the proposed method are higher than those of the other comparative methods in all cases of the test images from lines A and B. The comparison of the IoU between CM1 and PM suggests the effectiveness of performing the data augmentation for the training images. Additionally, the comparison of the IoU among CMs2–6 and PM suggests the effectiveness of SICAP for data augmentation for

distress detection. Furthermore, the comparison of the IoU of CM6 and PM suggests the effectiveness of focusing only on the regions where distress exists, especially for RICAP, in the data augmentation method. These results show that data augmentation focusing only on distressed regions without changing the original pixel values is effective in detecting subway tunnel distress.

We qualitatively validated the performance of the proposed method. Figures 5 and 6 show the detection results using the proposed method on lines A and B, respectively. The images in the left column are test images used in the experiment. The images in the center column are the pixel-level ground truth corresponding to the test images. The images in the right column are the estimated region using the proposed method. Figures 5 and 6 show that the proposed method can accurately estimate the distressed regions for the most part and can accurately detect a long and bold crack. These results suggest that data augmentation without changing the original pixel values effectively improves the performance of subway tunnel distress detection. Although there are many joints, most of them are not included in the detects horizontal cracks without detecting horizontal joints. The above experimental results confirm the effectiveness of the proposed method.

**Table 2.** Average of IoU when detecting "crack" and "others" based on each data augmentation method (DA method) in experiment I.

			Line A		Line B		
	DA Method	Crack	Others	Average	Crack	Others	Average
CM1	-	0.2600	0.4884	0.3741	0.2278	0.6201	0.4240
CM2	SamplePairing [49]	0.2878	0.3168	0.3023	0.2084	0.5862	0.3973
CM3	Selective SamplePairing	0.2763	0.3872	0.3318	0.1827	0.6471	0.4149
CM4	Mixup [48]	0.2751	0.4443	0.3597	0.2365	0.6073	0.4219
CM5	Selective Mixup	0.2479	0.2514	0.2497	0.1270	0.5254	0.3262
CM6	RICAP [50]	0.2781	0.5177	0.3979	0.2467	0.6229	0.4348
PM	SICAP	0.2983	0.5120	0.4052	0.2171	0.6547	0.4359

We demonstrate the effectiveness of the proposed method through a new evaluation that takes into account real-world applications. Specifically, since one of the most important aspects of tunnel maintenance is the density of cracks, a region-based evaluation is performed. For this evaluation, the image is divided into patches of 50 pixels of slide width and  $200 \times 200$  pixels. If cracks are present in 1% or more of the pixels in each patch, the patch is then considered positive; otherwise, it is considered negative. We adopted the f-value to compare the results of our methods with the ground truth. As an example, Figure 7 shows the correct response region in Figure 5C and the region estimated by the proposed method. This region-based comparison confirms that it is possible to qualitatively and adequately identify the cracked area. Using the f-value as the region-based evaluation in this image is about 80%. Therefore, it was confirmed that the performance is absolutely high quantitatively. This method has already reached the level of practical application and is being considered for introduction into the system of a collaborating company.

#### 3.4. Ablation Study: Performing SICAP for a Certain Kind of Distress

In this section, we describe the results of the ablation study and quantitative and qualitative evaluation of the proposed method. In this experiment, "PM-Crack" and "PM-Peeling" apply SICAP only to the area of "cracks" and "peeling", respectively. "PM-Others" applies SICAP only to the area of "others", including "desquamation chipping", "cold joint", "deposition", and "masonry joint". In each method, the number of patches used for training is determined experimentally. Table 3 presents the experimental results of the IoU and the number of patches used for training for each method when the test data are the distress images of lines A. As presented in Table 3, the average IoU values of

PM are higher than those of the other comparative methods in all cases where the test images are the distress images. Note that the methods showing the highest IoU values for "crack", "peeling", and "other distress" are "PM-Crack", "PM-Peeling", and "PM-Others", respectively. Thus, the proposed method effectively improves the detection performance of arbitrary distress in the images of distress in subway tunnels.



**Figure 5.** Examples of the results of the proposed method in line A. Figures (**A**–**D**) show the tunnel walls at various locations on line A.

Figure 8 shows the detection result using the proposed method performed only for "crack" (PM-Crack) on line A. The images in the left column are test images used in the experiment. The images in the center column are the pixel-level ground truth corresponding to the test images. The images in the right column are the estimated region using the proposed method. Figure 8 shows that the proposed method can accurately estimate the distress regions. Therefore, we improved the final detection performance by reducing the influence of the region of the distress other than the detection target by performing SICAP for each arbitrary distress. The above experimental results confirm the effectiveness of the proposed method when performed for a certain kind of distress.



**Figure 6.** Examples of the results of the proposed method in line B. Figures (**A**–**D**) show the tunnel walls at various locations on line B.

**Table 3.** Number of patches used for training after data augmentation and the average of IoU when detecting "crack", "peeling", and "others" based on each data augmentation method (DA method) in the ablation study.

		Number of Patches				IoU		
	DA Method	No Distress	Crack	Peeling	Others	Crack	Peeling	Others
CM1	-	75,265	19,304	3555	2483	0.3417	0.1018	0.4332
CM2	RICAP [50]	151,075	51,830	12,800	7730	0.3236	0.1147	0.4096
PM-All	SICAP	151,075	94,899	46,007	33,445	0.3094	0.1355	0.4982
PM-Crack	SICAP	94,569	38,608	3555	2483	0.3503	0.1242	0.4635
PM-Peeling	SICAP	91,014	19,304	19,304	2483	0.3219	0.1583	0.4098
PM-Others	SICAP	92,086	19,304	3555	19,034	0.3355	0.1423	0.5036



**Figure 7.** Examples of region-based results. The examples correspond to Figure 5C. Cracks in the image represent the ground truth. The closer the color in a patch is to red, the more positive patches are included.



**Figure 8.** Examples of the results of the proposed method in the ablation study. Figures (A–C) represent each test image.

## 4. Conclusions

In this paper, we proposed a new data augmentation method, SICAP, for distress detection to support the maintenance of subway tunnels. The proposed SICAP extracts only the distress regions from the images of distress of multiple lines and combines them to expand the data. The new image generated based on SICAP does not change the pixel values of the original image; thus, there is little loss of information. Additionally, since the expanded data invariably contain distressed regions, they are effective in building a robust model for various subway tunnel lines. Our new approach quantitatively outperformed the other data augmentation methods. We also confirmed the effectiveness of the proposed method by constructing a distress-specific detection model.

It is necessary to consider the introduction of other subway lines. This will be done as future work, as acquiring a sufficient amount of data is needed. Furthermore, we address the comparison with other data augmentation algorithms as future work.

**Author Contributions:** Conceptualization, K.M., S.T., T.H., R.T., T.O. and M.H.; methodology, K.M., S.T., T.H., R.T., T.O. and M.H.; software, S.T. and T.H.; validation, K.M., S.T., T.H., T.O. and M.H.; data curation, S.T. and T.H.; writing—original draft preparation, K.M., S.T. and T.H.; writing—review and editing, K.M., R.T., T.O. and M.H.; visualization, S.T. and T.H.; funding acquisition, K.M., T.O. and M.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** In this research, we utilized the inspection data that were provided by the Tokyo Metro Company Limited. This work was partly supported by JSPS KAKENHI Grant Number JP20K19856.

Data Availability Statement: Not applicable.

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

## References

- 1. Yang, Q.; Ji, X. Automatic pixel-level crack detection for civil infrastructure using Unet++ and deep transfer learning. *IEEE Sens. J.* **2021**, *21*, 19165–19175. [CrossRef]
- Shen, B.; Zhang, W.Y.; Qi, D.P.; Wu, X.Y. Wireless multimedia sensor network based subway tunnel crack detection method. *Int. J. Distrib. Sens. Netw.* 2015, 11, 184639. [CrossRef]
- 3. Dawood, T.; Zhu, Z.; Zayed, T. Deterioration mapping in subway infrastructure using sensory data of GPR. *Tunn. Undergr. Space Technol.* **2020**, *103*, 103487. [CrossRef]
- 4. Mehta, P.K.; Burrows, R.W. Building durable structures in the 21st century. Concr. Int. 2001, 23, 57–63.
- 5. Ministry of Land, Infrastructure Transport and Tourism. White Paper on Land, Infrastructure, Transport and Tourism in Japan, 2017 (online). 2018. Available online: http://www.mlit.go.jp/common/001269888.pdf (accessed on 13 November 2022).
- 6. Montero, R.; Victores, J.G.; Martinez, S.; Jardón, A.; Balaguer, C. Past, present and future of robotic tunnel inspection. *Autom. Constr.* **2015**, *59*, 99–112. [CrossRef]
- Bossio, A.; Monetta, T.; Bellucci, F.; Lignola, G.P.; Prota, A. Modeling of concrete cracking due to corrosion process of reinforcement bars. *Cem. Concr. Res.* 2015, 71, 78–92. [CrossRef]
- Woo, S.; Chu, I.; Youn, B.; Kim, K. Development of the corrosion deterioration inspection tool for transmission tower members. KEPCO J. Electr. Power Energy 2016, 2, 293–298. [CrossRef]
- 9. Pouliot, N.; Richard, P.L.; Montambault, S. LineScout technology opens the way to robotic inspection and maintenance of high-voltage power lines. *IEEE Power Energy Technol. Syst. J.* 2015, 2, 1–11. [CrossRef]
- Tešić, K.; Baričević, A.; Serdar, M. Non-destructive corrosion inspection of reinforced concrete using ground-penetrating radar: A review. *Materials* 2021, 14, 975. [CrossRef]
- 11. Bergquist, B.; Söderholm, P. Data analysis for condition-based railway infrastructure maintenance. *Qual. Reliab. Eng. Int.* **2015**, 31, 773–781. [CrossRef]
- 12. Xu, P.; Sun, Q.; Liu, R.; Souleyrette, R.R.; Wang, F. Optimizing the alignment of inspection data from track geometry cars. *Comput.-Aided Civ. Infrastruct. Eng.* **2015**, *30*, 19–35. [CrossRef]
- 13. Huang, Z.; Fu, H.; Chen, W.; Zhang, J.; Huang, H. Damage detection and quantitative analysis of shield tunnel structure. *Autom. Constr.* **2018**, *94*, 303–316. [CrossRef]
- 14. Liu, J.; Huang, Y.; Zou, Q.; Tian, M.; Wang, S.; Zhao, X.; Dai, P.; Ren, S. Learning visual similarity for inspecting defective railway fasteners. *IEEE Sens. J.* 2019, *19*, 6844–6857. [CrossRef]
- Ogawa, N.; Maeda, K.; Ogawa, T.; Haseyama, M. Correlation-aware attention branch network using multi-modal data for deterioration level estimation of infrastructures. In Proceedings of the IEEE International Conference on Image Processing, Anchorage, AK, USA, 19–22 September 2021; pp. 1014–1018.

- 16. Ogawa, N.; Maeda, K.; Ogawa, T.; Haseyama, M. Distress image retrieval for infrastructure maintenance via self-Trained deep metric learning using experts' knowledge. *IEEE Access* 2021, *9*, 65234–65245. [CrossRef]
- Calderón, L.S.; Bairán, J. Crack detection in concrete elements from RGB pictures using modified line detection kernels. In Proceedings of the Intelligent Systems Conference (IntelliSys), London, UK, 7–8 September 2017; pp. 799–805.
- Zhang, D.; Li, Q.; Chen, Y.; Cao, M.; He, L.; Zhang, B. An efficient and reliable coarse-to-fine approach for asphalt pavement crack detection. *Image Vis. Comput.* 2017, 57, 130–146. [CrossRef]
- Nguyen, H.N.; Kam, T.Y.; Cheng, P.Y. Automatic crack detection from 2D images using a crack measure-based B-spline level set model. *Multidimens. Syst. Signal Process.* 2018, 29, 213–244. [CrossRef]
- 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]
- Nayyeri, F.; Hou, L.; Zhou, J.; Guan, H. Foreground–background separation technique for crack detection. Comput.-Aided Civ. Infrastruct. Eng. 2019, 34, 457–470. [CrossRef]
- 22. Amhaz, R.; Chambon, S.; Idier, J.; Baltazart, V. Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2718–2729. [CrossRef]
- Qu, Z.; Guo, Y.; Ju, F.; Liu, L.; Lin, L. The algorithm of accelerated cracks detection and extracting skeleton by direction chain code in concrete surface image. *Imaging Sci. J.* 2016, *64*, 119–130. [CrossRef]
- 24. Ai, D.; Jiang, G.; Kei, L.S.; Li, C. Automatic pixel-level pavement crack detection using information of multi-scale neighborhoods. *IEEE Access* 2018, *6*, 24452–24463. [CrossRef]
- Hadjidemetriou, G.M.; Christodoulou, S.E.; Vela, P.A. Automated detection of pavement patches utilizing support vector machine classification. In Proceedings of the Mediterranean Electrotechnical Conference (MELECON), Limassol, Cyprus, 18–20 April 2016; pp. 1–5.
- 26. Turkan, Y.; Hong, J.; Laflamme, S.; Puri, N. Adaptive wavelet neural network for terrestrial laser scanner-based crack detection. *Autom. Constr.* **2018**, *94*, 191–202. [CrossRef]
- Basavaraju, A.; Du, J.; Zhou, F.; Ji, J. A machine learning approach to road surface anomaly assessment using smartphone sensors. *IEEE Sens. J.* 2019, 20, 2635–2647. [CrossRef]
- Maeda, K.; Takahashi, S.; Ogawa, T.; Haseyama, M. Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures. *Comput.-Aided Civ. Infrastruct. Eng.* 2019, 34, 654–676. [CrossRef]
- Maeda, K.; Takahashi, S.; Ogawa, T.; Haseyama, M. Estimation of deterioration levels of transmission towers via deep learning maximizing canonical correlation between heterogeneous features. *IEEE J. Sel. Top. Signal Process.* 2018, 12, 633–644. [CrossRef]
- 30. Ogawa, N.; Maeda, K.; Ogawa, T.; Haseyama, M. Deterioration level estimation based on convolutional neural network using confidence-aware attention mechanism for infrastructure inspection. *Sensors* **2022**, *22*, 382. [CrossRef]
- Maeda, K.; Takahashi, S.; Ogawa, T.; Haseyama, M. Neural network maximizing ordinally supervised multi-view canonical correlation for deterioration level estimation. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 919–923.
- Fan, X.; Wu, J.; Shi, P.; Zhang, X.; Xie, Y. A novel automatic dam crack detection algorithm based on local-global clustering. *Multimed. Tools Appl.* 2018, 77, 26581–26599. [CrossRef]
- Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic road crack detection using random structured forests. *IEEE Trans. Intell. Transp. Syst.* 2016, 17, 3434–3445. [CrossRef]
- 34. Guo, X.; Hao, P. Using a random forest model to predict the location of potential damage on asphalt pavement. *Appl. Sci.* **2021**, *11*, 10396. [CrossRef]
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* 2012, 25, 1097–1105. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26–30 June 2016; pp. 770–778.
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- Wang, A.; Togo, R.; Ogawa, T.; Haseyama, M. Detection of distress region from subway tunnel images via U-net-based deep semantic segmentation. In Proceedings of the IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, 15–18 October 2019; pp. 766–767.
- 39. Koch, C.; Georgieva, K.; Kasireddy, V.; Akinci, B.; Fieguth, P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Adv. Eng. Inform.* **2015**, *29*, 196–210. [CrossRef]
- 40. Tizani, W.; Mawdesley, M.J. Advances and challenges in computing in civil and building engineering. *Adv. Eng. Inform.* **2011**, 25, 569–572. [CrossRef]
- Dung, C.V. Autonomous concrete crack detection using deep fully convolutional neural network. *Autom. Constr.* 2019, 99, 52–58. [CrossRef]
- Zhang, W.; Zhang, Z.; Qi, D.; Liu, Y. Automatic crack detection and classification method for subway tunnel safety monitoring. Sensors 2014, 14, 19307–19328. [CrossRef]
- Khoa, N.L.D.; Anaissi, A.; Wang, Y. Smart infrastructure maintenance using incremental tensor analysis. In Proceedings of the ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 959–967.

- 44. Yang, X.; Li, H.; Yu, Y.; Luo, X.; Huang, T.; Yang, X. Automatic pixel-level crack detection and measurement using fully convolutional network. *Comput.-Aided Civ. Infrastruct. Eng.* 2018, 33, 1090–1109. [CrossRef]
- 45. Gong, Q.; Zhu, L.; Wang, Y.; Yu, Z. Automatic subway tunnel crack detection system based on line scan camera. *Struct. Control Health Monit.* **2021**, *28*, e2776. [CrossRef]
- Özgenel, Ç.F.; Sorguç, A.G. Performance comparison of pretrained convolutional neural networks on crack detection in buildings. In Proceedings of the International Symposium on Automation and Robotics in Construction, Berlin, Germany, 20–25 July 2018; Volume 35, pp. 1–8.
- 47. Li, Z.; Togo, R.; Ogawa, T.; Haseyama, M. A note on retrieval of visually similar distress regions in subway tunnel images: Introduction of deep features extracted by semantic segmentation network. *IEICE Tech. Rep.* **2020**, *119*, 65–68.
- Zhang, H.; Cisse, M.; Dauphin, Y.N.; Lopez-Paz, D. mixup: Beyond empirical risk minimization. *arXiv* 2017, arXiv:1710.09412.
   Inoue, H. Data augmentation by pairing samples for images classification. *arXiv* 2018, arXiv:1801.02929.
- Inoue, H. Data augmentation by pairing samples for images classification. *arXiv* 2018, arXiv:1801.02929.
   Takahashi, R.; Matsubara, T.; Uehara, K. Data augmentation using random image cropping and patching for deep cnns. *IEEE Trans. Circuits Syst. Video Technol.* 2019, 30, 2917–2931. [CrossRef]
- 51. Shorten, C.; Khoshgoftaar, T.M. A survey on image data augmentation for deep learning. J. Big Data 2019, 6, 60. [CrossRef]
- 52. Haruyama, T.; Maeda, K.; Togo, R.; Ogawa, T.; Haseyama, M. A note on improving performance of deep learning-based distress detection for supporting maintenance of subway tunnels Accuracy verification focusing on tunnel wall characteristics. *ITE Tech. Rep.* **2021**, *120*, 1–6.
- Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 40, 834–848. [CrossRef] [PubMed]
- 54. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [CrossRef]
- Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.