

Article

CNN-Based Road-Surface Crack Detection Model That Responds to Brightness Changes

Taehee Lee ¹, Yeohwan Yoon ¹, Chanjun Chun ² and Seungki Ryu ^{1,*}

¹ Future Infrastructure Research Center, Korea Institute of Civil Engineering and Building Technology (KICT), Goyang 10223, Korea; thlee420@gmail.com (T.L.); kicthyh@kict.re.kr (Y.Y.)

² Department of Computer Engineering, Chosun University, Gwangju 61452, Korea; cjchun@chosun.ac.kr

* Correspondence: skryu@kict.re.kr; Tel.: +82-31-910-0388

Abstract: Poor road-surface conditions pose a significant safety risk to vehicle operation, especially in the case of autonomous vehicles. Hence, maintenance of road surfaces will become even more important in the future. With the development of deep learning-based computer image processing technology, artificial intelligence models that evaluate road conditions are being actively researched. However, as the lighting conditions of the road surface vary depending on the weather, the model performance may degrade for an image whose brightness falls outside the range of the learned image, even for the same road. In this study, a semantic segmentation model with an autoencoder structure was developed for detecting road surface along with a CNN-based image preprocessing model. This setup ensures better road-surface crack detection by adjusting the image brightness before it is input into the road-crack detection model. When the preprocessing model was applied, the road-crack segmentation model exhibited consistent performance even under varying brightness values.

Keywords: road-surface cracks; image recognition; artificial intelligence; brightness; semantic segmentation



Citation: Lee, T.; Yoon, Y.; Chun, C.; Ryu, S. CNN-Based Road-Surface Crack Detection Model That Responds to Brightness Changes. *Electronics* **2021**, *10*, 1402. <https://doi.org/10.3390/electronics10121402>

Academic Editor: Danda B. Rawat

Received: 5 May 2021
Accepted: 8 June 2021
Published: 10 June 2021

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



Copyright: © 2021 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

To ensure driver safety and smooth vehicle operation, the road surface must always be maintained in a good condition. Maintenance of roads will become even more important with the development and commercialization of autonomous road vehicles. As irregular road-surface damage can affect the driving safety of autonomous vehicles and lead to accidents, a system that minimizes the time taken for damage detection and repair is required. Recently, a system for detecting the damaged portions on a road-surface image using a deep neural network has been developed.

Computer vision problems based on deep learning can be classified into image classification [1–4], object detection [5–7], semantic segmentation [8,9], instance segmentation [10–12], and panoptic segmentation [13,14]. In the image classification problem, the image is divided into small patches, and a neural network is trained to classify the cracks in each patch. Zhang et al. [15] classified the presence or absence of cracks in each image patch using a neural network composed of four convolutional layers and two fully connected layers. Pauly et al. [16] showed that the deeper the neural network, the better the crack image patch classification performance. Feng et al. [17] designed a deep residual network for image defect detection and classification. Eisenbach et al. [18] presented the German Asphalt Pavement Distress (GAPs) dataset, which includes road images classified into six types and applied them for deep-learning model training. Rateke et al. [19] presented a dataset for classifying the road-surface type and quality and reported the results of classification using convolutional layers.

The object detection technique mainly utilizes a method for determining the location of a crack in the entire image instead of splitting the image into small patches and considering each patch. Cha et al. [20] developed a technique to scan large images and mark the crack

locations by combining a convolutional neural network (CNN) trained with small patches and sliding window techniques. Maeda et al. [21] classified the road surface into eight types and constructed a CNN model to display bounding boxes indicating the location and type of damage.

The semantic segmentation technique, in which the CNN comprises an encoder and a decoder, is advantageous because it can determine the damage location as well as the geometry information of a crack. Schmugge et al. [22] proposed a crack segmentation method and presented the evaluation results for the remote visual inspection videos of nuclear power plants. Lau et al. [23] proposed a U-Net-based network architecture that uses a pretrained ResNet-34 neural network as the encoder to segment cracks on pavement surfaces. Rateke et al. [24] applied a semantic segmentation technique to segment the road-surface type, road-surface damage, and other information.

Instance segmentation combines object detection and semantic segmentation, thereby allowing pixels occupied by individual objects to be distinguished and visualized. In a crack detection problem, the location and shape of individual cracks in one image can be detected by instance segmentation models. Kim and Cho [25] developed a model based on Mask R-CNN for detection of multiple concrete damages. The model was trained using a total of 765 images, and 25 actual test images were used. Tane et al. [26] also developed a crack detector based on Mask R-CNN by training 352 original and annotated crack images. However, it appears that there are an inadequate number of image datasets suitable for training instance segmentation models for the crack detection problem. Panoptic segmentation is a combination of instance and semantic segmentation wherein all pixels in one image are classified into each object. However, panoptic segmentation hasn't been studied as thoroughly as other techniques since crack detection doesn't require segmenting every pixel in an image.

A common issue with all crack detection models is that they can provide meaningful results only for data that reflect the same conditions used for their training. Thus, even for the same surface damage at the same location, if the external environmental changes and leads to a change in brightness levels with respect to the training images, the damage may not be detected. To address this issue, in this study, we developed a fully convolutional network (FCN) model that semantically divides the damaged road pixels on a road-surface image and a CNN model that automatically provides an image-brightness-control variable to enable the developed FCN model to achieve the best detection performance. The road-surface image is input into the CNN model and preprocessed before it is input into the FCN-based road-crack segmentation model. Preprocessed and unprocessed images are input into the FCN model, and the effectiveness of the developed system is analyzed by comparing the detection performance between the two conditions.

2. Development of a Road-Crack Detection System That Responds to Brightness Changes

Figure 1 shows the overall structure of the developed road-surface crack detection system that responds to brightness changes. In a 1920×1080 image captured while looking ahead when driving on the road, the area corresponding to the road surface was cropped to a size of 1920×256 . To reduce the usage of computing resources, the cropped area was resized to 960×128 , and the RGB values of the image were placed in a $[960, 128, 3]$ matrix and input into the CNN-based image preprocessing model.

On receiving the input, the image preprocessing model calculates the enhancement factor (f_B), which adjusts the brightness of the image for the effective detection of road-surface cracks. In the HSB color space, the image is changed through multiplication by the brightness matrix $[B]$, and the brightness of the changed image is $[B']$.

$$[B'] = f_B [B] \quad (1)$$

The road-surface image with $[B']$ in the HSB color space is converted to the RGB color space; the converted RGB values are placed in a $[960, 128, 3]$ matrix and input into

the FCN-based road-surface damage inspection model. Finally, the damaged parts of the road-surface are labeled as “0” and “1” semantically.

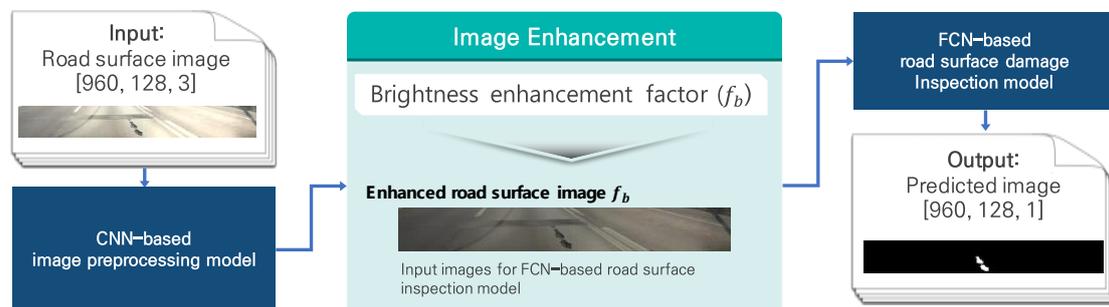


Figure 1. Overall flow of the developed road-surface inspection system.

3. Development of a Road-Surface Crack Detection Model

3.1. Structure of the Model

In this study, we developed a road-surface crack detection model based on the FCN architecture, which is extensively used in semantic segmentation. Figure 2 depicts the schematic of the road-surface crack detection FCN model used in this study. The auto-encoder structure [27,28] includes a convolution layer and deconvolution layer, and when a 960×128 image comprising RGB values is input into the three channels, a 960×128 image comprising “0” and “1” is finally output in one channel.

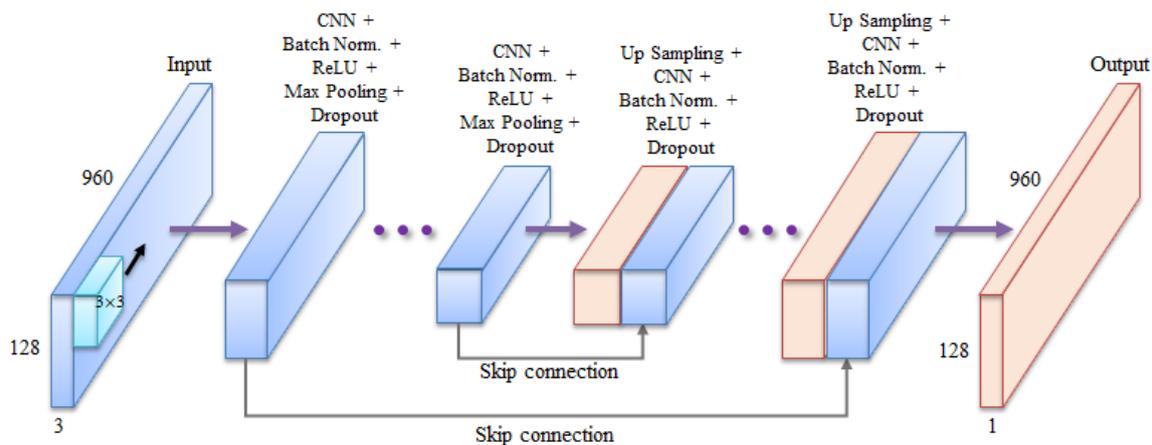


Figure 2. Schematic of the fully convolutional network (FCN)-based road-surface crack detection model.

A total of five FCN model structures with four, six, eight, ten, and twelve hidden layers excluding the input and output layers, respectively, were tested to select the final FCN model. In all the models, the convolution layers constituted half of the total hidden layers, whereas the deconvolution layers constituted the other half. The kernel size used in all the models was (3, 3). After zero padding, the composite product was multiplied by a stride of (1, 1). The number of kernels was 16 in the first convolution layer of each model, which increased by double each time it moved to the next layer. A batch normalization technique [29] was applied before passing the ReLU (rectified linear unit) [30], which was used as the activation function of the hidden layer. After applying the activation function to the convolution layer, the result was max-pooled with a size of (2, 2). In the deconvolution layer, up-sampling was performed using a (2, 2) filter before performing convolution, and the convolution layers at the symmetrical position were connected through skip connections. A dropout rate of 0.2 was applied during the learning of all the hidden layers. In the output layer, sigmoid was designated as the activation function, and the resulting value for each pixel was classified as “0” and “1”.

3.2. Model Training and Dataset Configuration for Testing

Figure 3 shows the dataset composition for training and testing the model. A camera was attached to the top of the vehicle's windshield facing the road surface, and, while driving at approximately 40–100 km/h, the road surface was periodically photographed. Subsequently, a total of 14,400 road-surface images were carefully prepared. In addition, images were classified based on six properties: artificial joints (such as the expansion joints at bridge junctions and the discontinuous surfaces between existing and new road surfaces present in repaired locations), road markings, roadside structures, shadows on the road, vehicle images, and road cracks. Figure 3a–e depict cases that can cause false positives when detecting road-surface cracks. A one-channel matrix of the same size as the original images was generated, and all the components were labeled as zero to indicate that there was no road-surface damage. In Figure 3f, which shows images displaying road-surface cracks, the pixels of the damaged area on the road are labeled as “1”, as shown in Figure 4. Among each image type, 1600 images were used as the training set, 400 as the validation set, and the remaining 400 as the test set. Therefore, the total number of images in the training, validation, and test sets was 9600, 2400, and 2400, respectively.

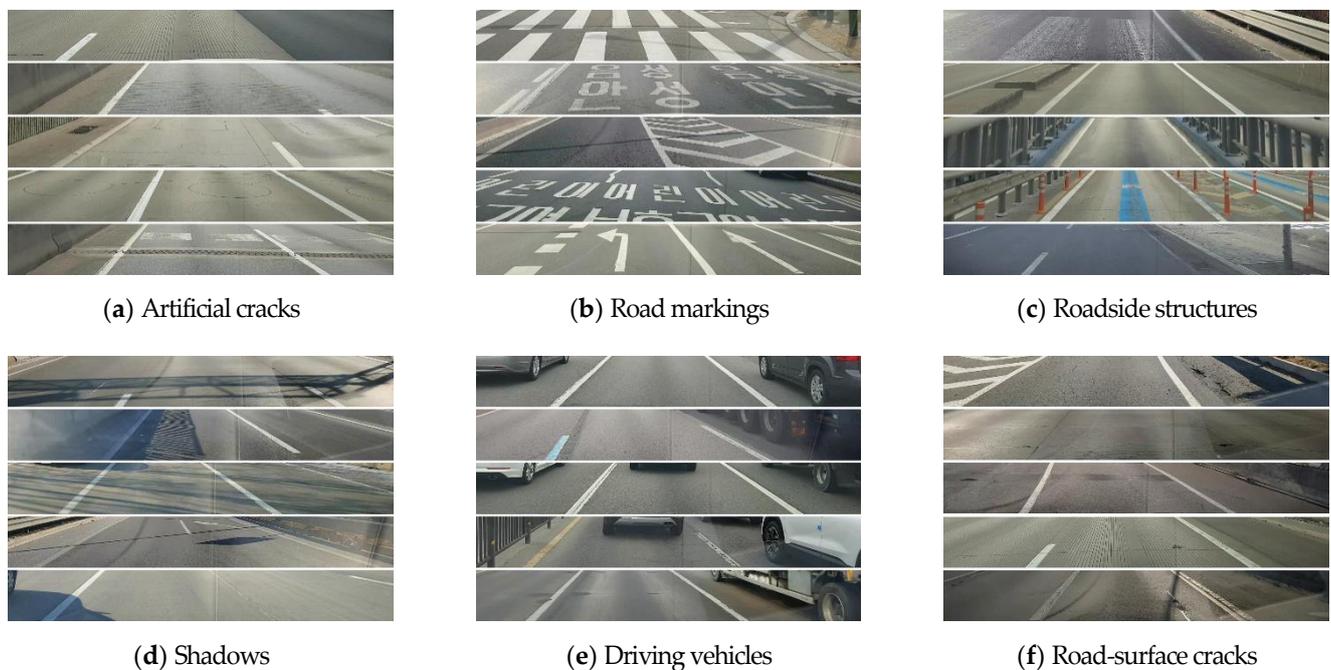


Figure 3. Datasets for training the road-surface crack segmentation model.



Figure 4. Examples of original and labeled images with road-surface cracks.

3.3. Learning Results by the Model

In the semantic segmentation model, the intersection over union (IoU) is generally used as an index to evaluate the prediction results [31]. The IoU value is obtained by dividing the intersecting area of the model-predicted pixels and the ground truth pixels by the union area of the model-predicted pixels and ground truth pixels. The closer the IoU to unity, the greater the coincidence of the predicted pixel region with the ground truth. It was assumed that if the trained model provides an IoU of 0.4 or more for the test images, the results were classified well predicted. After determining the prediction results of the learning model based on the IoU, the true positive, true negative, false positive, and false negative values were arranged in a confusion matrix to obtain the precision, recall, and F1-score [32].

Figure 5 shows the F1-score according to the number of hidden layers used in the learning model and the amount of the training data. If few hidden layers exist, the number of parameters required to calculate the output decreases; hence, installation on a mobile device becomes easy and can be used to segment road-surface cracks in real-time. However, if the number of parameters is less, and even if the training data are sufficiently prepared, the performance limit is low, and the model results are unreliable. In addition, even if the number of hidden layers is considerable, if the training data are insufficient, the parameters may not be sufficiently learned, resulting in low performance.

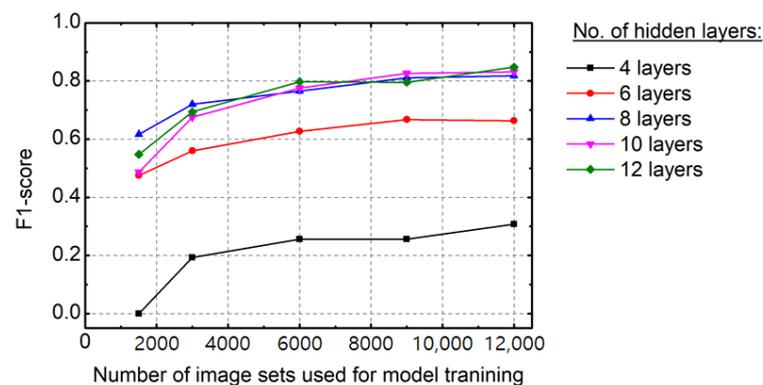


Figure 5. F1-score with different hidden layers and the number of model training images set.

As indicated in Figure 5, the model with four hidden layers obtained an F1-score of 0.31, even though all the images prepared for training and validation were used; if only 1500 images are used, the F1-score is zero, and it is not possible to predict the test set result. When the number of hidden layers increased to six, an F1-score of 0.66 was obtained using all the training and validation images, and the performance improved by two-fold in comparison with that of the four-layer model. In the case of the models with 8, 10, and 12 layers each, F1-scores of 0.81, 0.83, and 0.85, respectively, were obtained, and the performances tended to converge.

As the brightness of road surface was not constant, the same performance may not be obtained. Figure 6 depicts the F1-scores when the brightness was varied by adjusting f_B from 0.2 to 1.8 in intervals of 0.2 in the 2400 images used as the test set, which were input into the FCN-based road-surface crack segmentation models trained by a total of 9600 training and 2400 validation images. When $f_B = 1$, the brightness was the same as that of the original image. In all the models, when the brightness decreased, for $f_B \geq 0.6$, the F1-score was similar to that obtained when the original image was input; however, for $f_B < 0.6$, the F1-score rapidly decreased and approached zero at $f_B = 0.2$. As the brightness increased, in comparison with that of the original image, the F1-score also decreased. Even if the same image is used, it is not possible to achieve the same crack segmentation performance with different brightness levels.

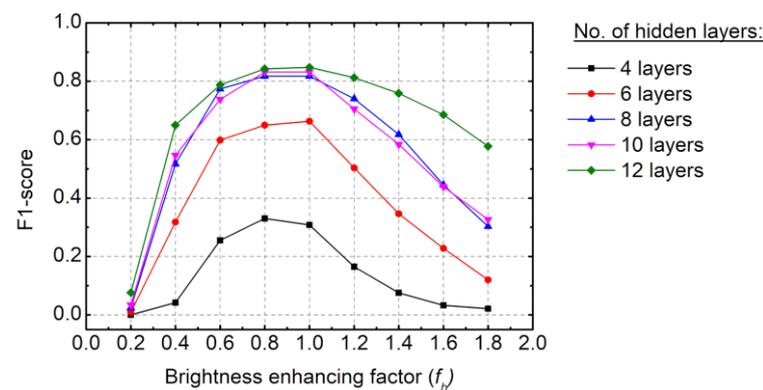


Figure 6. F1-score with different hidden layers and brightness enhancing factors.

4. Development of Preprocessing Model for Adopting Brightness

4.1. Structure of Preprocessing Model

Figure 7 depicts the configuration diagram of the image preprocessing model that receives an image with a matrix of (128, 960, 3) and delivers f_B approximating toward the brightness used in the training and validation sets. Excluding the input and output layers, it comprises a total of six CNN layers and five fully connected layers. Kernels sized (3, 3) in all the CNN layers were strided at intervals of (1, 1). The number of kernels was set to 16 in the first layer, which was designated to increase by double passing through the following CNN layer. In addition, the CNN performed convolution after zero padding, and max pooling was performed with a (2, 2)-sized filter. After flattening the component of the last CNN layer, the final f_B output was calculated using the five fully connected layers with 64, 32, 16, 8, and 1 nodes.

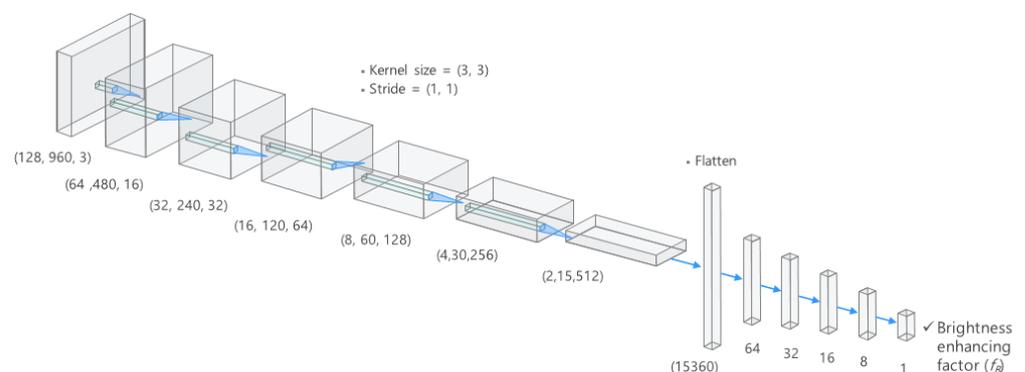


Figure 7. Schematic of the CNN-based image preprocessing model.

4.2. Training Dataset for the Image Preprocessing Model

Figure 8 shows the process of creating the dataset used for training the image preprocessing model depicted in Figure 7. The brightness of the 9600 training images and the 2400 validation images were changed by selecting a random f_B between 0.2 and 1.8. From a total of 12,000 images with randomly changed brightness values, the f_B values of 9600 training images were varied from 0.2 to 1.8 and input into the FCN model; the f_B exhibiting the maximum IoU was used as the target value to train the CNN-based image preprocessing model. Thus, the road-surface images are optimized for the crack detection model by using f_B , which is provided by the CNN-based image preprocessing model.

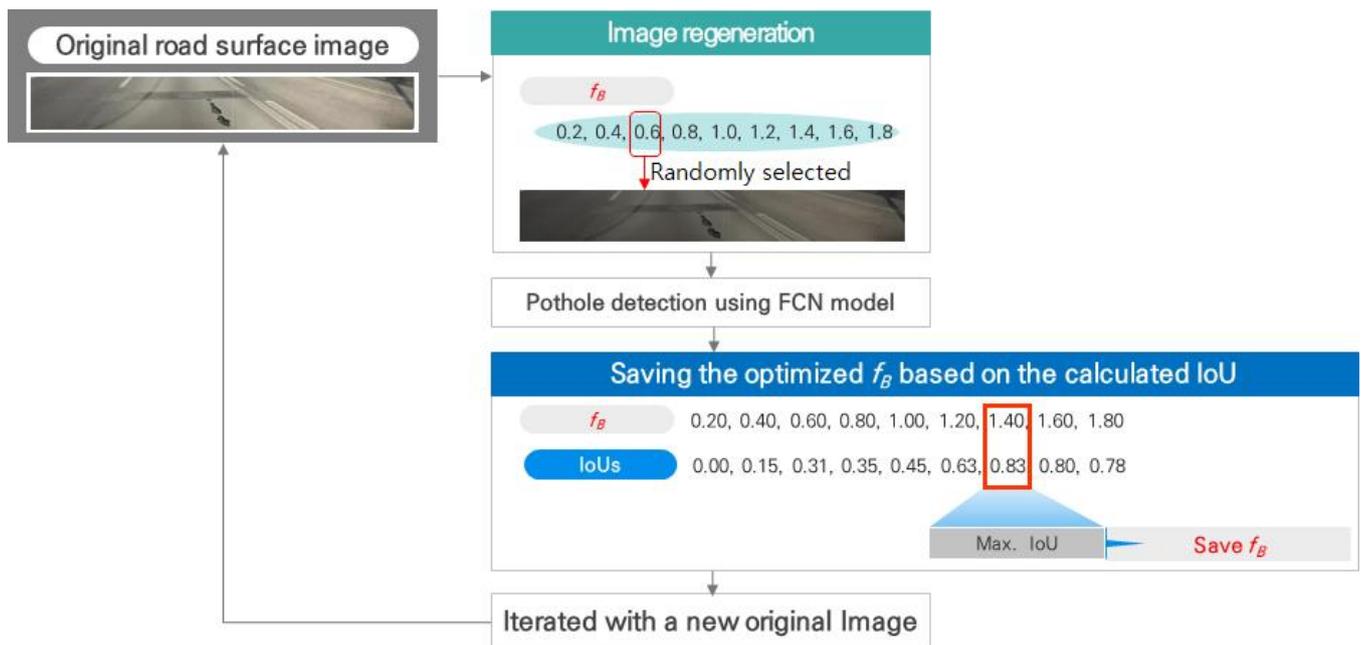
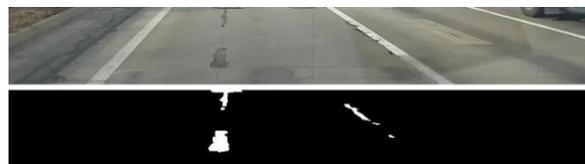


Figure 8. Dataset creation process for application in the image preprocessing model.

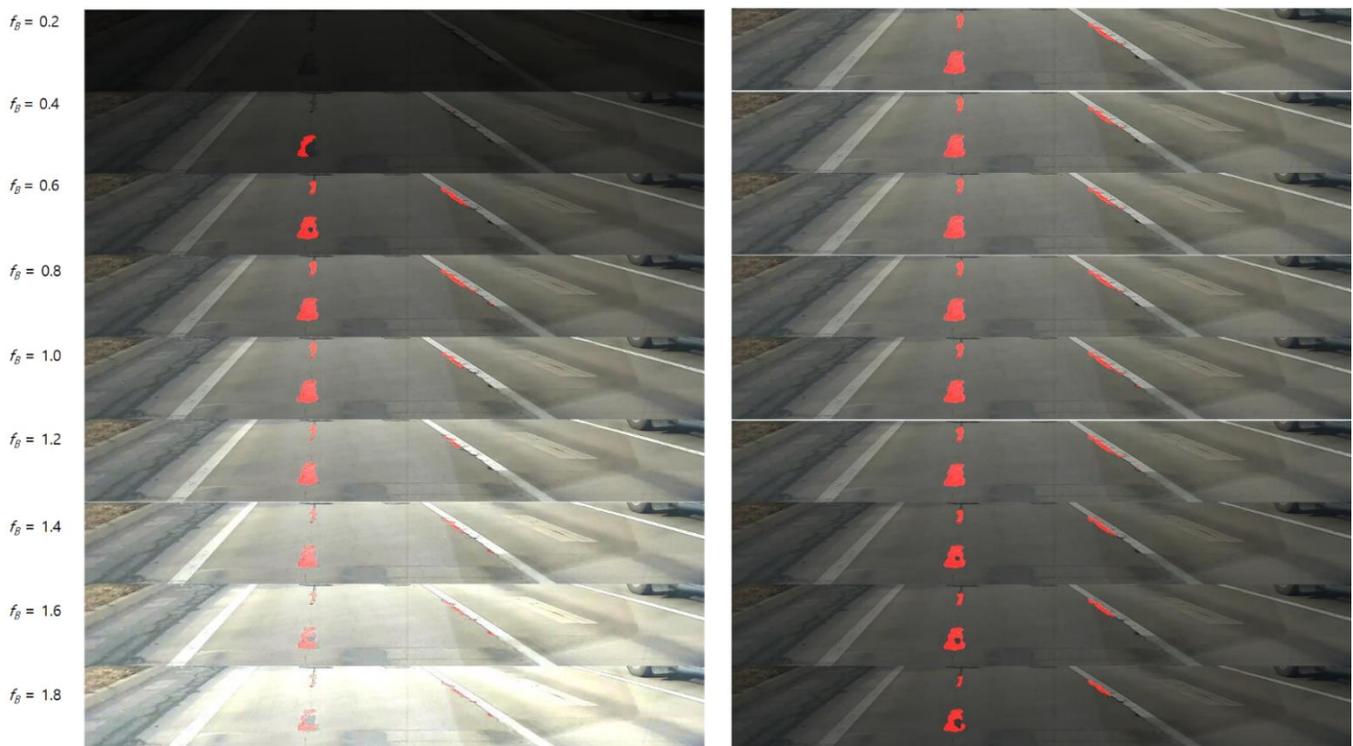
5. Performance Evaluation of the Road-Surface Crack Segmentation Model with Brightness Preprocessing

Figure 9a depicts one of the original and labeled images among the 2400 test data used for evaluating the result of linking the image preprocessing and road-crack segmentation models. In the original image in Figure 9a, f_B varied from 0.2 to 1.8 in increments of 0.2. When the image with the changed f_B was input into the FCN-based road-surface crack segmentation model without image preprocessing, the predicted crack areas are depicted in red, as shown in Figure 9b. In the figure, when f_B was 0.2, the brightness was the least, and it was difficult to recognize the condition of the road-surface; hence, no road-surface cracks could be detected on the image. Even when the road-surface image was bright, smaller regions were detected as road-surface cracks, degrading the model performance. Figure 9c shows the result of the road-surface crack detection model after the CNN-based preprocessing model optimized the image brightness. The image's brightness was converted such that the cracks of the road surface could be well identified, and the model performance was improved.



(a) Original and labeled images.

Figure 9. Cont.



(b) Predicted images without preprocessing.

(c) Predicted images with preprocessing.

Figure 9. Predicted images with and without the image preprocessing model.

Figure 10 depicts the comparison of the IoUs obtained before and after the application of the image preprocessing model to the brightness-converted image in Figure 9. If the image preprocessing model is not applied, the IoU decreases rapidly with the change in brightness; however, on applying the preprocessing model, the IoU maintained the stable values.

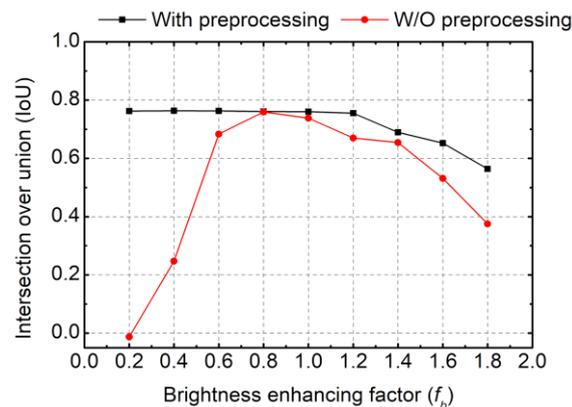


Figure 10. Comparison of the IoUs obtained with and without the image preprocessing model.

Figure 11 shows the histogram of the brightness when f_b of the original image (Figure 9) was changed, and that obtained when the image with the changed f_b was input into the image preprocessing model. In Figure 11a, the brightness of the original image was distributed between 100 and 150. However, when f_b was adjusted under 1.0, it was distributed as a lower brightness value within a narrower range; when f_b was increased than 1.0, the brightness was high over a wider range. However, when the f_b -adjusted image was input into the image preprocessing model, the image brightness converged to 50–150, which becomes similar to the original brightness of the image in the training set.

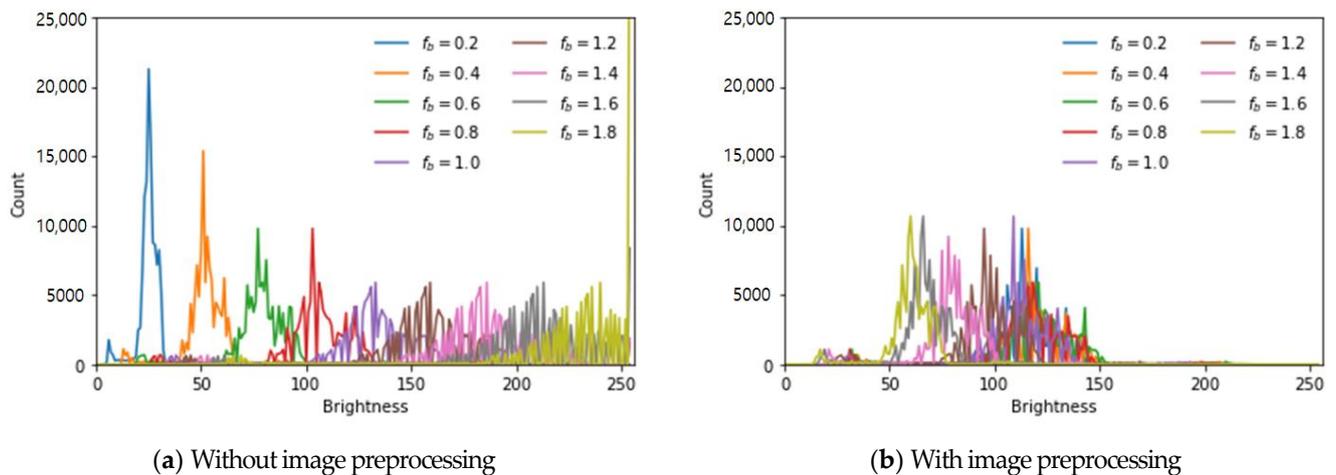


Figure 11. Histogram of the brightness with and without image preprocessing.

Figure 12 depicts the F1-score after applying the image preprocessing model to the five previously generated FCN-based models. As shown in Figure 6, as the f_B moved farther from the original image, the detection performance decreased significantly. However, when the preprocessing model was applied, road-surface cracks can be easily detected in the FCN model, and relatively constant F1-scores can be obtained.

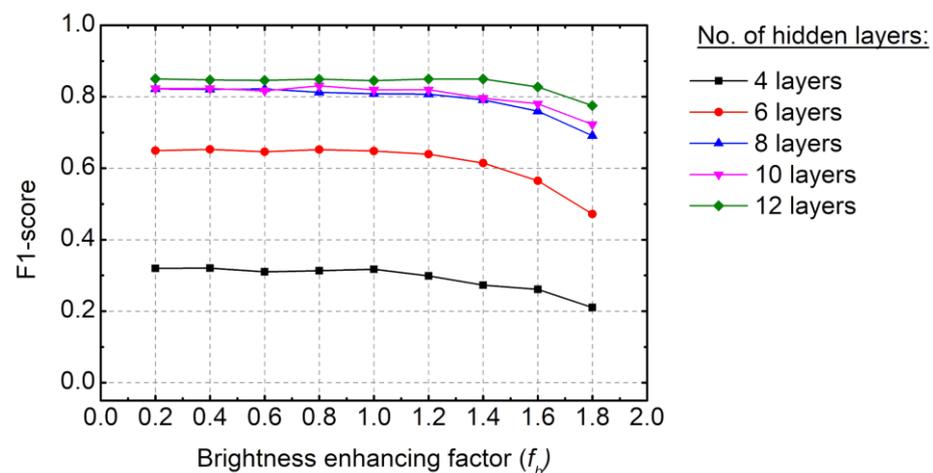


Figure 12. Improved F1-score on applying the image preprocessing model.

6. Conclusions

In this study, an FCN-based road-surface crack detection model and a CNN-based image's brightness optimizing model were developed. A total of 14,400 images were collected for six road conditions, including cracks on the road surface; among these, 9600 were used as the training dataset, 2400 were used as the validation dataset, and 2400 were used as the test set. When the test dataset was input into the FCN-based crack detection model with 12 layers, an F1-score of 0.85 was obtained. When the brightness of the original image was changed, the performance of the FCN-based crack detection model degraded rapidly, and there were cases in which the road-surface cracks could not be detected. However, when the CNN-based image preprocessing model was applied, which improved the detection of road-surface cracks by adjusting the brightness, the crack detection performance remained relatively constant.

Nevertheless, the crack images used in this study did not classify the types of cracks, and the FCN-based road-surface crack detection model was limited to expressing the damaged area alone. In addition, the 1920×256 -sized labeling images were disadvantageous when directly applied to the latest CNN model or applied for transfer learning. Therefore,

to develop a high-performance model that detects cracks and classifies the types of cracks, further research is needed to reconstruct the current labeled data.

Author Contributions: Writing—original draft preparation, methodology, and performance test, T.L., C.C. and Y.Y.; supervision and project administration, S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a grant (No. 19TLRP-B148886-02) [Commercial Vehicle-Based Road and Traffic Information System] from the Korea Agency for Infrastructure Technology Advancement (KAIA) and a grant (P0013941, 2020) [The Project of Conversion by the Past R&D Results] from the Ministry of Trade, Industry and Energy (MOTIE).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the implementation of further research.

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

References

1. Eigen, D.; Puhersch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS), Montreal, ON, Canada, 8–13 December 2014; pp. 2366–2374.
2. Iandola, F.; Moskewicz, M.; Karayev, S.; Girshick, R.; Darrell, T.; Keutzer, K. Densenet: Implementing efficient convnet descriptor pyramids. *arXiv* **2014**, arXiv:1404.1869.
3. Sun, Y.; Liang, D.; Wang, X.; Tang, X. Deepid3: Face recognition with very deep neural networks. *arXiv* **2015**, arXiv:1502.00873.
4. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–10 February 2017.
5. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 779–788.
6. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
7. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
8. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)] [[PubMed](#)]
9. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
10. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Araucano Park, Las Condes, Chile, 11–18 December 2015; pp. 1440–1448.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
12. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.
13. Kirillov, A.; He, K.; Girshick, R.B.; Rother, C.; Dollár, P. Panoptic Segmentation. *arXiv* **2018**, arXiv:1801.00868.
14. Xiong, Y.; Liao, R.; Zhao, H.; Hu, R.; Bai, M.; Yumer, E.; Urtasun, R. UPSNet: A Unified Panoptic Segmentation Network. *arXiv* **2019**, arXiv:1901.03784.
15. Zhang, L.; Yang, F.; Zhang, Y.D.; Zhu, Y.J. Road crack detection using deep convolutional neural network. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3708–3712.
16. Pauly, L.; Hogg, D.; Fuentes, R.; Peel, H. Deeper networks for pavement crack detection. In Proceedings of the 34th ISARC, Taipei, Taiwan, 28–31 June 2017; pp. 479–485.
17. Feng, C.; Liu, M.Y.; Kao, C.C.; Lee, T.Y. Deep active learning for civil infrastructure defect detection and classification. In Proceedings of the ASCE International Workshop on Computing in Civil Engineering 2017, Seattle, DC, USA, 25–27 June 2017; pp. 298–306.
18. Eisenbach, M.; Stricker, R.; Seichter, D.; Amende, K.; Debes, K.; Sesselmann, M.; Ebersbach, D.; Stoeckert, U.; Gross, H. How to get pavement distress detection ready for deep learning? In A systematic approach. In Proceedings of the 2017 International Joint Conference on Neural Networks IJCNN, Anchorage, AK, USA, 14–19 May 2017; pp. 2039–2047.
19. Rateke, T.; Justen, K.A.; von Wangenheim, A. Road surface classification with images captured from low-cost camera-road traversing knowledge (rtk) dataset. *Rev. Inf. Teórica Appl.* **2019**, *26*, 50–64. [[CrossRef](#)]
20. Cha, Y.J.; Choi, W.; Büyüköztürk, O. Deep learning-based crack damage detection using convolutional neural networks. *Comput.-Aided Civ. Infrastruct. Eng.* **2017**, *32*, 361–378. [[CrossRef](#)]
21. Maeda, H.; Sekimoto, Y.; Seto, T.; Kashiya, T.; Omata, H. Road damage detection using deep neural networks with images captured through a smartphone. *arXiv* **2018**, arXiv:1801.09454v1.

22. Schmugge, S.J.; Rice, L.; Lindberg, J.; Grizziy, R.; Joffey, C.; Shin, M.C. Crack Segmentation by Leveraging Multiple Frames of Varying Illumination. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, 24–31 March 2017; pp. 1045–1053.
23. Stephen, L.; Chong, L.H.; Edwin, K.P.; Xu, T.; Wang, X. Automated Pavement Crack Segmentation Using U-Net-Based Convolutional Neural Network. *IEEE Access* **2020**, *8*, 114892–114899.
24. Rateke, T.; Von Wangenheim, A. Road surface detection and differentiation considering surface damages. *Auton. Robot.* **2021**, *45*, 299–312. [[CrossRef](#)]
25. Kim, B.; Cho, S. Automated Multiple Concrete Damage Detection Using Instance Segmentation Deep Learning Model. *Appl. Sci.* **2020**, *10*, 8008. [[CrossRef](#)]
26. Tan, C.; Uddin, N.; Mohammed, Y.M. Deep Learning-Based Crack Detection Using Mask R-CNN Technique. In Proceedings of the 9th International Conference on Structural Health Monitoring of Intelligent Infrastructure, St. Louis, MO, USA, 4–7 August 2019; pp. 1484–1490.
27. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.-A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
28. Lu, X.; Tsao, Y.; Matsuda, S.; Hori, C. Speech enhancement based on deep denoising autoencoder. *Proc. Interspeech* **2013**, *1*, 436–440.
29. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning (ICML), Lille, France, 6–11 July 2015; pp. 448–456.
30. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel, 21–24 June 2010; pp. 807–814.
31. Liu, X.; Deng, Z.; Yang, Y. Recent progress in semantic image segmentation. *Artif. Intell. Rev.* **2019**, *52*, 1089–1106. [[CrossRef](#)]
32. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Proceedings of the 27th European Conference on Advances in Information Retrieval Research (ECIR), Santiago de Compostela, Spain, 21–23 March 2005; pp. 345–359.