Anomaly detection using local regions in road images acquired from a hand-held camera

Takuya Kamitani[1], Syoji Fujimoto[1], Hiroki Yoshimura[1,2], Masashi Nishiyama[1,2] and Yoshio Iwai[1,2]
[1] Graduate School of Engineering, Tottori University, Japan
[2] Cross-informatics Research Center, Tottori University, Japan
Email: nishiyama@tottori-u.ac.jp

***Abstract*—We investigate a method to accurately detect road images with anomalies using the local regions generated from a small number of reference images. There are few datasets of road images with labeled anomalies acquired by hand-held cameras that are large enough to train an accurate detector. We hence evaluated whether an anomaly road image detector using local regions trained on a small dataset of reference images can increase performance. Experimental results show that the use of local regions instead of whole images significantly improves detection performance on a road dataset collected by the local government of Tottori prefecture.**

***Index Terms*—Anomaly detection, Road images, Local regions**

I. Introduction

There is a high demand to maintain roads so that drivers and pedestrians can safely travel on them. When an anomaly occurs on a road surface, a management organization such as a local government needs to quickly detect and appropriately repair it. However, because roads exist over a wide region, it is costly to continuously detect anomalies. In particular, local governments face the serious problem of how to reduce the burden of maintaining roads as further population decreases are expected in the future. In the current situation, observers working in cooperation with local governments manually detect anomalies on road surfaces. To assist such observers, a system that determines whether or not road images contain anomalies is required.

To develop an anomaly detection system, existing methods [1]–[5] search for cracks and potholes in pavement images acquired from a camera embedded in a special vehicle. The performance of crack and pothole detection is highly accurate under a constraint that fixes the relative positions of the road surface and a camera. However, because a special vehicle is very expensive, local governments cannot easily use it. Thus, a system that can help to detect anomalies from road images acquired from hand-held cameras is required. In this case, backgrounds other than the road surface such as natural objects and buildings sometimes appear in the road images because an observer is not able to keep his/her camera angle constant. Thus, existing methods [1]–[5] cannot be directly applied to detect anomalies in road images acquired by a hand-held camera.

To eliminate the constraint on the relative positions of the camera and road surface, a road segmentation [6]–[9] using a convolutional neural network (CNN) is effective. However, existing methods do not consider how to detect whether or not there are anomalies on the road surface. If we collected a large number of reference images with labeled road anomalies, we would be able to improve detection accuracy by retraining the network models of existing methods [6]–[9]. Recently, a public dataset with a vehicle-mounted camera has been collected [10]. However, there are few datasets with labeled road anomaly images acquired from hand-held cameras. A collection of reference images requires a great deal of cost for a local government. We thus must develop a method to effectively utilize a small number of reference images acquired from a hand-held camera.

In this paper, we focus on the use of local regions generated from a small number of reference images to train a CNN instead of using whole images. Recently, with respect to object detection, existing methods [11]–[13] have exploited local regions by combining them with a CNN. Our method simply exploits a sliding window search technique with a CNN. We investigate whether or not local regions have the ability to detect road anomaly images acquired using a hand-held camera. Experimental results on a dataset of road images collected by a local government show that our method using local regions is clearly superior to one using whole images. The remainder of the paper is organized as follows. Section II describes the details of our method, Section III presents the detection performance using local regions, and Section IV gives our concluding remarks.

II. Road anomaly image detection using local regions

A. Overview

To develop a method for detecting road images with anomalies, there are two approaches: the first one is segmentation, e.g., [6]–[9], the second one is localization, e.g., [11]–[13]. The use of local regions is categorized as a localization approach. Our method generates many local regions from a single whole image acquired from a hand-held camera. We aim to increase the number of training samples for training a CNN. We believe that the use of local regions rather than whole images is more robust because the influence caused by the variation in poses of a hand-held camera is reduced. Our method is described in detail below.

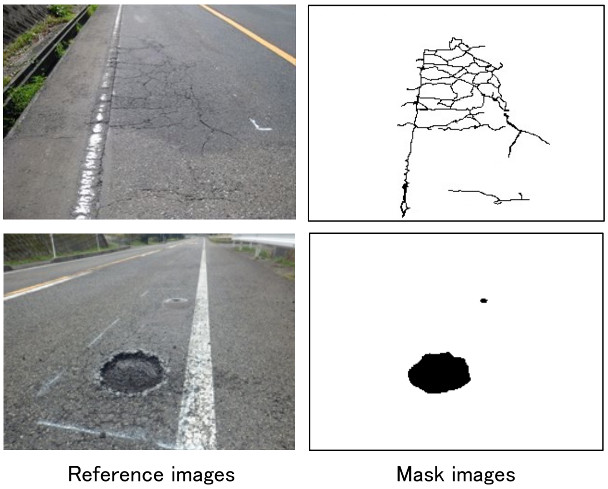

Fig. 1. Examples of reference and mask image pairs.

B. Generating local regions for training a CNN

To generate local regions for training a CNN, we collected a small number of reference images with manually labeled anomalies such as cracks and potholes. We constructed a pair of images consisting of a reference image acquired from a hand-held camera and a mask image that indicates the anomaly labels for each pixel. Figure 1 shows examples of reference and mask image pairs. Black pixels represent anomalies and white pixels represent the normal road surface.

We generate local regions for reference and mask image pairs, as illustrated in Figure 2. Our method exploits a grid sampling technique. As described in [14], a grid sampling technique divides the reference image using equal intervals and defines a local region around each intersection. The interval of grid sampling is G_1 pixels, and the size of a local region is $S \times S$ pixels. Furthermore, our method generates the local regions of the mask images using the same grid sampling. For each local region in the mask image, we compute the ratio of anomaly pixels, i.e., the number of labeled pixels in the local region of the mask image, to the total number of pixels in that local region. We use thresholds R_1 and $R_2\%$ ($R_1 \geq R_2$) and regard a local region for which the ratio of anomaly pixels is equal to or more than R_1 as a positive sample; if this ratio is equal to or less than R_2, it is a negative sample. Our method employs a margin defined by R_1 and R_2 to eliminate local regions where the boundary between anomalies and normal road surface are ambiguous because such local regions sometimes cause failure when training a CNN.

C. Detecting a road anomaly image

Our method exploits a grid search technique, which performs detection process at fixed intervals, for detecting a test image containing anomalies. With respect to object detection, a selective search technique [15] is sometimes exploited. This technique determines candidates for a detection process by grouping small regions where textures are similar. Selective search works well for object detection because the textures of each object are different. However, we believe that it does not work well for the detection of road anomaly images because the textures of road images are very similar. Thus, we simply use a grid search technique.

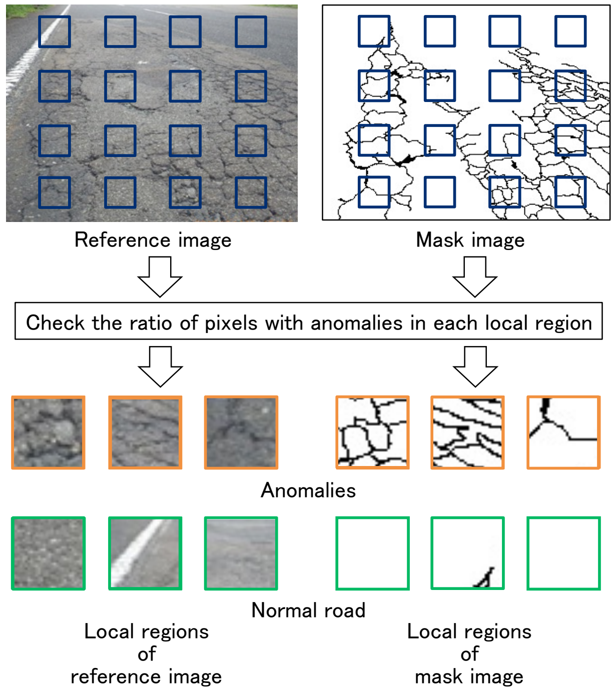

Fig. 2. Overview of the generation of local regions.

Figure 3 illustrates an overview of anomaly detection in a test image using local regions. Our method trains a CNN in advance using the positive and negative samples generated as described in Section II-B. Given a test image, we apply a grid search technique using local regions of size $S \times S$ pixels. The interval of the grid search is G_2 pixels.

To design a practical application, we need to consider how to alert an observer to a test image containing local regions with anomalies. To determine whether or not our method should alert an observer, we use a ratio consisting of the number of local regions containing CNN-detected anomalies to the number of searched local regions. Our method alerts the observer when the ratio of local regions with anomalies to total regions in a test image equals or is more than threshold $R_3\%$.

III. Evaluation

A. Dataset

We used a dataset of road images collected by a local government. The target roads for the test lie in the central part of Tottori prefecture. Observers who cooperated with the Torrori prefectural office acquired road images using consumer digital cameras. The data was collected over four years, from 2014 to 2018. They collected 159 images from 119 different locations at which the observers judged that a repair to the road surface was needed. The number of road

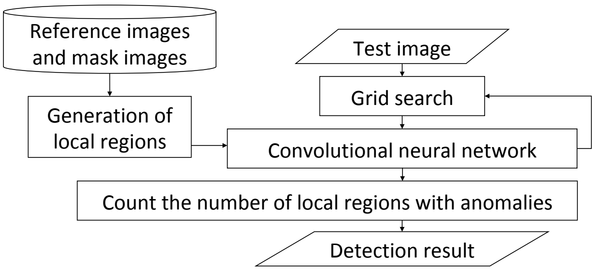

Fig. 3. Overview of road anomaly image detection using local regions.

Fig. 4. Examples of reference images collected by the local government of Tottori prefecture for the dataset.

images with anomalies was 119, and the number of normal road images was 40. The average size of a road image was $311.8 \pm 32.9 \times 236.0 \pm 26.1$ pixels. Figure 4 shows examples of road images included in our dataset. We manually labeled the pixels of cracks and potholes in the road images. Figure 5 shows examples of local regions generated from the dataset.

B. Basic detection performance for local regions with anomalies

To determine the parameters of our method, we evaluated detection performance of the system on local regions with anomalies. We applied eight-fold cross validation to the dataset of the road images described in Section III-A. We generated local regions in each validation set. We randomly selected local regions such that the numbers of positive and negative training samples were equivalent. We used a deep learning technique (the layer architecture was *Mini*-CNN, as described in [16]). We set $G_2 = 3$ to generate the test samples. We used F-score,

Fig. 5. Examples of local regions with anomalies or normal road surface for evaluating detection performance.

defined using precision and recall, to evaluate the detection performance.

Figure 6 shows the average and standard deviation of the F-scores for detecting local regions with anomalies when a certain parameter is fixed and the others are changed. We can see that $G_1 = 5$ is a better value than 10 or 15, $S = 32$ and 40 yields better performance than 48, $R_1 = 20$ is better than 15 or 25, and $R_2 = 5$ is a better value than 0, 10, or 15. We obtained the best performance using $G_1 = 5$, $S = 40$, $R_1 = 20$, and $R_2 = 5$.

C. Evaluation of the road anomaly image detection system

We evaluated the performance of the road anomaly image detection system using the results of local regions. Using threshold R_3, our method determines whether or not a test image contains an anomaly. We used the classifier for local regions presented in Section III-B and evaluated its F-score performance.

We compared the performance of our method with that of a baseline method that used a CNN with whole images instead of local regions. Table I shows the F-scores for road anomaly image detection using $R_3 = 5$. The results show that the performance of our method is superior to that of the baseline method. We believe that local regions have the potential to correctly detect road anomaly images, especially when compared with whole images, when the number of reference images is small.

IV. Conclusions

We proposed a method for detecting a road image with anomaly by generating many local regions from reference

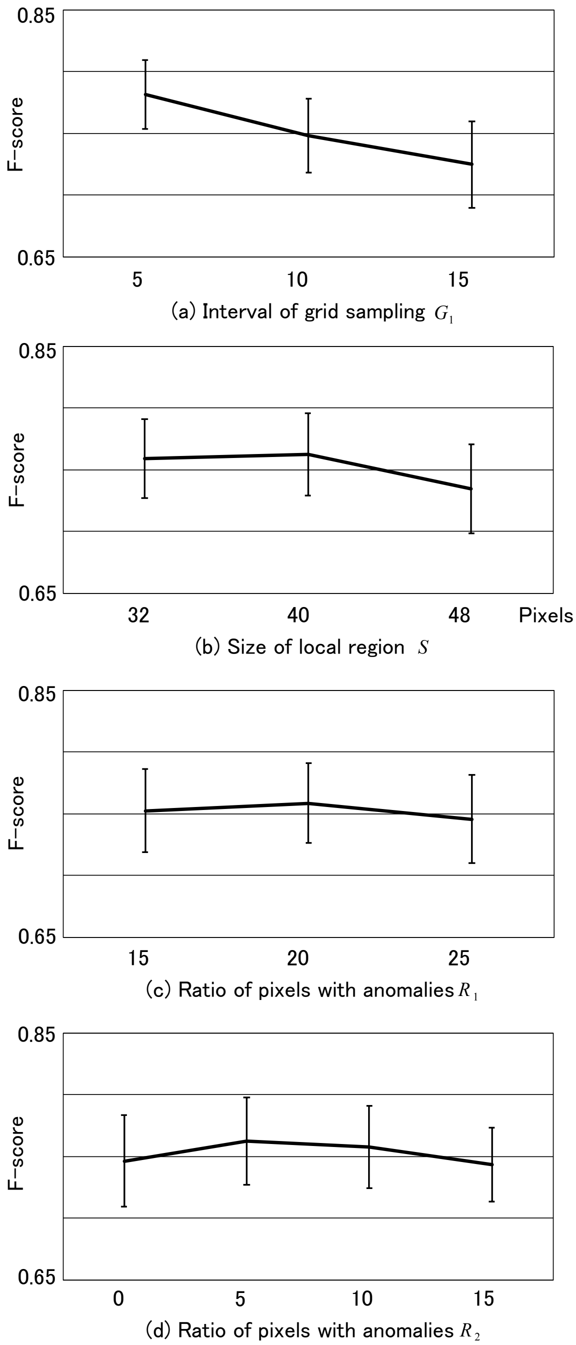

Fig. 6. System F-score performance when one parameter is fixed and the others are changed.

TABLE I
F-SCORES FOR DETECTING ROAD ANOMALY IMAGES USING THE RESULTS OF LOCAL REGIONS.

CNN input	Accuracy (%)
Our method	$\mathbf{0.74 \pm 0.09}$
Baseline	0.55 ± 0.06

images. We organized a dataset of road images acquired from hand-held cameras under the cooperation with the local government. Experimental results show that the use of local regions significantly improves detection performance compared with that of whole images. In future work, we will expand datasets of road images collected in several areas and intend to evaluate detection performance.

Acknowledgement: This work was partially supported by SIP, Infrastructure Maintenance, Renovation, and Management granted by JST. We would like to express our gratitude to Tottori prefecture for their cooperation with collecting the dataset.

REFERENCES

[1] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, "Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2718–2729, 2016.

[2] M. Gavilan, D. Balcones, O. Marcos, D. F. Llorca, M. A. Sotelo, I. Parra, M. Ocana, P. Aliseda, P. Yarza, and A. Amirola, "Adaptive road crack detection system by pavement classification," *Sensors*, vol. 11, no. 10, pp. 9628–9657, 2011.

[3] Y. Jo and S. Ryu, "Pothole detection system using a black-box camera," *Sensors*, vol. 15, no. 11, pp. 29316–29331, 2015.

[4] D. Zhang, Q. Li, Y. Chen, M. Cao, L. He, and B. Zhang, "An efficient and reliable coarse-to-fine approach for asphalt pavement crackdetection," *Image and Vision Computing*, vol. 57, no. C, pp. 130–146.

[5] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," in *Proceedings of the 4th International Conference on Transportation Information and Safety*, 2017, pp. 917–924.

[6] J. M. Alvarez, T. Gevers, Y. LeCun, and A. M. Lopez, "Road scene segmentation from a single image," in *Proceedings of 12th European Conference on Computer Vision Proceedings*, 2012, pp. 376–389.

[7] J. Fritsch, T. Kuhnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *Proceedings of 16th International Conference on Intelligent Transportation Systems*, 2013, pp. 1693–1700.

[8] C. C. T. Mendes, V. Fremont, and D. F. Wolf, "Exploiting fully convolutional neural networks for fast road detection," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2016, pp. 3174–3179.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[10] Hiroya Maeda, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, and Hiroshi Omata, "Road damage detection using deep neural networks with images captured through a smartphone," in *arXiv:1801.09454*, 2018.

[11] W. Liu, D. Anguelov, and D. Erhan, "Ssd: Single shot multibox detector," in *Proceedings of European Conference on Computer Vision*, 2016, pp. 21–37.

[12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[13] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6517–6525.

[14] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proceedings of 9th European Conference on Computer Vision*, 2006, pp. 490–503.

[15] J. R. Uijlings, K. E. Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013.

[16] G. Antipov, S.A. Berrani, N. Ruchaud, and J.L. Dugelay, "Learned vs. hand-crafted features for pedestrian gender recognition," in *Proceedings of the 23rd ACM International Conference on Multimedia*, 2015, pp. 1263–1266.