Autonomous Road Potholes Detection on Video

Jia Juang Koh[1], Timothy Tzen Vun Yap[1(✉)], Hu Ng[1], Vik Tor Goh[2], Hau Lee Tong[1], Chiung Ching Ho[1] and Thiam Yong Kuek[3]

[1] Faculty of Computing & Informatics, Multimedia University, 63100 Cyberjaya, Malaysia
[2] Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Malaysia
[3]Faculty of Business & Finance, Universiti Tunku Abdul Rahman, 31900 Kampar, Malaysia
jiahuangkoh@gmail.com, {timothy, nghu, vtgoh, hltong, ccho}@mmu.edu.my, kuekty@utar.edu.my

Abstract. This research work explores the possibility of using deep learning to produce an autonomous system for detecting potholes on video to assist in road monitoring and maintenance. Video data of roads was collected using a GoPro camera mounted on a car. Region-based Fully Convolutional Networks (R-FCN) was employed to produce the model to detect potholes from images, and validated on the collected videos. The R-FCN model is able to achieve a Mean Average Precision (MAP) of 89% and a True Positive Rate (TPR) of 89% with no false positive.

Keywords: Road Surface Defects, Object Identification, Video Data, Machine Learning, Deep Learning

1 Introduction

Road defects have been a concern for many drivers as they can cause unnecessary accidents and casualties. The accidents are mainly due to road defects such as potholes, sunken or elevated manholes which are extremely common in many big cities or rural roads. Potholes may cause damage to vehicles such as flat tires, torn-off bumpers, bent wheel rims, and damaged shock absorbers. These defects that could have been the cause of many accidents and they can be avoided if the authorities such as the local councils can be quickly notified for repair.

Nevertheless, the degeneration of road is unavoidable because of constant usage and poor weather conditions. In Malaysia, the government has constantly spent a great deal of money to improve Malaysian roads. For instance, the Selangor state government has spent over half a billion ringgits on improving Selangor roads in 2014 [1]. However, allocation of resources for road maintenance proves to be a challenge. It is costly and time consuming for councils to constantly monitor conditions of roads. Thus, this research seeks to create a model to identify road potholes on video to assist in the maintenance of roads, perhaps aided by autonomous drones that monitor roads in future.

© Springer Nature Singapore Pte Ltd. 2019

R. Alfred et al. (eds.), *Computational Science and Technology*, Lecture Notes in Electrical Engineering 481, https://doi.org/10.1007/978-981-13-2622-6_14

2 Literature Review

In terms of road surface defects detection, primarily potholes, cracks and patches, a few research has been performed for image and video. In 2016, Shen et al., have developed road crack recognition application using MATLAB [2]. The software he developed had successfully extracted a distinct road crack feature from images by employing threshold segmentation and edge detection.

Huidrom et al., proposed a Critical Distress Detection, Measurement and Classification (CDDMC) algorithm for automated detection and measurement of potholes, crack, and patches of a series of road surface condition video frames [3]. The CDDMC algorithm had successfully detected and measured these three specific road surface conditions effectively and precisely in one pass.

Kawai et al., proposed a distinction technique for night-time road surface conditions employing a car-mounted camera [4]. They concentrated on the dissimilarity of features in road surfaces condition. Sun et al., proposed a road image status detection technique using a video camera and developed a Naïve Bayesian classifier to classify the road surface condition image [5].

Zhao-zheng et al., proposed a method to estimate visibility distance based on the contrast of road surface condition with distance information using traffic video-surveillance system [6], while Raj et al., developed an algorithm that detects road surface types such as asphalt, cement, sandy, grassy, and rough based on video data taken from a car-mounted camera [7]. These researches employed image processing techniques with machine learning on video, however, none have applied deep learning approaches, which is investigated in this research work.

3 Methodology

The dataset of images used for the development of the detection model is provided by Nienaber et al., for their research work in South Africa [8, 9]. These images were captured by a GoPro Hero 3+ camera in a vehicle travelling at roughly 40 km/h. Each image has its resolution set to 3680×2760 in JPG format. The dataset contains two different sets, one is considered to be simple (easily recognizable potholes) while another is more complex, their file sizes are 10.8 GB and 16.4 GB respectively. Each of the set consist of folders containing the training images as well as a collection of positive test images. The training images consist of positive data (images that contains potholes) and negative data (images without potholes).

Dataset of on road videos for validation were collected using a GoPro Hero 4 Silver camera mounted on the front of the car. Each video has its resolution set to 1920×1080 and its frame rate set to 30 frames per second. Every video of the dataset is in MP4 format. The entire dataset consists of videos of sudden stops, potholes, smooth roads, speed bumps, uneven roads, corner roads, and rumble strips, with a total file size of 55.4 GB.

The TensorFlow Object Detection API [10] was employed to develop the identification model. TensorFlow [11] is an open source library for deep learning developed by Google. Region-based Fully Convolutional Networks (R-FCN) was chosen as the training model for its precise and highly effective object detection capability. It uses position-sensitive score maps to cope with the dilemma of translation-invariance and translation-variance in image classification and object detection respectively [12].

The performance evaluators employed in this research include the Mean Average Precision (MAP), True Positive Rate (TPR), and False Positive Rate (FPR). The MAP is the percentage form of cases where the potholes were detected over all tested cases, given by

$$MAP = \frac{T_p}{T} \times 100\% \tag{1}$$

where T_p is the number of cases that where the object is detected and T is the total number of tested cases.

The TPR is the percentage form of cases that are detected as positive over all positive cases given by

$$TPR = \frac{T_p}{P} \times 100\% \tag{2}$$

where T_p is the number of cases that the object is detected and P is the number of positive tested cases.

The FPR is the percentage of cases that the object is detected wrongly as positive cases over all negative cases given by

$$FPR = \frac{F_p}{N} \times 100\% \tag{3}$$

where F_p is the number of cases with objects wrongly detected as positive cases and N is the number of negative tested cases.

4 Design of Experiments

A set of images that contained potholes was constructed as the positive data. 1000 positive images were selected from the positive dataset to be labeled. A tool, LabelImg [13], was used to annotate the potholes in the images. A XML file was produced for each labeled image, containing data such as coordinates of the bounding boxes, and their height and width. Correspondingly, a set of 1000 images without any potholes were selected as negative images. Fig. 1 shows samples of the positive (top) and negative images (bottom).

The two sets of positive and negative images were further divided into training and testing sets. The training set contained 90% of all positive and negative data respectively, while 10% of the images were chosen from both positive and negative data respectively to make up the testing set.

Fig. 1. Sample images of the training and testing set. Top: positive; bottom: negative.

The XML files of both the training data and validation data were converted into a TFRecord, the data object for labeled training data used by the TensorFlow Object Detection API. In addition, a pbtxt file (textual representation of the TensorFlow graph) was also created as the label map.

Training was initiated using the pre-trained models of the R-FCN and its checkpoint alongside the TFRecord of the training data, as well as the label map. Default paramaters (learning rate, number of layers, number of neurons, number of iterations, Lambda L2-regularization parameter, etc) were used, with softmax as the activation function and an epoch of 1. Parameter optimization will be considered in future research work.

The post-trained model of the R-FCN was then used to export the inference graph, from which the final model was generated. The final model was then applied to the test set to obtain the accuracy result. Finally, the model was validated using the videos taken and the detected potholes in the videos were labeled with bounding boxes for visual inspection. A flowchart depicting the process is shown in Fig. 2.

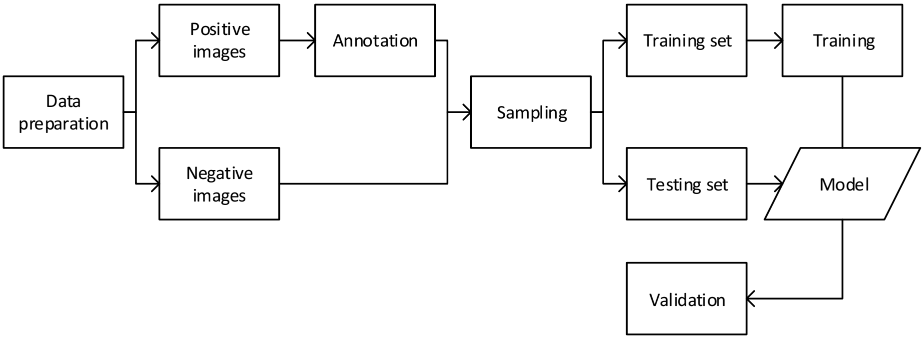

Fig. 2. Flowchart of the experiment process.

5 Results and Discussions

The results of the performance measurement of the model is shown in Table 1. The model is able to achieve a MAP of 89%, with a TPR of 89% and no false positive. In addition, Fig. 3 shows sample results of positive images from the model. The R-FCN can detect most medium sized pothole in bright image successfully, but it has difficulty in detecting potholes in dark or unilluminated areas of the image (Fig.3 – top, right). It also fails to detect small potholes (Fig. 3 – bottom, right).

The R-FCN is able to produce a model with fairly accurate capabilities to detect potholes, but in order to achieve this result, sufficient labeled data has to be available in addition to computing power for continuous training.

Table 1. The MAP and TPR of the model.

MAP	TPR
89%	89%

6 Conclusions

R-FCN was employed using TensorFlow to train a model from images to detect potholes in videos. The training was successful and the R-FCN model is able to achieve 89% MAP and 89% TPR with no false positive. With sufficient labeled data and computational power, R-FCN can achieve high accuracy on pothole detection for use in videos, with limitations in detecting small potholes or potholes in dark or unilluminated areas.

Fig. 3. Positive images with correctly identified potholes.

Acknowledgements. Financial support from the Ministry of Higher Education, Malaysia, under the Fundamental Research Grant Scheme with grant number FRGS/1/2015/SG07/MMU/02/1, as well as the Multimedia University Capex Fund with Project ID MMUI/CAPEX170008, are gratefully acknowledged.

References

1. The Star Online, https://www.thestar.com.my/news/community/2014/03/10/crumbling-roads-exhausting-funds-huge-allocation-comes-with-hopes-of-better-maintenance, last accessed 2018/2/10.
2. Shen, G.: Road crack detection based on video image processing. 3RD INTERNATIONAL CONFERENCE ON SYSTEMS AND INFORMATICS 2016, pp. 912–917, IEEE, Shanghai (2016).
3. Huidrom, L., Das, L. K., Sud, S.: Method for automated assessment of potholes, cracks and patches from road surface video clips. Procedia - Social and Behavioral Sciences, 312–321 (2013).
4. Kawai, S., Takeuchi, K., Shibata, K., Horita, Y.: A method to distinguish road surface conditions for car-mounted camera images at night-time. 12TH INTERNATIONAL CONFERENCE ON ITS TELECOMMUNICATIONS 2012, pp. 668–672, IEEE (2012).

5. Sun, Z., Jia, K.: Road surface condition classification based on color and texture information, 9TH INTERNATIONAL CONFERENCE ON INTELLIGENT INFORMATION HIDING AND MULTIMEDIA SIGNAL PROCESSING, pp. 137–140), IEEE (2013).
6. Zhao-Zheng, C., Jia, L., Qi-Mei, C.: Real-time video detection of road visibility conditions. WORLD CONGRESS ON COMPUTER SCIENCE AND INFORMATION ENGINEERING 2009, pp. 472–476. IEEE (2009).
7. Raj, A., Krishna, D., Priya, H., Shantanu, K., Devi, N.: Vision based road surface detection for automotive systems. 2012 INTERNATIONAL CONFERENCE ON APPLIED ELECTRONICS, pp. 223–228, IEEE (2012).
8. Nienaber, S., Booysen, M.J., Kroon, R.S.: Detecting potholes using simple image processing techniques and real-world footage. SATC 2015, Pretoria, South Africa (2015).
9. Nienaber, S., Kroon, R.S., Booysen M.J.: A comparison of low-cost monocular vision techniques for pothole distance estimation. IEEE CIVTS 2015, IEEE, Cape Town, South Africa (2015).
10. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., Murphy, K.: Speed/accuracy trade-offs for modern convolutional object detectors. CVPR 2017 (2017).
11. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv:1603.04467 [cs.DC], (2016).
12. LabelImg, https://github.com/tzutalin/labelImg, last accessed 2018/2/10.
13. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: Object detection via region-based fully convolutional networks. arXiv:1605.06409 [cs.CV], (2016).