ASPHALT POTHOLE DETECTION IN UAV IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

Yuri V. Furusho Becker[1], *Henrique Lopes Siqueira*[1], *Edson Takashi Matsubara*[2], Wesley Nunes Gonçalves[2], Jose Marcato Jr.[1]

[1]FAENG, [2]FACOM, Universidade Federal do Mato Grosso do Sul, Campo Grande, Brazil

ABSTRACT

Transportation infrastructure needs constant maintenance. Pavement management systems requires reliable and detailed data of the current state of the roads to make effective decisions. Currently, pavement condition evaluation methods are mostly performed manually with visual inspection and interpretations in situ, which is labor intensive, time consuming and expensive. In this paper an experiment was conducted where a set of different configurations and parameters for Convolutional Neural Networks (CNNs) were applied to automatically detect potholes from images taken by an Unmanned Aerial Vehicle (UAV). Results showed that the pre-trained Faster-RCNN Inception ResNet model with reduced anchor box stride and image augmentation applied provides better accuracy compared to several other models tested, obtaining accuracy for this experiment of 70.4% across five-fold cross validation.

Index Terms— Remote Sensing, convolutional neural networks, pavement evaluation, road survey.

1. INTRODUCTION

Considering the importance of the road network for the economy and logistic of a country, it is essential to ensure good pavement performance and safe driving conditions. Therefore, periodic road health monitoring surveys are required to collect information about the current road quality to better manage infrastructure maintenance. These surveys are usually manual visual and they suffer heavily from the associated subjectivity of human decision making [1]. High labor cost, time consumption and dangers intrinsic to road inspections are other factors that reduce the periodicity of manual survey procedures [2].

In comparison, automated surveying systems can be fast, accurate and eliminates the subjectivity involved in human inspections [1]. Combined with the ever-increasing availability of low cost and high-quality Unmanned Aerial Vehicles (UAVs) and digital cameras in the last decade, the necessity of automation enabled research, creation and deployment of many different computer vision systems for automated surveys [3].

In order to take advantage of the increasing affordability of aerial images, on account of the popularization of UAVs, a method based on CNNs is proposed. This method automatically learns features from the images [4], in contrast to the classical hand-craft computed vision algorithms that needs precise tuning of parameters and thresholds [5], specially when related to UAV images, as there is great variation in spatial resolution, illumination, viewpoint and scale compared to standard remote sensing images [6]. It learns features by feeding large quantity of data to the network, which can be easily obtainable with UAVs.

2. PROPOSED METHOD

The objective of this study is to investigate CNN different parameters and models in automatically detecting the presence of potholes in images taken by an UAV. The following parameters were tested to check which produces better results: image resolution, data augmentation, detection algorithm and pre-trained models.

An experiment was conducted where an UAV (DJI Phantom 4 Advanced) was used to collect images from roads that contained several potholes along its length. The chosen pavement segments had elements which were expected to be found in real case scenarios such as: trees casting shadows, parked cars covering part of the asphalt, unclear pavement boundary with the sidewalk and presence of low vegetation. Images were collected from different dates and heights: 2, 9, 25 and 50 meters above ground.

A total of 300 images were captured in different heights, light conditions and contexts. For each image, potholes were manually labeled using a free open-source software named LabelImg. After all potholes were labeled, data augmentation was applied in some data subsets using another free open-source Python library: imgaug. Data augmentation is important in machine learning experiments as it increases the amount of training data [7] and avoids over-fitting, as the model generalizes better [8, 9]. Each image was copied 10 times and augmentations were randomly applied for each image. The operations used were translation, horizontal flip, Gaussian blur and single channel brightness amplifier.

In addition to image augmentation, copied subsets had their resolution reduced. Each image taken by the sensor is an 8.34 MB RGB jpeg file with 5472 x 3648 pixels. The images were resized to 0.5 and 0.8 of their original resolution using bi-linear interpolation. Image resolution impacts accuracy significantly, as reducing image size by half in width and height lowers the accuracy by around 16% [8].

Research in studies using CNNs in small objects helped defining which variables are more expressive when tuning pre-trained models and hyperparameters. A detection model determines how the algorithm solves object detection. It defines the methods taken from receiving the input image to returning the object class probability and location. Models vary greatly in precision and speed. As speed in inference is not a concern in this application, Faster R-CNN (Region-based Convolutional Neural Networks) was the detection algorithm chosen for this work. It is the most accurate system in comparison to SSD (Single Shot MultiBox Detector) or R-FCN (Region-based Fully Convolutional Networks), specially for small objects [8]. This algorithm is composed of two networks. The Region Proposal Network (RPN) generates region proposals where objects might be found. Faster RCNN default parameters are not optimal for objects with reduced pixel size and obtained better results reducing anchor box stride for the Region Proposal Network [10]. For each different configuration dataset 80% was used for training and 20% for testing. As common practice [11], transfer learning was used to modify only the last layer of an already pre-trained model. These procedures and models were not used to report final accuracy, as they were only used to find optimal parameters. After the model with highest precision was found, 5-fold cross-validation was used to report final detection performance. When given an input image, the detector returns bounding boxes around the predicted areas with the highest probability of truly being the object desired. The metric used to evaluate such prediction was Average Precision at 50% Intersection Over Union (AP@0.50IoU). This metric is a recurrent used evaluation formula where the intersection area between the box predicted and the ground-truth box is divided by the union of both boxes [11]. If the IoU proportion is over 0.5, it is considered a correct detection. Further, precision and recall are calculated counting true positives and false positives. Summing all scores for each recall level results in the Average Precision (AP). This was done for each of the 5 models created during the 5-fold cross validation and the reported accuracy of the model is the mean AP from all models.

3. RESULTS

Comparing the same dataset and model, applying data augmentation improves precision by 4% compared to not applying augmentation and decreasing image resolution by 0.5. As for region proposal parameters, it was found a 2% increase when reducing anchor box stride to 8 px compared to 16 px. Lastly, the better performing pre-trained model found for this application was the Faster-RCNN Inception ResNet v2 Atrous pre-trained on the COCO dataset.

With these configurations, 5-fold cross validation obtained a mean value of 70.4% and 7% standard deviation.

4. CONCLUSION

Road pavement inspections are essential for maintenance programs, but new procedures must be adopted to automate the data collection. This work investigates machine learning techniques applied for potholes detection in pavements using UAVs as a cheap and fast way to obtain images. This work found the following: potholes were detected using convolutional neural networks with 70.4% accuracy. Practices such as using the pre-trained Faster-RCNN Inception ResNet model, reducing anchor box stride and image augmentation resulted in better accuracy.

5. ACKNOWLEDGEMENTS

The present work was carried out with a research project financed by CAPES through the program Print-CAPES.

6. REFERENCES

[1] M. Salman, S. Mathavan, K. Kamal, and M. Rahman, "Pavement crack detection using the Gabor filter," *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, , no. Itsc, pp. 2039–2044, 2013.

[2] Yong Shi, Limeng Cui, Zhiquan Qi, Fan Meng, and Zhensong Chen, "Structured Forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 1–12, 2016.

[3] Christian Koch, Kristina Georgieva, Varun Kasireddy, Burcu Akinci, and Paul Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," 2015.

[4] Y. Long, Y. Gong, Z. Xiao, and Q. Liu, "Accurate object localization in remote sensing images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 2486–2498, May 2017.

[5] S Varadharajan, S Jose, K Sharma, L Wander, and C Mertz, "Vision for road inspection," in *IEEE Winter Conference on Applications of Computer Vision*, 3 2014, pp. 115–122.

[6] Y. Bazi and F. Melgani, "Convolutional svm networks for object detection in uav imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3107–3118, June 2018.

[7] Luis Perez and Jason Wang, "The effectiveness of data augmentation in image classification using deep learning," *CoRR*, vol. abs/1712.04621, 2017.

[8] Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy, "Speed/accuracy trade-offs for modern convolutional object detectors," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 3296–3305, 2017.

[9] Yun Ren, Changren Zhu, and Shunping Xiao, "Small Object Detection in Optical Remote Sensing Images via Modified Faster R-CNN," *Applied Sciences*, vol. 8, no. 5, pp. 813, 2018.

[10] Chenchen Zhu, Ran Tao, Khoa Luu, and Marios Savvides, "Seeing Small Faces from Robust Anchor's Perspective," vol. 1, 2018.

[11] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016.