Training a tandem semantic segmentation and disparity map network for road profile calculation

Training a Tandem Semantic Segmentation and Disparity Map Network for Road Profile Calculation

Lydon, D. [1], Taylor, S.E.[2], Early, J.[1]Sonebi, M[2]
[1]Mechanical and Manufacturing Engineering, Queen's University Belfast, N.Ireland BT7 1NN
[2]Natual and Built Environment Queen's University Belfast, N.Ireland BT7 1NN
Email : D.Lydon@qub.ac.uk

ABSTRACT: Inspection of bridge structures is an important consideration for governmental bodies as early detection of structural damage can facilitate early repairs, theoretically reducing maintenance costs for underfunded infrastructure management departments. This paper investigates the use of a computer vision-based system for road profile analysis, which can then be used as part of a larger damage detection solution. A tandem semantic segmentation and disparity map generation convolutional neural network is developed and trialled on datasets collated for this research. A discussion of the obtained results followed by recommendations for future work concludes the paper.
KEY WORDS: Computer Vision, Deep Learning, Road Profile Mapping

1 INTRODUCTION

A functional road network is key to the social and economic development of a nation [1]. 79% of freight transported in the UK in 2019 was by road, with 75% of commutes to work also occurring by road [2]. The volume of traffic on our road networks is increasing yearly, with an increase in car journeys of 7.2%, an increase in Light Commercial Vehicle journeys of 17.6% and an increase of 7.5% in HGV journeys compared to 5 years ago [3]. Bridges are a critical element of the road network, and are particularly vulnerable to damage from vehicle strikes, flooding and overloading. The estimated cost to bring just the substandard bridges on the network to perfect condition is £1.12bn, last year the budget allocation was £93m [4]. These reduced budgets mean that careful consideration must be given to allocation of funds. A report from the National Audit Office [5] emphasised the importance of unbiased, consistent data collection to facilitate the proper allocation of limited budgets and develop a clear picture of the current state of transport infrastructure. Structural Health Monitoring (SHM) can provide an objective and data driven means of collecting information about bridge structures to be used to help decision making for asset managers/owners. This paper will provide a brief overview of current methods for semantic segmentation and disparity calculation in road profile analysis before concluding with the methodology and development of a convolutional neural network for road profile analysis.

2 LITERATURE REVIEW

There have been numerous studies in applied deep learning to road analysis. A study by [6] used transfer learning, a process where a CNN is initially trained on a very large general dataset, and then the later layers are retargeted for another application, to develop a system for segmentation of road surface to detect cracks on a pixel level. Segmentation is the process of assigning each pixel in an image with a predefined class label. The output of a segmentation algorithm with only two classes for example would be a binary image, where pixels with value 0 belong to one class and pixels with value 1 belong to the other. Segmentation of a disparity map was carried out in [7], in this case a thresholding method was used in place of CNN for the segmentation process. A disparity map is the pixel difference between two rectified stereo images and is used to calculate depth or distance from the camera in images. A tandem method where segmentation and classification CNN's were combined to detect and classify pavement cracks was laid out in [8], this dual purpose network was the inspiration for the research developed in this paper. The novel contribution from this paper is the use of real time 3d content creation paired with advanced data augmentation to create a diverse dataset for training a tandem CNN in road profile calculation, to the knowledge of the authors this has not been attempted in this manner previously.

3 METHODOLOGY

3.1 Data Collation and Augmentation

The datasets for this paper were collected from the following sources: The Cityscapes Dataset for Semantic Urban Scene Understanding [9] , Cambridge Labeled Video Database [10], the KITTI Object Scene Flow for Autonomous Vehicles Dataset [11] and the Middlebury Stereo Vision dataset[12]. These datasets were chosen because they offered depth map information in addition to semantic segmentation labels for each scene. The datasets were pre-processed to ensure they were of the same dimensions as CNN structures expect images to all be of uniform size. An additional portion of the data engineering pipeline was changing the segmentation masks in the collated training datasets from multiclass to binary, i.e. road surface and non-road surface. This was done because our network is only being used to calculate road profile, so calculating disparity of other portions of the image would be a waste of computational resources.

3.2 Synthetic Data Generation

In order to develop a more diverse collection of images for training the CNN structure, it was decided to investigate using a real time 3D content creation engine to generate realistic images. The engine chosen was Unity [13], a games engine that has recently developed a simulation suite for this purpose. Using a games engine drastically reduces the overhead of labelling elements in a scene and allows for accurate depth map generation as this can be accessed from the engine through in-built scripting functionality. The scenes are created in the engine as normal, any assets that are loaded are tagged with a label for segmentation that are accurate to pixel level. Changes such as varying lighting/weather conditions etc that would be difficult to capture in real world environments can be easily created in game engines, resulting in a more diverse dataset that will enavle the CNN to perform well in field testing.

3.3 Data Augmentation

Another method for increasing the size of a dataset for CNN training is data augmentation. This involves performing mathematical operations on the images such as translation, rotation, cropping, zooming, flipping & colour space modifications. For this research, the augmentation was carried out using the NVIDIA DALI [14] pipeline creation tool. DALI enables significantly faster augmentation operations to be carried out as it performs all operations on the GPU, which allows for data parallelism to be employed, increasing data throughput.

3.4 CNN Structure & Training

The developed CNN was constructed in two parts, one feeding into the other. The first portion of the network (SemCNN) was a semantic segmentation network based on U-Net. The U-net architecture used in this research was pre-trained on the Google OpenImages v4[15] dataset. U-net was originally developed for biomedical research but has been successfully applied in numerous other segmentation domains. The Vgg-19 [16] backbone was used in this implementation of U-Net. The left and right rectified images were fed into the SemCNN and binary masks identifying road profiles were generated. The masked images were then fed into the disparity map generation CNN (DMapCNN). The DMapCNN was loosely based on the structure from [17], but with some resizing of the convolution filter sizes and with leakyReLU as the activation function for all layers. This was done to speed up training and prevent the issue with reLU losing some units due to zero activation.

4 RESULTS

The dataset was split on a 60-20 -20 training – validation - test outline. The network was trained using a randomized search hyperparameter optimisation process for 120 epochs with early stopping, the optimal values for the cost function were determined at epoch 97. On the test set, a mean Intersection over Union (mIoU) of 0.53 when compared to ground truth was the final result on the test set after training. A sample output from the network is shown in Figure 1.

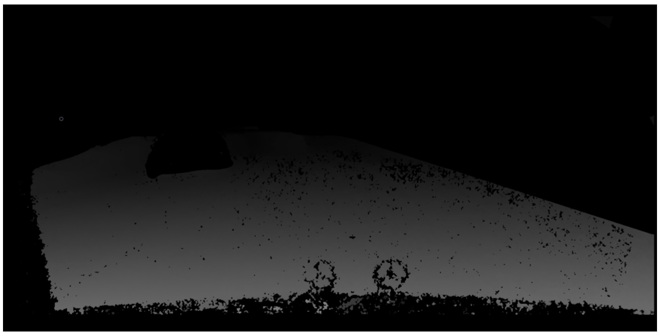

Figure 1 Output from Proposed CNN

5 CONCLUSIONS & FUTURE WORK

The proposed network has proven successful in limited cases to be able to segment and accurately generate disparity maps at a resolution useful for road profile generation. However, the relatively low mIoU on the whole dataset is cause for concern. Future work in this area will consist of further structural experimentation on the internal layers of the dataset in addition to more analysis on data output to attempt to determine where the errors in the outputs are occurring. Other work that will be carried out is the creation of additional scenes in Unity to increase the amount of data available to the CNN for training. If this work is successful, the authors intend to collect real world data and trial the network in the field in real time on a fleet of zero -emissions busses being developed alongside this project.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of the EPSRC Prosperity Partnership (EP/S036695/1) for their support on this project.

REFERENCES

[1] E. Ivanová and J. Masárová, “IMPORTANCE OF ROAD INFRASTRUCTURE IN THE ECONOMIC DEVELOPMENT AND COMPETITIVENESS,” *Econ. Manag.*, vol. 18, no. 2, pp. 263–274, Aug. 2013, doi: 10.5755/j01.em.18.2.4253.

[2] Department for Transport, “Transport Statistics Great Britain 2019,” London, 2019.

[3] Department for Transport, “Provisional Road Traffc Estimates Great Britain 2019,” London, 2019.

[4] RAC Foundation, “Bridge maintenance backlog remains static,” *RAC Foundation*, 2020. [Online]. Available: https://www.racfoundation.org/research/economy/number-of-substandard-road-bridges. [Accessed: 10-Feb-2021].

[5] National Audit Office, “Maintaining strategic infrastructure: roads,” London, 2015.

[6] M. D. Jenkins, T. A. Carr, M. I. Iglesias, T. Buggy, and G. Morison, “A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks,” in *European Signal Processing Conference*, 2018, vol. 2018-Septe, pp. 2120–2124, doi: 10.23919/EUSIPCO.2018.8553280.

[7] R. Fan and M. Liu, “Road Damage Detection Based on Unsupervised Disparity Map Segmentation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 11, pp. 4906–4911, Nov. 2020, doi: 10.1109/TITS.2019.2947206.

[8] H. Majidifard, Y. Adu-Gyamfi, and W. G. Buttlar, “Deep machine learning approach to develop a new asphalt pavement condition index,” *Constr. Build. Mater.*, vol. 247, p. 118513, Jun. 2020, doi: 10.1016/j.conbuildmat.2020.118513.

[9] M. Cordts *et al.*, “The Cityscapes Dataset for Semantic Urban Scene Understanding,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 3213–3223, Apr. 2016, doi: 10.1109/CVPR.2016.350.

[10] G. J. Brostow, J. Fauqueur, and R. Cipolla, “Semantic object classes in video: A high-definition ground truth database,” *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, Jan. 2009, doi: 10.1016/j.patrec.2008.04.005.

[11] M. Menze and A. Geiger, “Object scene flow for autonomous vehicles,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 3061–3070, 2015, doi: 10.1109/CVPR.2015.7298925.

[12] H. Hirschmüller and D. Scharstein, “Evaluation of cost functions for stereo matching,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383248.

[13] Unity3d, “Unity Real-Time Development Platform | 3D, 2D VR & AR Engine,” *Unity Technologies*, 2020. [Online]. Available: https://unity.com/. [Accessed: 19-May-2022].

[14] NVIDIA, “NVIDIA Developer Data Loading Library (DALI),” *developer.nvidia.com*, 2022. [Online]. Available: https://developer.nvidia.com/dali. [Accessed: 19-May-2022].

[15] A. Kuznetsova *et al.*, “The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale,” *Int. J. Comput. Vis.*, vol. 128, no. 7, pp. 1956–1981, 2020, doi: 10.1007/s11263-020-01316-z.

[16] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” 2014, doi: 10.1.1.740.6937.

[17] V. M. Mendoza Guzmán, J. M. Mejía Muñoz, N. E. Moreno Márquez, P. I. Rodríguez Azar, and E. Santiago Ramírez, “Disparity map estimation with deep learning in stereo vision,” in *CEUR Workshop Proceedings*, 2018, vol. 2304, pp. 27–40.