

Identifying Potholes under Challenging Weather Conditions

Amita Dhiman and Reinhard Klette

School of Engineering, Computer and Mathematical Sciences

Department of Electrical and Electronic Engineering

Auckland University of Technology, Auckland, New Zealand

Email: amita.dhiman@aut.ac.nz

Abstract—This paper addresses the problem of identifying potholes on roads. The model discussed in this paper is able to identify dry and wet potholes. The proposed model employs a transfer-learning-based approach; using Mask-RCNN, we use weights trained for the CCSAD model to identify dry potholes. This approach provides promising results on four different datasets recorded under challenging weather conditions, from bright-sunny to dark-cloudy days. The aim of this research is to identify potholes at instance level segmentation. Evaluated results demonstrate very high accuracy of detected potholes compared to our previous model trained by using one dataset consisting of dry potholes only.

I. INTRODUCTION

Dangerous potholes not only cause a damage to a vehicle; they are also a threat to human life. In 2011, the *Christchurch Infrastructure Rebuild Team* was formed to rebuild Christchurch’s earthquake damage. After seven years, Christchurch is still mentioned as the “pothole capital” of New Zealand, as published in one of the regional newspaper [1]. In 2017, different city councils in New Zealand received a number of complaints about potholes such as 276 requests in Auckland only, for compensation for damages or injuries that have been related to road surface conditions; Christchurch’s council received 994 complaints about potholes, and Wellington’s Council received 394 [2].

Depending on the severity of potholes, they are a nuisance of varying degree both for passengers and to the country’s economy. For all the received pothole complaints in 2017, Christchurch council spent 525,000, Wellington council 12,782, Invercargill city council 60,000, and Dunedin city council around 27,000 [1].

Regarding driverless cars, major companies such as Tesla, Toyota, Ford, or BMW announced to be able to deliver autonomous cars by about 2020 [3]. However, research in the field of automatic identification of road potholes still needs to follow the successes in other subject areas. In the already defined digital world, the reporting and identification of potholes still depends mainly on public reporting [4], [5].

The greatest challenge to solve the problem of identification of potholes is the absence of robust and advanced methods using automatic identifications of potholes. Automated detection methods can save large sums spend on vehicle damage, and can also make road repair systems more efficient.

Current methods use a variety of sensors such as inertial measurements, 3-dimensional (3D) scanners, and optical sensors. Vibration-based methods use an accelerometer for measuring the vibrations caused in a vehicle due to hitting a pothole or other road damages [6], [7].

Other techniques such as 2D-vision based methods [8], [9] rely heavily on manual processing; it appears impractical to define 2D features of potholes due to their irregular shape.

Other sensors such as ground penetrating radar [10], 2D laser scanners [11], 3D laser [12], or stereo vision cameras [13] offer an accurate option based on 3D reconstruction [14], [15]. Ground penetrating radar is mainly used for special purposes. Stereo vision cameras or a 3D laser are the logical choice for identifying a pothole from a distance, where a 3D laser is (still) an expensive sensor, also usually not applicable when a pothole is filled with water.

The current state-of-art methods employ deep-learning-based methods (see Fig. 1) to identify road damage. However, to the best of our knowledge, there is not yet any report on focused research based on deep learning in fields related to pothole identification. To develop such a deep-learning-based model to identify potholes, 65% of the total efforts are generally spent on the collection and preparation of a road-pothole dataset; see Fig. 1. The purpose of this study is to introduce four datasets while providing insights into related findings based on experiments

The rest of the paper is organised as follows. Section II presents a review of related literature. Sections III gives the introduction to four different datasets that we used. Section IV demonstrates experimental results of this study. Section IV concludes.

II. RELATED WORK

Pothole-detection approaches can be divided into two types - first, conventional image-analysis-based algorithms, and second, deep-learning-based models.

Based on conventional computer-vision algorithms, Dhiman et al. [16] proposed a method for identification of potholes using a disparity map. The algorithm models a road manifold in 3D space using RANSAC for iterative optimization. An elevation map is built for the road surface following the identification of the road manifold. Pixels representing points

below ground-manifold level are considered for a connected-component analysis. Finally, morphological operations are used to find salient pothole regions in the disparity image. Authors of [17] used a multi-frame fusion technique (based on visual odometry) for detecting potholes. The technique works by accumulating multiple-frame 3D reconstructions which are properly aligned into a road-centred coordinate system. A 3D plane-fitting technique is first carried out to approximate the road manifold at the beginning of the accumulation, followed by the construction of a digital elevation model (for the road being analysed) from integrated multiple frames.

Akarsu et al. [22] classified defects in road-surface images into three types: *Horizontal*, *vertical*, or *crocodile*; Zalama et al. [23] classified only into vertical or horizontal. Chun et al. [24] proposed an automated asphalt pavement crack detection technique using a combination of a naive Bayes-based method and image processing.

Kulkarni et al. [25] used an android smart phone to identify changes in acceleration to detect potholes; the authors used a high frequency filter and a neural network approach. Eriksson et al. [26] proposed a pothole patrol system to detect potholes which gathers data from vibration and GPS sensors and processes this data to access road-surface conditions.

Bhatt et al. [27] used a machine-learning approach to detect potholes, and assess the road condition. Hsu et al. presented in [28] a multi-sensor approach to detect potholes and measure a road-quality index by integrating laser, camera, GPS, and *inertial measurement unit* (IMU) sensors into an experimental golf-cart system. The analysis is based on imagery, laser-scanned 3D data, and the data of the other sensors.

As examples for the second type, Zhang et al. [19] suggested a CrackNet to predict class scores for all the pixels in existing road damage, and Maeda et al. [18] employed a state-of-the-art CNN to train a model for detecting road-surface damage using a large-scale dataset of road images collected in Japan using a smart phone by the authors. The authors used SSD Inception V2 [20] and SSD MobileNet [21] to classify eight types of road damages. Detected road damages are identified within enclosed bounding boxes.

Staniek [29] used a form of a recurrent artificial neural network (i.e. a Hopfield neural network) to solve the complex problem of matching points in stereo images to perform depth

analysis. The author provides an introduction into a variety of issues faced in case of stereo vision, ranging from acquiring images of roads to calculate 3D point clouds. The author proposed a fusion of sensors, such as stereo vision cameras, inertial sensors, and a GPS, for solving depth discontinuities. Depth discontinuities may occur due to sudden significant changes in image intensities.

To the best of our knowledge, no example of an application has been published so far to detect potholes at pixel level. Extensive research has been carried out already for image segmentation using CNNs such as RefineNet [30], PSPNet [31] or Large-Kernel-Matters [32]. CNNs for image segmentation may also be of relevance, such as the *fully convolutional neural network* (FCN) by Long et al. [33], in which a final fully-connected layer is replaced by another convolutional layer for a large receptive field to capture the global context of a scene. However, this results into coarse segmentation maps by upsampling layers of the FCN.

Badrinarayan [34] proposes *Segnet*, a multiclass deep-encoded-decoder-based CNN, that is more memory-efficient than the FCN and performs semantic pixelwise segmentation. Segnet eliminates the need of upsampling, as this decoder uses pooling indices, computed in the max-pooling step of the corresponding encoder, for non-linear upsampling.

One more class of CNNs, which uses dilated or atrous convolutions, is proposed in DeepLab by Chen [35]. The dilated convolutions in DeepLab help to increase the field of view exponentially without increasing the number of parameters. However, this type of convolutions is computationally very expensive because of its application for high-resolution feature maps.

There is currently strong progress towards object detection and recognition based on deep learning; a common issue is the lack of training data. Therefore, reflecting this common problem in our paper, we detect potholes by pixel-level segmentation. We use a state-of-the-art deep learning approach to achieve instance segmentation to delineate the boundary of potholes at pixel level in an image. Because we have a (very) limited amount of labelled data only (considering the unlimited diversity of pothole appearances), we initially developed a model to identify dry potholes using a transfer learning based approach.

Transfer learning can be defined as a method where a model, developed for one task, can be reused as a starting point for a model for another task, such as when proceeding from a large source domain dataset like COCO [36] or Imagenet [37] to a smaller target domain.

Generally, in transfer learning the feature learnt in source data is exploited to improve generalization in the target domain dataset. Feature spaces usually differ between source and target domain data, and the aim is to boost the performance rapidly in the target domain. We used as source domain the COCO dataset, trained using Mask R-CNN [38], and as target domain we used different datasets, to be specified below. The aim of this study is actually to put forward different datasets for road pothole identification as well as a novel method for

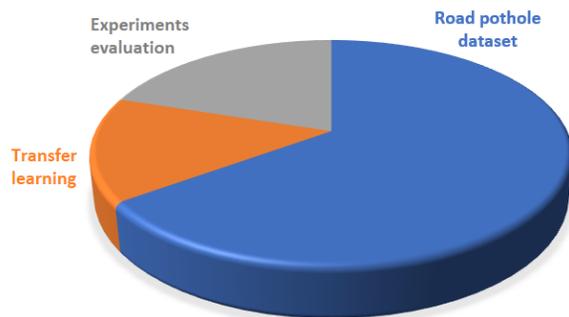


Fig. 1. Efforts required to develop a deep-learning-based model.

detecting potholes at pixel level.

As a base network, we used Mask R-CNN. Mask R-CNN was extended in 2017 by using Faster R-CNN [39], to predict segmentation masks for each region of interest, along with classification and bounding box regression. To delineate the boundary of each instance at pixel level, Mask R-CNN uses a small mask branch FCN, applied for each ROI to predict a mask.

III. DATASETS

1. CCSAD. Hayet et al. [40] introduced an image dataset (recorded in Mexico), called *challenging sequences for autonomous driving* (CCSAD), that is useful for executing methods to detect damaged road surfaces; see Fig. 2, top left, for an example. The dataset has been acquired at 20 fps by using two Basler Scout *scA1300-32fm* firewire greyscale cameras. Each image is of dimension $1,018 \times 765$.

The CCSAD dataset comes in four different categories: Colonial Town Streets, Urban Streets, Avenues and Small Roads, and Tunnel Network. CCSAD accounts for 500 GB of data that incorporate calibrated and rectified pairs of stereo images, videos, and meta-data.

2. DLR. This dataset has been recorded while using the *integrated positioning system* (IPS) [41], [42], developed by the German Aerospace Centre (DLR), installed on a car. The dataset has 48,913 images, recorded by a GoPro camera, mounted behind a car’s windshield. The car was moving at an average speed of 40 km/h while scanning the road surface. Each image has a size of $1,360 \times 1024$; see Fig. 2, top right, for an example.



Fig. 2. Examples from four data sets. Top, left: CCSAD. Top, right: DLR. Bottom, left: Japan. Bottom, right: Sunny.

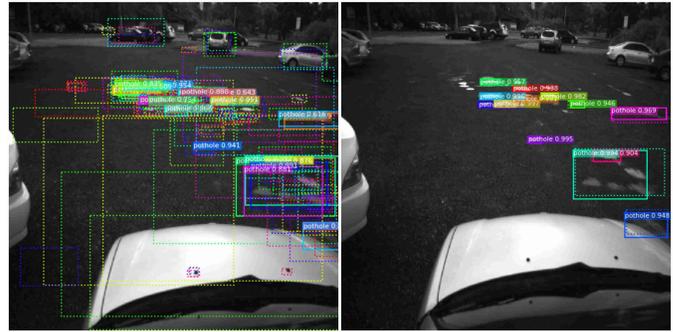


Fig. 3. Illustration of detection steps.

3. Japan. This dataset comprises of 163,664 road images of dimension 600×600 collected in seven different cities of Japan [18]; see Fig. 2, bottom left, for an example. The dataset contains 9,053 damaged-road images and 15,435 instances of damaged road surfaces such as cracks or potholes. The authors used a smartphone on the dashboard of a moving car driving at an average speed of 40 km/h. Images are captured at an interval of one second under different weather and lighting conditions.

4. Sunny. Authors of [44] provided a dataset of 48,913 images of size $3,680 \times 2,760$ recorded using a GoPro camera, mounted behind a car’s windshield. The camera was set to a 0.5 second time lapse mode. The car was moving at an average speed of 40 km/h while scanning the road surface; see Fig. 2, bottom right, for an example.

IV. EXPERIMENTS AND DISCUSSION

For transfer learning, we used 143 images of the CCSAD, DLR, Japan and Sunny dataset collectively. We labelled pothole masks manually in JSON format. We used a batch size of 2; for 2,000 iterations it took 12 hours on a Ge Force GTX GPU. In this study, we used RESNET101 as a backbone architecture and a learning rate of 0.001. We train the network using stochastic gradient descent and a learning momentum of 0.9.

As we used images from four different datasets, the image dimensions were different. To keep an aspect ratio of uniform size $1,024 \times 1,024$, zero padding is added to the top and bottom of an image. We have two classes in our dataset, one



Fig. 4. Original and masked image.

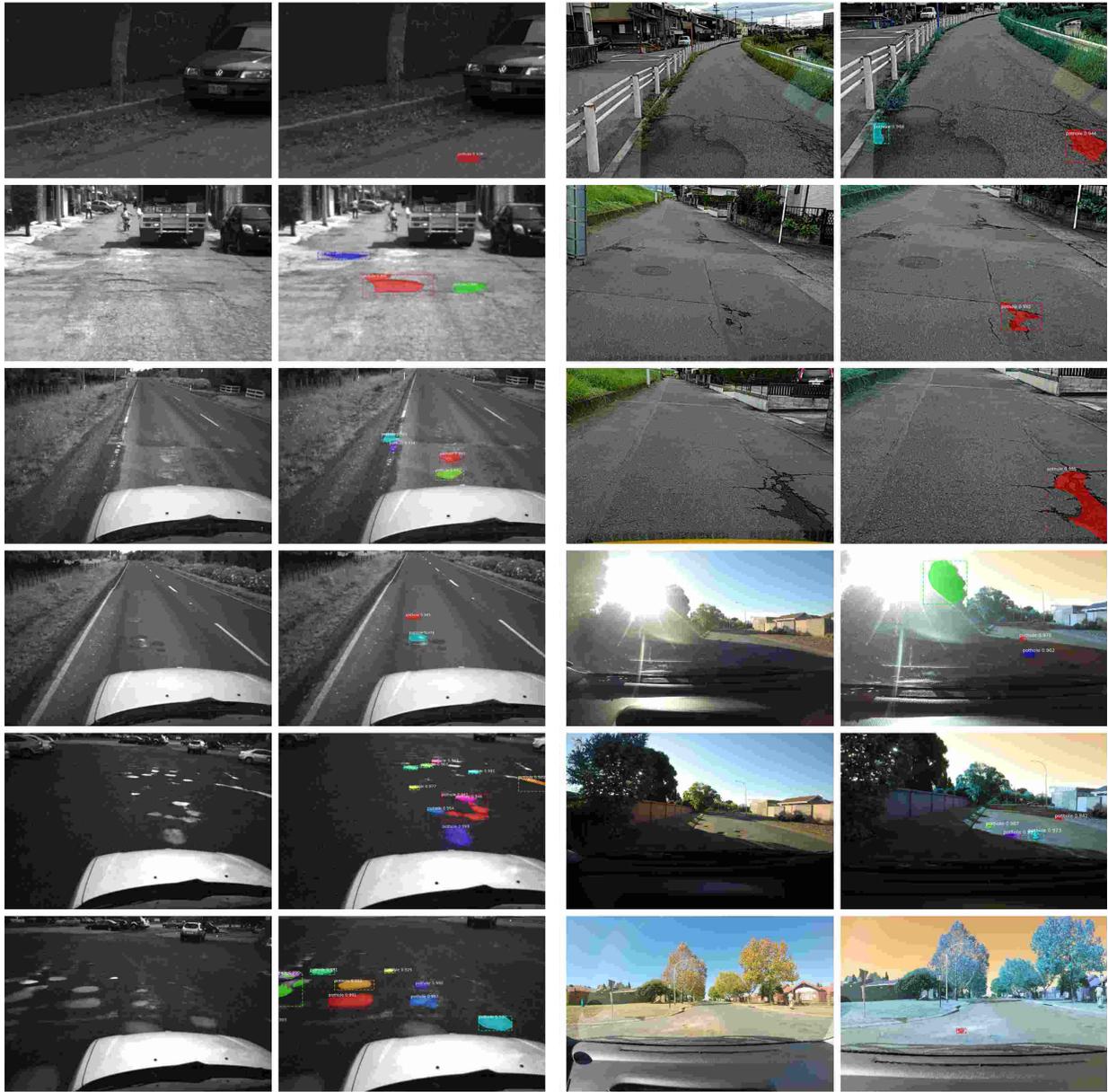


Fig. 5. Detected “potholes” based on transfer learning using Mask R-CNN, shown in two columns with original image on the left and labelled results on the right. *Top to bottom, left to right*: Twelve frames in order as listed in Table I.

for “background” and one for “pothole”. Transfer learning with Mask R-CNN is a two-stage framework.

Stage 1: Classification and bounding box refinement.

During the first stage, the whole training image is scanned to generate anchor proposals by fine tuning RPN from end-to-end. RPN is a lightweight neural network that scans over the backbone’s feature map, using a sliding window to generate anchors. Anchors are typically bounding boxes in the image to predict multiple regions while a small $n \times n$ window slides over the convolved feature map of the entire training image. As the sliding-window operation is convolutional in nature, so it is handled fast on a GPU. This stage generates a maximum of 256 anchors per image and bounding-box refinements. This

stage outputs a grid of anchors (see Fig. 3, left) at different scales. $\text{IoU} > 0.7$ define positive samples, and $\text{IoU} < 0.3$ define negative samples, respectively.

The bounding box refinement step accepts a refined grid of anchors from the RPN and classifies the anchors precisely. It maps anchor bounding boxes, as shown in dotted lines in Fig. 3, left, into final boxes as shown in solid lines in Fig. 3, right. Mask-RCNN refines the ROIAlign layer by removing a harsh quantization of the RoIPool layer, to properly encapsulate the extracted features with the input.

Stage 2: Mask generation. The mask branch is a CNN that accepts positive regions as input, generated by the classifier during Stage 2, and generates a low resolution 28×28 soft

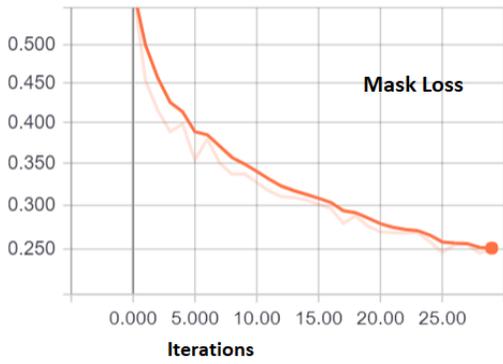


Fig. 6. Decreasing mask loss in dependency of iterations.

mask for it. A soft mask differs from a binary mask as these are represented by float numbers and represent more details. We fine-tune the layer of Mask R-CNN, according to our object class “pothole”, and see the final result as shown in Fig. 3, right. To calculate a loss mask, average cross-binary entropy [45] is used, in which only the k -th mask is included if the region is associated with the ground truth:

$$M_{Loss} = \frac{-1}{m^2} \sum_{1 \leq i, j \leq m} [y_{ij} \log \hat{y}_{ij}^G + (1 - y_{ij}) \log(1 - \hat{y}_{ij}^G)] \quad (1)$$

where y_{ij} is the label of an anchor of dimension i, j for the true mask of size $m = 28$. We calculate the loss by a predicted value \hat{y}_{ij}^G of the same anchor in a mask learned for ground truth class G .

For visualization purposes, a loss is calculated after completion of every epoch. As loss acts as a penalty on the network, it is clear from Fig. 6 that the loss decreases in each iteration.

We first developed a model based on transfer learning using Mask R-CNN with CCSAD frames and tested our model on sequence 1 and 2 of this dataset. For each selected frame, the ground truth is established manually. However, the developed model was able to correctly identify dry potholes in properly illuminated scenes as shown in Fig. 5, second row, first image. But such a model is not applicable to identify wet potholes in real-world scenarios. Thus, to identify potholes under varying (challenging) conditions, we collected and used a variety of

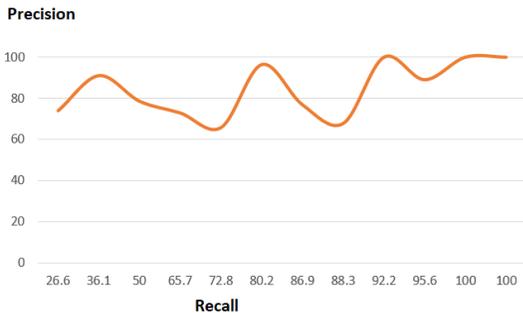


Fig. 7. Precision-Recall curve.

datasets and trained the network using frames from all the different datasets, collectively.

In Fig. 5, twelve tested frames are shown in two columns from the four used datasets introduced in Section III. Left column, the two top-most frames are from CCSAD sequences 1 and 2, identifying dry potholes under different illuminating conditions. Left column, the four other tested frames are from the DLR data showing wet potholes. Right column, the three tested frames on the top are from the Nagakute and Numazu data in the Japan dataset; identified potholes are emerging from big cracks. Right column, the other three tested frames are from the Sunny dataset; they show two images at the bottom of the column where networks of pothole instances are correctly identified while a false positive has been detected in the third image (from the bottom) – as a pothole is of arbitrary shape, under bright sunshine a tree is miss-classified as a pothole in this case (this could be excluded by identifying a ground manifold first).

As we are interested in identifying potholes at pixel level, we provide a quantitative analysis using calculated common classification measures:

$$\text{Precision} = \frac{tp}{tp + fp} \quad (2)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (3)$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} \quad (4)$$

where tp denotes the number of true positives, fp of false positives, fn of false negatives, and tn of true negatives. Results are listed in Table I.

TABLE I
EVALUATION MEASURES (IN %) FOR RESULTS SHOWN IN FIG 5.

Dataset	Frame	Precision	Recall	Accuracy
CCSAD	547	89.0	95.6	99.7
CCSAD	345	96.4	80.2	99.9
DLR	472000	67.7	88.3	99.7
DLR	572000	100.0	100.0	100.0
DLR	749000	100.0	92.2	99.8
DLR	449000	91.0	36.1	95.6
Japan	20170912135214	72.8	65.7	99.3
Japan	20170906135035	76.8	86.9	99.8
Japan	20170906135037	96.4	72.8	99.2
Sunny	G0010116	73.9	26.6	98.9
Sunny	G0010118	100.0	100.0	100.0
Sunny	G0011873	78.5	50.0	99.9
Overall		87.0	81.3	99.7

In some cases, our model is able to identify potholes at 100% of ground truth pixels; the potholes in these cases are usually wet potholes and easy to label manually. Our model also achieves an overall Precision of 87% and Recall of 81% while achieving an of Accuracy of 99.7%. To visualize the relationship between classification measures, a Precision-Recall plot is shown in Fig. 7.

V. CONCLUSIONS

More variations in training data improve the detection accuracy. More benchmark datasets with exact ground-truth

labelling are needed for universal types of experiments. Having already data from four different countries certainly contributed to the documented good results. The trained network even identified “crack patterns” of emerging potholes which were difficult to see by a human. The novel pixel-level identification of potholes proved to be a useful approach, both for academic and practical purposes. Our main purpose was to identify road distress by exact masks, not just by bounding boxes.

REFERENCES

- [1] Christchurch report, www.stuff.co.nz/the-press/news/100847641/christchurch-the-pothole-capital-of-new-zealand/, 2018.
- [2] Auckland report, www.neighbourly.co.nz/e-edition/east-bays-courier/article/2027329, 2018.
- [3] Driverless car market watch, www.driverless-future.com/?page_id=384, 2018.
- [4] B.-H. Lin and S.-F. Tseng, “A predictive analysis of citizen hotlines 1999 and traffic accidents: A case study of Taoyuan city,” in *Proc. Int. Conf. Big Data Smart Computing*, 2017, pp. 374–376.
- [5] D. Santani, J. Njuguna, T. Bills, A. W. Bryant, R. Bryant, J. Ledgard, and D. G. -Perez, “CommuniSense: Crowdsourcing road hazards in Nairobi,” in *Proc. Int. Conf. Human-Computer Interaction Mobile Devices Services*, 2015, pp. 445–456.
- [6] C.-W. Yi, Y. -T. Chuang, and C. S. Nian, “Toward crowdsourcing-based road pavement monitoring by mobile sensing technology,” *IEEE Trans. Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1905–1917, 2015.
- [7] A. Mednis, G. Stardins, R. Zviedris, G. Kanonirs, and L. Selavo, “Real time pothole detection using smartphones with accelerometers,” in *Proc. Int. Conf. Distributed Computing Sensor Systems Workshops*, 2011, pp. 1–6.
- [8] H. Kong, J. Y. Audibert, and J. Ponce, “General road detection from a single image,” *Image Processing*, vol. 19, no. 8, pp. 2221–2220, 2010.
- [9] C. Koch and I. Brilakis, “Pothole detection in asphalt pavement images,” *Advanced Engineering Informatics*, vol. 25, no.3, pp. 507–515, 2011.
- [10] U. Spagnolini, and V. Rampa “Multitarget detection/tracking for monostatic ground penetrating radar: Application to pavement profiling,” *IEEE Tran. Geoscience Remote Sensing*, vol. 37, no. 1, 383–394, 1999.
- [11] M. Xiangyang, L. Lin, T. Nan, C. Liu, and J. Xiuhan “Laser-based system for highway pavement texture measurement,” *Proc. IEEE Int. Conf. Transportation Systems*, 2003, 2, pp. 1559–1562.
- [12] J. Laurent, M. Talbot, and Doucet “Road surface inspection using laser scanners adapted for the high precision 3D measurements of large flat surfaces,” in *Proc. Int. Conf. Recent Advances*, 303–310, 1997.
- [13] R. Klette, “*Concise Computer Vision: An Introduction into Theory and Algorithms*,” London: Springer, 2014.
- [14] X. Ai, Y. Gao, J.G. Rarity, and N. Dahnoun, “Obstacle detection using U-disparity on quadratic road surfaces,” in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 2013, pp. 1352–1357.
- [15] F. Oniga, S. Nedeveschi, M. M. Meinecke, and T. B. To, “Road surface and obstacle detection based on elevation maps from dense stereo,” in *Proc. IEEE Int. Conf. Intelligent Transportation Systems*, 2007, pp. 859–865.
- [16] A. Dhiman, H.-J. Chien, and R. Klette, “Road surface distress detection in disparity space,” in *Proc. Int. Conf. Image Vision Computing New Zealand*, IEEE online, 2017.
- [17] A. Dhiman, H.-J. Chien, and R. Klette, “A multi-frame stereo vision-based road profiling technique for distress analysis,” in *Proc. Int. Symp. Pervasive Systems Algorithms Networks*, IEEE online, 2018.
- [18] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road damage detection using deep neural networks with images captured through a smartphone,” *arXiv preprint*, arXiv:1801.09454, 2018.
- [19] A. Zhang, K. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J. Q. Li, and C. Chen, “Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network,” *Computer-Aided Civil Infrastructure Engineering*, 32(10): 805–819, 2017.
- [20] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy “Speed/accuracy trade-offs for modern convolutional object detectors”, in *Proc. Computer Vision Pattern Recognition*, 2017
- [21] A. G. Howard, M.Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient convolutional neural networks for mobile vision applications”, *arXiv preprint*, arXiv:1704.04861, 2017.
- [22] B. Akarsu, M. Karaköse, K. Parlak, E. Akin, and A. Sarimaden, “A fast and adaptive road defect detection approach using computer vision with real time implementation”, *Int. J. Applied Mathematics Electronics Computers*, 4: 290–295, 2016.
- [23] E. Zalama, J. G. -G Bermejo, R. Medina, and J. Llamas “Road crack detection using visual features extracted by Gabor filters”, *Computer-Aided Civil Infrastructure Engineering*, 29(5), 342–358, 2014.
- [24] P. -J. Chun, K. Hashimoto, N. Kataoka, N. Kuramoto, and M. Ohga, “Asphalt pavement crack detection using image processing and naive Bayes based machine learning approach”, *J. Japan Society Civil Engineers, Ser. EI (Pavement Engineering)*, 70(3), 2015.
- [25] A. Kulkarni, N. Mhalgi, S. Gurnani, and N. Giri “Pothole detection system using machine learning on android”, *Int. J. Emerging Technology Advanced Engineering*, 4(7): 360–364, 2014.
- [26] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and H. Balakrishnan, “The pothole patrol: Using a mobile sensor network for road surface monitoring”, in *Proc. Int. Conf. MSAS*, 29–39, 2008.
- [27] U. Bhatt, S. Mani, E. Xi, and J. Z. Kolter “Intelligent pothole detection and road condition assessment”, *arXiv preprint*, arXiv:1710.02595v2, 2017.
- [28] Y. -W. Hsu, J. W. Perng, and Z. -H. Wu, “Design and implementation of an intelligent road detection system with multisensor integration”, in *Proc. Int. Conf. Machine Learning Cybernetics*, 219–225, 2016.
- [29] M. Staniek, “Neural networks in stereo vision evaluation of road pavement condition”, in *Proc. Int. Symp. Non-Destructive Testing Civil Engineering* 15–17, 2015.
- [30] G. Lin, A. Milan, C. Shen, and I. D. Reid, “RefineNet: Multi-path refinement networks for high-resolution semantic segmentation”, *CoRR*, 1611.06612, 2016.
- [31] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network”, *CoRR*, 1612.01105, 2016.
- [32] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, “Large kernel matters - Improve semantic segmentation by global convolutional network”, *CoRR*, 1703.02719, 2017.
- [33] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation”, in *Proc. Computer Vision Pattern Recognition*, pp. 3431–3440, 2015.
- [34] V. Badrinarayanan, A. Kendall, and R. Cipolla “SegNet: A deep convolutional encoder-decoder architecture for image segmentation”, *CoRR*, 1511.00561: 2015.
- [35] L. -C Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs”, *IEEE Trans. Pattern Analysis Machine Intelligence*, 40(4): 834–848, 2018.
- [36] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft COCO: Common objects in context”, in *Proc. European Conf. Computer Vision*, 2014.
- [37] J. Deng, W. Dong, R. Socher, L. -J. Li, K. Li, and F.-F. Li, “Imagenet: A large-scale hierarchical image database”, in *Proc. Computer Vision Pattern Recognition*, pp. 248–255, 2009
- [38] H. Kaiming, G. Georgia, D. Piotr, and G. Ross, “Mask R-CNN”, *CoRR*, 1703.06870, 2017.
- [39] R. Shaoqing, H. Kaiming, G. Ross, and S. Jian, “Faster R-CNN: Towards real-time object detection with region proposal networks”, *CoRR*, 1506.01497, 2015.
- [40] R. Guzmán, J. -B. Hayet, and R. Klette, “Towards ubiquitous autonomous driving: The CCSAD dataset,” in *Proc. Int. Conf. Computer Analysis Images Patterns*, 2015, pp. 582–593.
- [41] A. Brner, D. Baumbach, M. Buder, A. Choinowski, I. Ernst, E. Funk, D. Griebach, A. Schischmanow, J. Wohlfeil, and S. Zuev, “IPS - A vision-aided navigation system” in *Proc. Advanced Optical Technologies*, 6(2):121–129, 2017.
- [42] D. Griebach, D. Baumbach, and S. Zuev, “Stereo-vision-aided inertial navigation for unknown indoor and outdoor environments” in *Proc. Indoor Positioning Indoor Navigation IEEE Xplore*, 2014.
- [43] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset.” *Int. J. Robotics Research*, 2013.
- [44] S. Nienaber, M.J. Booysen, and R.S. Kroon, “Detecting potholes using simple image processing techniques and real-world footage,” in *Proc. Southern African Transport Conference*, July 2015.
- [45] www.tensorflow.org/api_docs/python/tf/nn/sigmoid_cross_entropy_with_logits