A First-Person Database for Detecting Barriers for Pedestrians

Zenonas Theodosiou[1], Harris Partaourides[1], Tolga Atun[1], Simoni Panayi[1] and Andreas Lanitis[1,2]
[1]*Research Centre on Interactive Media Smart Systems and Emerging Technologies, Nicosia, Cyprus*
[2]*Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus*
{*z.theodosiou, h.partaourides*}*@rise.org.cy, tolga.atun@yahoo.com, simoni.panayi@gmail.com, a.lanitis@rise.org.cy*

Keywords: Pedestrians, Safety, Visual Lifelogging, Egocentric Vision, First-person View, Dataset

Abstract: Egocentric vision, which relates to the continuous interpretation of images captured by wearable cameras, is increasingly being utilized in several applications to enhance the quality of citizens' life, especially for those with visual or motion impairments. The development of sophisticated egocentric computer vision techniques requires automatic analysis of large databases of first-person point of view visual data collected through wearable devices. In this paper, we present our initial findings regarding the use of wearable cameras for enhancing the pedestrians' safety while walking in city sidewalks. For this purpose, we create a first-person database that entails annotations on common barriers that may put pedestrians in danger. Furthermore, we derive a framework for collecting visual lifelogging data and define 24 different categories of sidewalk barriers. Our dataset consists of 1796 annotated images covering 1969 instances of barriers. The analysis of the dataset by means of object classification algorithms, depict encouraging results for further study.

1 INTRODUCTION

Walking is the most basic and highly popular form of transportation and it is evident today that it is getting more dangerous. According to the World Health Organization (WHO, 2019), 270K pedestrians per year lose their lives around the world. Contemporary cities have to deal with the various problems caused by the increasing amount of technical barriers and damages that occur on the footpaths which endanger the lives of pedestrians (Sas-Bojarska and Rembeza, 2016). Guaranteeing everyday urban safety has always been a central theme for local authorities, addressing remarkable human, social, and economic aspects. The need for clear paths in urban sidewalks, free of barriers, continuous, and in a well-maintained condition is of great importance. Thus, the automatic detection of obstructions and damages can have a positive impact on the sustainability and safety of citizens' communities. The pedestrian detection (Szarvas et al., 2005) is one of the main research areas as an ultimate aim to develop efficient systems to eliminate deaths in traffic accidents. The safety in roads has attracted a large interest in the last years and a number of studies have been presented for both pedestrians (Nesoff et al., 2018; Wang et al., 2012) and drivers (Timmermans et al., 2019). A study on pothole detection was presented by Prathiba el al. (Prathiba et al., 2015) for the identification of different types of cracks on road pavements. Wang et al. (Wang et al., 2012) developed the WalkSafe, a smartphone application for vehicles recognition to help pedestrians cross safely roads. Jain et al.(Jain and Gruteser, 2017) presented an approach based on smartphone images for recognizing the texture of the surfaces in pedestrians' routes to be used for safety purposes. A mobile application which uses phone sensors was also presented to enhance the safety of the distracted pedestrians (Tung and Shin, 2018). On the other hand, Maeda et al (Maeda et al., 2018) proposed an approach for the detection of several road damages in smartphones using convolution networks.

In the realm of safety, the practicality and efficient use of wearable cameras can effectively help increase the safety of pedestrians. The continuous visual data acquisition can lead to the real-time detection of obstructions, warning the wearers of the potential dangers and alerting the authorities for taking maintenance or corrective actions for ensuring the elimination of dangerous spots for pedestrians. Due to the broad use of deep learning algorithms in analysing visual lifelogging data, the existence of large annotated datasets are more essential than ever before. Although there are available datasets created by wearable or smartphone cameras refer to road safety, none of them is dedicated specifically for the safety of pedestrians in sidewalks. This work outlines a first-person database, which can be used for the development of techniques for automatic detection of barriers and other damages that pose safety issues to the pedestrians. In addition, we present initial results on the performance of the dataset in a classification scheme using a well-known deep Convolutional Neural Network (CNNs) as a baseline and elaborate on the promising outcomes.

The structure of the rest of the paper is as follows: Section 2 presents the current state of the art focusing on visual lifelogging, datasets and image interpretation techniques. Section 3 is dedicated on the created dataset explaining the method we have followed to collect and annotate the lifelogging data. The methodology we used to evaluate the performance along with the initial results are shown in Section 4. Finally, conclusions and future work are drawn in Section 5.

2 BACKGROUND

2.1 Visual lifelogging

Visual lifelogging has been broadly used nowadays due to the advances in wearable and sensing technologies (Theodosiou and Lanitis, 2019). The small size and light weight of wearable cameras in addition to the broad use of smartphone devices allow the 24/7 uninterrupted acquisition of the carrier's daily life (Bolaños et al., 2015). The interpretation of lifelogs can lead to useful results which can be exploited to enhance health, protection, security, and to analyze lifestyle and daily habits. The automatic analysis of visual lifelogging data combining both computer vision and machine learning techniques, is known as Egocentric or First-person camera Vision.

Several image analysis methodologies have been proposed dedicated on visual lifelogging for both indoor and outdoor applications. In (Bolaños and Radeva, 2016) a two-step method is presented for food detection and recognition in lifelogging images. A recognition of personal locations in daily activities of the wearer is studied in (Furnari et al., 2017) while social interactions and lifestyle patterns are analyzed in (Bano et al., 2018) and (Herruzo et al., 2017) respectively. Visual lifelogging has also been used in ambient assisted living applications (Climent-Pérez et al., 2020) such as fall detection, monitoring, etc.

Wearable cameras can play the role of a digital memory helping people with memory problems to improve the quality of their daily life (Oliveira-Barra et al., 2019). Thus, the recording, storage and retrieval of inaccessible memories through visual lifelogging has been extensively studied the last years (Silva et al., 2016). In addition, navigation and safety applications have also been developed using wearable cameras, enhancing the quality of living for several groups of people. Jiang et al. (Jiang et al., 2019) proposed a vision sensors based system for assisting people with vision impairments while in (Maeda et al., 2018) a smartphone application of road damages' detection is presented for drivers protection. A variation of the latter system can be adopted to detect and recognize the barriers in sidewalks for pedestrians' safety.

2.2 Egocentric Databases

The creation of annotated databases is a crucial step for the development of new egocentric vision techniques. The current trends in automatic analysis and interpretation of lifelogs collected from wearable camera devices relate with deep learning methods that perform better when are trained and tested on qualitative and quantitative data. Mayor and Murray (Mayol and Murray, 2005) created the first egocentric vision dataset with 600 frames captured by a wearable camera installed on the left shoulder of the wearer. The dataset was used for training systems to recognize hand actions. The abundance and availability of wearable cameras and smartphones have led to the creation of several first person datasets the last years (Bolaños et al., 2015) including datasets for object recognition (Bullock et al., 2015), activity recognition (Gurrin et al., 2016), social interaction analysis (Bano et al., 2018), etc. However, not all datasets are publicly available to the academic community.

Epic-kitchens (Damen et al., 2018) is the largest publicly available egocentric dataset with a total of 55 hours of recordings collected by a head-mounted high-definition camera. The dataset consists of 11.5 million frames covering daily activities in 32 kitchens. KrishnaCam (Singh et al., 2016) is another example of a large available egocentric dataset dedicated on daily outdoor activities captured using Google Glass. It consists of 460 unique video recordings, each ranging in length from a few minutes to about a half hour of video, making up 7.6 million frames in total.

Concerning the road safety, several datasets have been created for road and pavement cracks detection (Gopalakrishnan, 2018). The dataset presented by Zhang et al. (Zhang et al., 2016) was created using a low-cost smartphone. The dataset consists of 500 images and was used to detect cracks on pavements with the aid of deep convolutional networks. A large-scale dataset focused on road damages was created by Maeda et al. (Maeda et al., 2018). This dataset was made up of 9,053 road damage images captured with a smartphone installed on a car. A large dataset was created by merging recordings captured by 8 pedestrians while walking in 4 large cities (Jain and Gruteser, 2017). The dataset used to detect different surfaces in daily walking paths. Although the dataset has many possibilities to be used in several applications, it's not annotated according to the relevant barriers and its not publicly available yet.

2.3 Image Interpretation

The interpretation of visual lifelogs requires flexible algorithms that can address their specific features such as, large number of objects, blurring, motion artifacts, lighting variations, etc. Due to the difficulties of the traditional algorithms to cope with these limitations, deep learning algorithms have been successfully used the last years to analyze the visual content of data collected through wearable devices.

CNNs have been established as the most prominent strain of neural networks within the field of computer vision due to their efficiency in capturing spatial dependencies in images. They have achieved great strides in fundamental tasks for image interpretation such as classification, localization and object detection. The radical advancements in CNNs has been possible by the abundance of large public image repositories in the likes of ImageNet, Pascal VOC and MS COCO (Russakovsky et al., 2015; Lin et al., 2014; Everingham et al., 2007) which serve as platforms for enhancing generations of architectures in a bid to achieve state-of-the-art performances (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014; He et al., 2016).

Since their conception (LeCun et al., 1998), CNNs have evolved into numerous different architectures with allow for an increased network depth and complexity but, at their core, their basic components have remained very similar. Even though ResNet is approximately 20 times deeper than AlexNet and 8 times deeper than VGGNet (Gu et al., 2018), all three architectures consist of three types of core layers: convolutional, pooling and fully connected layers.

In general, a CNN is considered as a hierarchical feature extractor where the convolutional layers are responsible for creating feature maps, the pooling layers are used for reducing the resolution of the feature maps and the fully connected layers perform the high-level feature learning, i.e. image interpretation tasks.

3 BARRIER DETECTION DATASET

Based on our background research, we observed that there is no available dataset to effectively tackle issues regarding the safety of pedestrians on city sidewalks. The detection and identification of barriers that may endanger pedestrians' lives are facing several challenges such as: the barrier's type (i.e. cracks, objects, etc.), the small differences between the harmless and dangerous barriers a pedestrian encounters in sidewalks, the fact that a harmless object can become dangerous in the case of pedestrians, and the fact that a barrier is considered dangerous when it is in the vicinity of the pedestrian. To this end, we design a methodology for collecting and annotating such data which then utilize to populate our dataset. In this section, we present in more detail the proposed methodology, the data acquisition and annotation process, followed by an overview of the created dataset.

3.1 Proposed Methodology

Our approach for populating the pedestrians' barriers dataset consists of three core tasks. Initially, a person walks around the city limits sidewalks collecting in frequent time intervals snapshots with the rear camera of a mid-range smartphone. The smartphone is placed on the chest of the wearer with a slight downwards angle to accommodate our target area which is near the pedestrian's feet. Our approach considers a mid-range smartphone which is used by the majority of pedestrians. This is followed, by a pre-processing step that blurs faces, license plates, and brands to solve privacy issues and comply with the EU General Data Protection Regulation (GDPR).

Finally, the annotation process consists of placing bounding boxes around the categories derived by our analysis of the common barriers existing in urban areas. Specifically, the annotator must consider the danger area around a pedestrian. This translates to a region of 2m around the pedestrian. Considering the downwards angle of the camera this is broadly defined in the lower half of the collected images. Additionally, the annotation process considers up to three (most imminent) barriers which are placed up to the middle of the image. The barriers are categorized in 24 classes spanning in 3 main categories and 7 barrier types covering a broad range of possible barriers on the city sidewalks that affect pedestrians' safety. See Table 1 for more details.

3.2 Dataset

During the first stage of the acquisition process, an individual collected 1796 first-person images by walking 12.3Km in the city center of Nicosia, Cyprus. The smartphone used was an iPhone 7 with snapshot resolution of 1512x2016 pixels . The collected images were manually annotated using the VGG Image Annotator tool (Dutta and Zisserman, 2019) into the 24 different classes. A total of 1969 bounding boxes were found. 472 of the bounding boxes are related to barriers of the Infrastructure, 909 are related to barriers of Physical Condition and the remaining 588 to the category indicating Temporary barriers. *Tree* was the most popular Barrier class with 360 bounding boxes while *Boulder*, *Chair*, *Table*, *Mail Box* and *Bench* were the less popular classes (1). Examples of the annotation on different types of barriers are depicted in Fig. 1.

Table 1: Barrier types in our dataset.

Category	**Barrier Type**	**Detail**	**Class**	**Instances**
Physical Condition	Damage	Crack	B00	28
		Hole/Pot-hole	B01	323
		Paver (bro-ken)	B02	27
	Layout	Narrow Pave-ment	B10	54
		No Pave-ment	B11	40
Infrastructure	Street Furniture	Bench	B20	2
		Light	B21	80
		Bin	B22	70
		Parking Meter	B23	53
		Plant Pot	B24	88
	Street Decor	Tree	B30	360
		Shrub	B31	52
		Parking Pre-ven-tion	B32	2
		Mail Box	B33	202
Temporary	Vehicles	4-wheels	B50	85
		2-wheels	B51	61
	Construction	Boulder	B61	8
		Safety Sign	B62	74
		Fence	B63	161
		Traffic Cone	B64	151
	Other	Litter	B70	16
		Chair	B71	6
		Table	B72	9
		Advert. Sign	B73	17

4 EXPERIMENTAL RESULTS

To evaluate our dataset, we deploy a typical CNN deep network using using Tensorflow (Chollet et al., 2015) and train in an end-to-end manner using stochastic gradient descent.

Specifically, we utilize a variant of the VGG-16 model architecture with 13 convolutional layers, 5 max pooling layers followed by two dense layers and a dropout with a ratio set to 0.5, initialized with the pretrained weights of ImageNet.

As part of the classification process, image regions containing barriers, as indicated in the annotation process, were cropped in order to isolate the objects as individual images. We then perform a simple classification on each individual object found within our dataset using the VGG network.

Before the images are passed through the network, we perform preprocessing steps to avoid overfitting during training. The preprocessing steps include subsampling to the fixed-size of 256x256 pixels, shuffling the training examples, normalize the images in the range [0, 1], varying the brightness of images as well as some image geometric transformations such as random rotations, width and height shifts, horizontal flips and image magnifications.

In order to balance our dataset, we conducted training and evaluation using a reduced version of our dataset where we removed any class which contained 30 examples or less. Our revised dataset included 15 out of the original 24 barrier classes which subsequently yielded a more balanced class representation. Additionally, we randomly divided our data in a ratio of 70% training data (1297 images) and 30% validation data (557 images).

For this experiment, our classifier was trained for 200 epochs with a batch size of 64. During training the model achieved an accuracy of 65% and a weighted average of 59% whereas during validation it reached 55% for both accuracy and weighted average. Table 2 depicts the confusion matrix on the training data. We enhance the matrix by including the precision and recall for each class.

Looking at the confusion matrix, we can observe that there is a general trend which suggests that as the number of examples within a class increases so do its performance metrics. This is because the more

(a) (b) (c) (d)

Figure 1: Sample images from the created dataset: (a) No Pavement, (b) Parking Meter, (c) 2-Wheels, (d) 4-Wheels.

Table 2: The table represents the confusion matrix produced when the baseline model is used make predictions on the validation set.

	B01	B10	B11	B21	B22	B23	B24	B30	B31	B33	B50	B51	B62	B63	B64	**Recall**
B01	105	0	0	0	0	0	0	33	0	10	1	2	3	24	52	0.46
B10	0	5	1	0	0	0	0	11	0	2	0	0	0	18	0	0.14
B11	0	0	18	0	0	0	0	2	0	0	8	0	0	0	0	0.64
B21	0	0	0	16	0	0	0	12	1	0	14	3	4	0	0	0.32
B22	2	0	0	1	20	0	0	12	0	1	8	1	2	0	0	0.43
B23	0	0	0	1	0	18	0	5	1	3	0	0	6	0	0	0.53
B24	0	0	4	1	0	1	16	19	0	9	1	11	4	0	0	0.24
B30	2	0	0	1	12	8	0	214	0	6	2	1	1	0	0	0.87
B31	0	0	0	3	0	0	0	12	12	1	0	1	5	0	0	0.35
B33	10	0	1	0	1	0	0	20	0	100	4	3	0	1	0	0.71
B50	0	0	3	0	0	0	0	8	0	0	45	3	0	0	0	0.76
B51	0	0	0	0	0	0	0	7	0	0	3	32	3	0	0	0.71
B62	1	0	3	1	0	0	0	13	0	0	0	5	26	0	2	0.51
B63	26	1	0	0	0	0	0	6	0	1	1	0	3	66	11	0.57
B64	55	0	0	0	0	0	0	4	0	0	0	0	0	3	52	0.46
Precision	0.52	0.83	0.60	0.67	0.61	0.67	1.00	0.57	0.86	0.75	0.52	0.52	0.46	0.59	0.44	

examples a classifier is given, the more robust it can become in identifying the features of a specific class. It can also be seen that for some classes, our baseline model has difficulty distinguishing them and consequently mislabels most of their examples. For example, class B64 has most of its images misclassified in B01 which suggests that either there are not enough B64 examples or that the two barriers have very similar features. Similarly, B63 has approximately half of its examples classified in B01.

5 CONCLUSIONS & FUTURE WORK

This work presents the preliminary results of our on going study on creating a new first-person dataset on pedestrians' barriers while walking on sidewalks. To the best of our knowledge, this is the first dataset dedicated on sidewalk's barriers. Currently the dataset consists of 1796 images including 3156 instances of barriers categorized into 24 different classes. The performance of the dataset was evaluated using convolutional networks for object classification. The VGG-16 architecture was used as a baseline classifier for 15 barriers objects.

Future work involves collecting more images especially for classes with small number of instances and repeating the annotation process by two different annotators, as well as, using segmentation masks to finalize the dataset. Additionally, more experiments will be conducted to evaluate the performance of the final dataset as a training set in deep learning schemes for both detection and recognition tasks. Last but not least, the future work includes the preparation of the dataset to become publicly available and its dissemination to the research community.

The ultimate aim of this work is to develop an accurate real time pedestrian barrier detection system that will be incorporated in an integrated smart city platform that aims to provide services and enhanced safety for pedestrians. Our belief is that this system can be adopted by Municipalities and can be used as an immediate report system that can communicate any safety concerns for pedestrians for them to be repaired, thus working towards achieving the goals set by the World Health Organization on reducing the deaths of pedestrians.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 739578 complemented by the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

REFERENCES

Bano, S., Suveges, T., Zhang, J., and McKenna, S. (2018). Multimodal egocentric analysis of focused interactions. *IEEE Access*, 6:1–13.

Bolaños, M., Dimiccoli, M., and Radeva, P. (2015). Toward storytelling from visual lifelogging: An overview. *IEEE Transactions on Human-Machine Systems*, 47:77–90.

Bolaños, M. and Radeva, P. (2016). Simultaneous food localization and recognition. In *23rd International Conference on Pattern Recognition (ICPR)*, pages 3140–3145.

Bullock, I. M., Feix, T., and Dollar, A. M. (2015). The yale human grasping dataset: Grasp, object, and task data in household and machine shop environments. *The International Journal of Robotics Research*, 34(3):251–255.

Chollet, F. et al. (2015). Keras. https://keras.io.

Climent-Pérez, P., Spinsante, S., Michailidis, A., and Flórez-Revuelta, F. (2020). A review on video-based active and assisted living technologies for automated lifelogging. *Expert Systems with Applications*, 139:112847.

Damen, D., Doughty, H., Farinella, G. M., Fidler, S., Furnari, A., Kazakos, E., Moltisanti, D., Munro, J., Perrett, T., Price, W., and Wray, M. (2018). Scaling egocentric vision: The epic-kitchens dataset. In *European Conference on Computer Vision (ECCV)*.

Dutta, A. and Zisserman, A. (2019). The VIA annotation software for images, audio and video. *arXiv preprint arXiv:1904.10699*.

Everingham, M., Van Gool, L., Williams, C. K., Winn, J., and Zisserman, A. (2007). The pascal visual object classes challenge 2007 (voc2007) results.

Furnari, A., Farinella, G. M., and Battiato, S. (2017). Recognizing personal locations from egocentric videos. *IEEE Transactions on Human-Machine Systems*, 47(1):6–18.

Gopalakrishnan, K. (2018). Deep learning in data-driven pavement image analysis and automated distress detection: A review. *Data*, 3(3):28.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., Cai, J., et al. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377.

Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., and Albatal, R. (2016). Overview of ntcir-12 lifelog task.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Herruzo, P., Portell, L., Soto, A., and Remeseiro, B. (2017). Analyzing first-person stories based on socializing, eating and sedentary patterns. In Battiato, S., Farinella, G. M., Leo, M., and Gallo, G., editors, *New Trends in Image Analysis and Processing – ICIAP 2017*.

Jain, S. and Gruteser, M. (2017). Recognizing textures with mobile cameras for pedestrian safety applications. *IEEE Transactions on Mobile Computing*, PP.

Jiang, B., Yang, J., Lv, Z., and Song, H. (2019). Wearable vision assistance system based on binocular sensors for visually impaired users. *IEEE Internet of Things Journal*, 6(2):1375–1383.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., and Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images: Road damage detection and classification. *Computer-Aided Civil and Infrastructure Engineering*, 33.

Mayol, W. W. and Murray, D. W. (2005). Wearable hand activity recognition for event summarization. In *Ninth IEEE International Symposium on Wearable Computers (ISWC'05)*, pages 122–129.

Nesoff, E., Porter, K., Bailey, M., and Gielen, A. (2018). Knowledge and beliefs about pedestrian safety in an urban community: Implications for promoting safe walking. *Journal of Community Health*, 44.

Oliveira-Barra, G., Bolaños, M., Talavera, E., Gelonch, O., Garolera, M., and Radeva, P. (2019). *Lifelog retrieval for memory stimulation of people with memory impairment*, pages 135–158.

Prathiba, T., Thamaraiselvi, M., Mohanasundari, M., and Veerelakshmi, R. (2015). Pothole detection in road using image processing. *International Journal of Management, Information Technology and Engineering*, 3(4):13–20.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252.

Sas-Bojarska, A. and Rembeza, M. (2016). Planning the city against barriers. enhancing the role of public spaces. *Procedia Engineering*, 161:1556 – 1562.

Silva, A. R., Pinho, M., Macedo, L., and Moulin, C. (2016). A critical review of the effects of wearable cameras on memory. *Neuropsychological rehabilitation*, 28:1–25.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Singh, K. K., Fatahalian, K., and Efros, A. A. (2016). Krishnacam: Using a longitudinal, single-person, egocentric dataset for scene understanding tasks. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9.

Szarvas, M., Yoshizawa, A., Yamamoto, M., and Ogata, J. (2005). Pedestrian detection with convolutional neural networks. pages 224 – 229.

Theodosiou, Z. and Lanitis, A. (2019). Visual lifelogs retrieval: State of the art and future challenges. In *2019 14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*.

Timmermans, C., Alhajyaseen, W., Reinolsmann, N., Nakamura, H., and Suzuki, K. (2019). Traffic safety culture of professional drivers in the state of qatar. *IATSS Research*.

Tung, Y. and Shin, K. G. (2018). Use of phone sensors to enhance distracted pedestrians' safety. *IEEE Transactions on Mobile Computing*, 17(6):1469–1482.

Wang, T., Cardone, G., Corradi, A., Torresani, L., and Campbell, A. T. (2012). Walksafe: A pedestrian safety app for mobile phone users who walk and talk while crossing roads. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems & Applications*, HotMobile '12, pages 5:1–5:6.

WHO (2019). Pedestrian safety: a road safety manual for decision-makers and practitioners. https://www.who.int/roadsafety/projects/manuals/pedestrian/en/. Accessed: 2019-07-30.

Zhang, L., Yang, F., Zhang, Y., and Zhu, Y. (2016). Road crack detection using deep convolutional neural network.