INTELLIGENT ROAD SURFACE DEEP EMBEDDED CLASSIFIER FOR AN EFFICIENT PHYSIO-BASED CAR DRIVER ASSISTANCE

F. Rundo[1], *R. Leotta*[2], *V. Piuri*[3], *A. Genovese*[3], *F. Scotti*[3],*S. Battiato*[2]

[1]STMicroelectronics, ADG Central R&D
francesco.rundo@st.com
[2]University of Catania, IPLAB - Computer Science Department
leotta.rob@gmail.com, battiato@dmi.unict.it
[3]University of Milan, Computer Science Department
{*vincenzo.piuri, angelo.genovese, fabio.scotti*}@*unimi.it*

ABSTRACT

Car driving safety represents one of the major targets of the ADAS (Advanced Driver Assistance Systems) technologies deeply investigated by the scientific community and car makers. From intelligent suspension control systems to adaptive braking systems, the ADAS solutions allows to significantly improve both driving comfort and safety. The aim of this contribution is to propose a driving safety assessment system based on deep networks equipped with self-attention Criss-Cross mechanism to classify the driving road surface combined with a physio-based drowsiness monitoring of the driver. The retrieved driving safety assessment performance confirmed the effectiveness of the proposed pipeline.

Index Terms— ADAS, Automotive, Deep Learning, Road Classification, Intelligent Suspension

1. INTRODUCTION

The ADAS technologies are able to accomplish several tasks to assist the vehicle's driver leveraging different level of automation: from car driving assistance to fully autonomous driving or In-vehicle-Infotainment-Systems (IVIS) [1]. The recent ADAS technologies include automotive embedded systems suitable to provide ad-hoc warnings and alerts to the driver such as the Intelligent Speed Adaptation, collision warning systems or car driver drowsiness monitor [2]. Moreover, recent ADAS solutions combined visual information inside and outside the car with physiological assessment of the driver [2, 3]. In this context, the authors propose an innovative fully automated ADAS application which combines an efficient physiological car driver's drowsiness monitor driven by adaptive road surface risk assessment. The use of self-attention layers with temporal convolutional deep dilated architectures makes the proposed pipeline robust and efficient in monitoring driving risk. About road segmentation and classification, several approaches have been proposed. In [4] the authors described the development of a nice performer automated algorithms for extracting road features from Mobile Laser Scanning point cloud data. In [5] the authors proposed an interesting strategy to identify cracks on images captured during road pavement surveys. It adopted an efficient segmentation procedure, after appropriate image smoothing, followed by ad-hoc binary classification. Deep learning based solutions both supervised and unsupervised have been implemented for addressing the issue of a robust road segmentations [6, 7, 8, 9]. About driver attention monitoring systems, the authors have been deeply investigated the topic providing several scientific contributions and surveys [10, 11, 12, 13, 14, 15]. Several further approaches confirmed that physiological signal, especially the Photoplethysmography (PPG), can be efficiently used to monitor the car driver's attention level [15, 16, 17]. The proposed pipeline will be described in detail in the next paragraphs.

2. THE ROAD SURFACE SEGMENTATION AND CLASSIFICATION

The first sub-system of the proposed pipeline embeds a road segmentation and classification algorithm. In Fig. 1 is reported the scheme of the implemented approach. As schematized in Fig. 1, a Mask-R-CNN embedding a DenseNet-201 as feature generator backbone is proposed [18]. Mask-R-CNN is widely used in the automotive field [18]. The advantage of this architecture is that it provides a pixel-based segmentation of the driving frames as well as the generation of a bounding-box that characterizes the Region of Interest (ROI) on which to perform post-processing. The segmented road (ROI) will be fed as input of the enhanced downstream ResNet-101 in which we have embedded a Recurrent Criss-Cross Attention (RCCA) layer. The attention mechanism based on Criss-Cross processing was firstly proposed in [19] showing very promising performance in several tasks including semantic segmentation. More in detail, for each source image/feature

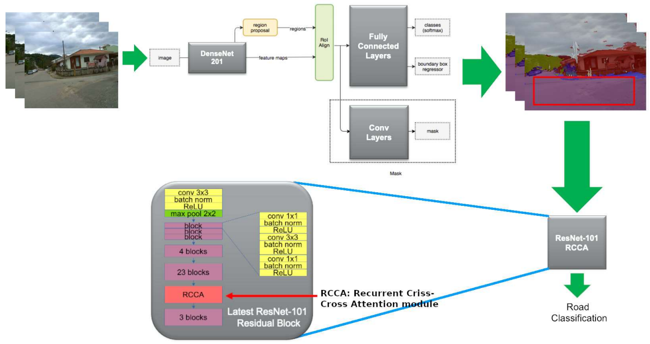

Fig. 1. The proposed Road Surface Classifier: Mask-R-CNN with a Recurrent Criss-Cross Attention (RCCA) enhanced ResNet-101

pixel, an innovative Criss-Cross attention module computes the contextual information of all the correlated pixels on its Criss-Cross path. This attention pre-processing combined with further recurrent operations allow the Criss-Cross method to leverage the full-image dependencies during the learning session of the deep network [19]. Lets formalize the attention processing embedded in the Criss-Cross layer we have implemented. Given a local feature map $H \in R^{C \times W \times H}$ where C is the original number of channels while $W \times H$ represents the spatial size of the generated feature map trough a Deep Convolutional Network. The Criss-Cross layer applies two preliminary 1×1 convolutional layers on H in order to generate two feature maps F_1 and F_2, which belong to $R^{C' \times W \times H}$ and in which C' represents the reduced number of channels due to dimension reduction with respect to original (C). Lets define an *Affinity* function suitable to generate the Attention-Map $A_M \in R^{(H+W-1) \times (W \times H)}$. The Affinity operation is so defined. For each position u in the spatial dimension of F_1, we extract a vector $F_{1,u} \in R^{C'}$. Similarly, we define the set $\Omega_u \in R^{(H+W-1) \times C'}$ by extracting feature vectors from F_2 at the same position u. So that, $\Omega_{i,u} \in R^{C'}$ is the i-th element of Ω_u. Taking into account the above operations, we can define the introduced *Affinity* operation as follows:

$$\delta^A_{i,u} = F_{1,u} \Omega^T_{i,u} \tag{1}$$

where $\delta^A_{i,u} \in D$ is the affinity potential i.e. the degree of correlation between features $F_{1,u}$ and $\Omega_{i,u}$ for each $i = [1, ..., H + W - 1]$, and $D \in R^{(H+W-1) \times (W \times H)}$. Then, we apply a softmax layer on D over the channel dimension to calculate the attention map A_M. Finally, another convolutional layer with 1×1 kernel will be applied on the feature map H to generate the re-mapped feature $\vartheta \in R^{C \times W \times H}$ to be used for spatial adaptation. At each position u in the spatial dimension of ϑ, we can define a vector $\vartheta_u \in R^C$ and a set $\Phi_u \in R^{(H+W-1) \times C}$. The set Φ_u is a collection of feature vectors in ϑ having the same row or column with position u. At the end, the final contextual information will be obtained by an *Aggregation* operation defined as follows:

$$H'_u = \sum_{i=0}^{H+W-1} A^{i,u}_M \Phi_{i,u} + H_u \tag{2}$$

where H'_u is a feature vector in $H' \in R^{C \times W \times H}$ at position u while $A^{i,u}_M$ is a scalar value at channel i and position u in A_M. The so defined contextual information H'_u is then added to the given local feature H to augment the pixel-wise representation and aggregating context information according to the spatial attention map A_M. The Criss-Cross processing fails to process the connections there are among one pixel and its around. For this reason, a Recurrent Criss-Cross processing was embedded in the proposed pipeline (with $R = 2$ i.e. Criss-Cross operations can be unrolled into 2 loops)[19]. In order to enhance the deep classifier, we have included a Criss-Cross layer in the latest residual block of the ResNet-101 as reported in Fig. 1. The proposed pipeline has been trained and tested in the RTK dataset [20] trying to discriminate four types of road: asphalt, paved, potholes and unpaved. The output of the Criss-Cross enhanced ResNet-101 (having a softmax as latest layer) is a binary mask of four bits [asphalt, paved, potholes, unpaved]. The bits set to 1 confirm that the segmented input frames contains this kind of road surface. The performance results will be showed in the next sections.

3. THE PPG BASED CAR DRIVER DROWSINESS MONITORING SYSTEM

As introduced, the second block of the proposed ADAS pipeline is the physio-based car driver drowsiness monitoring system. Specifically, we proposed a car-driver attention level monitoring based on the usage of the driver's Photoplethysmographic (PPG) signal. Through a deep PPG signal features analysis [3] a non-invasive blood volume dynamic assessment can be retrieved. More in detail, a common PPG waveform consist of a pulsatile physiological signal ('AC') that embedding blood volume information overlapped with slowly varying component ('DC') that represents information correlated to the skin tissues (where the PPG sensor is placed), respiration and thermoregulation. With a device consisting in a light-emitter and a detector placed on the skin that measure the amount of light either transmitted or reflected we can detect the blood volume changes occuring with the heart pressure pulse. The correlation between blood volume changes and the Autonomic Nervous System that manage the alert levels of the subject and cardiac activity allows to consider the PPG an excellent indirect detector of the subject's level of attention [3, 10, 11]. In addition, the correct level of attention required for safe driving is computed and adjusted according to the driving context (speed, pavement conditions, adjacent vehicles, and so on) [14]. In this work, the Silicon Photomultiplier sensor [10, 11] was used as PPG detector.

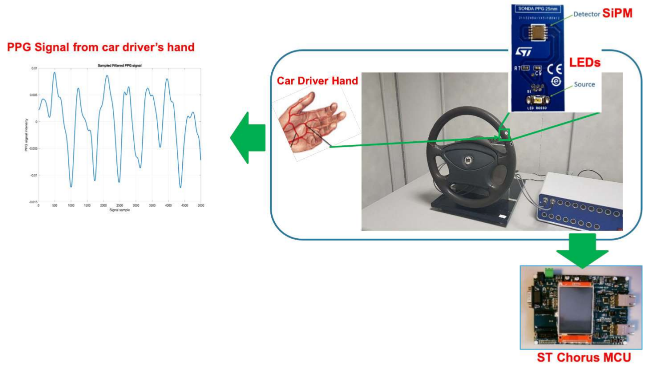

Fig. 2. The PPG sensing device embedded in the car steering.

More in detail, the suggested PPG probes consists in a large area of n-on-p Silicon Photomultipliers (SiPMs) fabricated at STMicroelectronics [10, 11]. The SiPMs array has a total area of $4.0 \times 4.5mm^2$ and 4871 square microcells with $60\,\mu m$ pitch, packaged in a surface mount housing (SMD) with about $5.1 \times 5.1mm^2$ total area [11]. Furthermore, on the SMD package was glued a Pixelteq dichroic bandpass filter by means the use of Loctite 352TM adhesive. The aforementioned bandpass filter was set with a pass band centered at about $540\,\mathrm{nm}$ with a Full Width at Half Maximum (FWHM) of $70\,\mathrm{nm}$ and an optical transmission higher than $90 - 95\%$ in the pass band range. As light emitter we have used the OSRAM LT M673 LEDs in SMD package that emits at $830\,\mathrm{nm}$ which is based on InGaN technology [11]. More in detail, the aforementioned LEDs devices have an area of $2.3 \times 1.5mm^2$, spectral bandwidth of $33\,\mathrm{nm}$, viewing angle of $120°$ and lower power emission (mW) in the standard operation range. The authors have designed a printed circuit board (PCB) in order to make the PPG probe easily to use. More implementation details can be found in [11]. In Fig. 2 we report an overall scheme of the proposed PPG sensing framework (SiPM + LEDs) embedded in the car steering. As reported in Fig. 2, the filtering and stabilization of the raw PPG signal collected from the car driver hand will be performed by the developed algorithms running as firmware in the ST Chorus MCU [11, 12, 13, 14, 15, 16]. After that, the hyper-filtering approach we have implemented and patented [11], [16] will be applied to the collected stabilized PPG data in order to retrieve an evaluation of the attention level of the driver from which the PPG signal is sampled. We have configured the hyper-filtering approach for the application herein described. Specifically, the proposed hyper-filtering system has been inspired by the widely-accepted idea of hyper-spectral processing used in 2D imaging [16]. Hyper-spectral imaging gather visual information through the whole electromagnetic spectrum, in order to retrieve the so called "frequency spectrum of each pixel" [15]. Thus, using the same method, the authors considered the information set retrieved from such "hyper-filtered" signals i.e. the set of signals obtained by different frequency filtering of the source time-serie (PPG in our use-case). With the proposed hyper-filtering approach we are able to collect valuable information about the frequency spectrum of the car driver's PPG signal and then about the correlated driver's attention level (Drowsiness monitoring). More in detail, we have divided the valuable PPG frequency range $0.5\,\mathrm{Hz}$–$10\,\mathrm{Hz}$ in several sub-ranges in which we have applied the Butterworth pass-band filter (high-pass and low-pass filters) as described in [16, 15]. Thus, we have configured two layers of hyper-filtering systems which are able to modulate the frequencies in the low-pass application, meanwhile preserving the cut-off frequency of the high-pass filter (Hyper low-pass filtering layer) and vice-versa (Hyper high-pass filtering layer). The applied hyper-filtering frequency setup is reported in Table 1 and Table 2. We proposed the usage of Butterworth filters in both hyper-filtering setup since they do not create modulations or distortions in the bandwidth [14, 15, 16]. We retrieved the frequency values reported in Table 1 and Table 2 through a Reinforcement Learning algorithm with a reward function correlated to the car driver drowsiness classification accuracy [15, 16]. Once the hyper-filtering setup has been assessed, the collected car driver PPG raw signal will be processed accordingly. Specifically, from the collected source PPG driver signal, a subset of hyper-filtered signals will be generated through the frequency setup as per Table 1 and Table 2. Lets define $W_i^{PPG}(t,k)$ the single segmented waveform of the i-th hyper-filtered PPG time-serie. For each sample $s^i(t_k)$ of the segmented PPG waveform $W_i^{PPG}(t,k)$, we will compute a signal-pattern representing the dynamic of the sample $s^i(t_k)$ for each i-th $W_i^{PPG}(t,k)$ waveforms. Consequently, we collect a large dataset of hyper-filtered signal patterns [14, 15, 16]. As soon as the driver put the hand over the PPG sensing probe embedded on the steering wheel, the hyper-filtering pipeline starts to work generating the signal-patterns to be fed as input to the Deep Learning block as detailed in Fig. 3. Specifically, the designed classifier is a Deep 1D Temporal Dilated Convolutional Neural Network (1D-TCNN) with residual blocks [15]. The temporal convolutional network is mainly characterized by a causal convolution layer [15]. The designed 1D-TCNN is composed as follows: 25 residual blocks with 3×3 kernel filters, where such of them contains dilated

F	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**
HP	0.5	/	/	/	/	/	/	/	/	/	/
LP	0.0	1.1	3.2	3.5	3.8	3.9	4.0	4.1	5.0	5.1	6.3

Table 1. Hyper Low-pass filtering setup (in Hz).

F	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**
HP	0.1	1.1	2.3	2.4	3.2	3.5	4	4.2	5	5.3	6.4
LP	7.0	/	/	/	/	/	/	/	/	/	/

Table 2. Hyper High-pass filtering setup (in Hz).

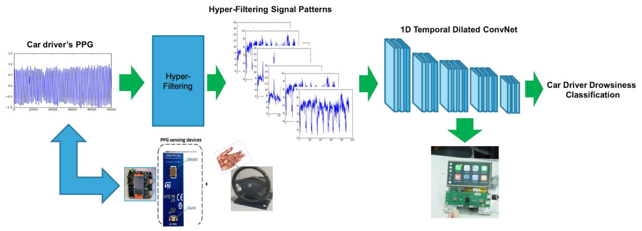

Fig. 3. The Physio-based Car Driver Drowsiness Monitoring System

convolution operations in which the dilation factor size starts from 2 and increase (power of 2) till to 16, normalization, ReLU activation, spatial dropout layers and a downstream softmax layer. The so designed 1D-TCNN is able to classify the input hyper-filtered PPG patterns coming from a drowsy or wakeful driver ($(0.0-0.5)$, $(0.51-1.0)$ respectively). As reported in Fig. 3, the designed 1D TCNN is running over the STA1295A Accordo5 MCU [15, 16].

4. EXPERIMENTAL RESULTS AND CONCLUSION

We tested the implemented pipeline, validating the single subsystems and then arranging a composite scenario of a road surface-driven risk assessment (driving safety monitoring system). Specifically, we considered the following risk assessment: asphalt (low driving risk), paved (medium driving risk due to certain braking problems) unpaved / potholes (high driving risk). More in detail, if high or medium risk level is assessed by the road classification system (Mask-R-CNN with Criss-Cross ResNet-101 downstream) the driving safety monitoring system will check that 1D-TCNN confirms a corresponding "wakeful" attention classification. Otherwise, acoustic alert-signal will be emitted by the Audio underlying System (STA1295A MCU) in order to alert the drowsy driver. If the driver's PPG signal is not available for some reasons (for instance: the driver does not have his/her hand placed over the PPG sensing devices in the steering wheel), the authors have developed a system for ad-hoc visual reconstruction of the PPG signal by means of an innovative motion magnification technique applied to specific driver's facial landmarks [12]. Now, follows more details about the performance of the proposed sub-systems. About the proposed road surface classification deep pipeline, we validated and compared our pipeline using the RTK dataset and related algorithms [20, 9]. We arranged the dataset into 80% for training and validation while the remaining 20% for testing. The following Table 3 shows some performance benchmarks. Regarding the driver's physio-based drowsiness assessment, we have tested the suggested pipeline by gathering various PPG measurements of several subjects in different scenarios (Drowsy vs Wakeful drivers) under clinical study covered by the *Ethical Committee CT1 authorization.113/2018/PO*.

Method	Road Surface Classification Performance		
	Low Risk (Asphalt)	**Medium Risk (Paved)**	**High risk (unpaved / potholes)**
Proposed	**93%**	92%	**97% / 97%**
Proposed w/o Criss-Cross	92%	89%	89% / 92%
Proposed w/o ResNet-101	88%	88%	84% / 82%
[20, 9]	92%	**94%**	94% / **97%**

Table 3. Road Surface Classification Performance.

We collected data from 70 patients with different features such as ages, gender, etc. Furthermore, simultaneously with the PPG signals we also acquired EEG signal to be able to verify the attention level (alpha and beta waves) [13]. We have sampled the PPG signals of the subjects by means of the hardware setup detailed in this contribution with a sampling frequency equal to $1\,\mathrm{kHz}$. We gathered 5 minutes of PPG signals for both condition (Drowsiness and Wakefulness). The so collected PPG time-series, have been organized as follow: 70% was used for the training and validation phases while the remaining 30% was used for testing. For the training phase of the 1D-TCNN we set an initial learning rate equal to 0.001 and dropout factor equal to 0.5. Furthermore, a classic SGD algorithm was used. The following Table 4 reports the performance obtained with the aforementioned pipeline compared to similar pipeline based on deep learning [16]. The collected performance results (related to the single subsystems) confirm that the overall proposed pipeline performs very well allowing an adaptive, robust and innovative fully automated assessment of the driving risk based on road surface classification. As confirmed by the results reported in Table 3, the use of Criss-Cross enhanced downstream ResNet-101 classifier allow to obtain significantly improvement in terms of classification performance. This research was supported by the National Funded Program 2014-2020 under grant agreement n. 1733, (ADAS + Project).

Method	Driver Drowsiness Monitoring	
	Drowsy Driver	**Wakeful Driver**
Proposed	**98.71%**	**99.03%**
[16]	96.50%	98.40%

Table 4. Car Driver Drowsiness Classification Performance.

5. REFERENCES

[1] Ryosuke Okuda, Yuki Kajiwara, and Kazuaki Terashima, "A survey of technical trend of adas and autonomous driving," in *Technical Papers of 2014 International Symposium on VLSI Design, Automation and Test*. IEEE, 2014, pp. 1–4.

[2] Chang Wang, Qinyu Sun, Yingshi Guo, Rui Fu, and

Wei Yuan, "Improving the user acceptability of advanced driver assistance systems based on different driving styles: A case study of lane change warning systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 10, pp. 4196–4208, 2019.

[3] Nicoleta Minoiu Enache, Saïd Mammar, Mariana Netto, and Benoit Lusetti, "Driver steering assistance for lane-departure avoidance based on hybrid automata and composite lyapunov function," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 28–39, 2009.

[4] Haiyan Guan, Jonathan Li, Yongtao Yu, Michael Chapman, and Cheng Wang, "Automated road information extraction from mobile laser scanning data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 194–205, 2014.

[5] Henrique Oliveira and Paulo Lobato Correia, "Road surface crack detection: improved segmentation with pixel-based refinement," in *2017 25th IEEE EUSIPCO Proceedings*. IEEE, 2017, pp. 2026–2030.

[6] 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," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.

[7] Jin Tian, Jiazheng Yuan, and Hongzhe Liu, "Road marking detection based on mask r-cnn instance segmentation model," in *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*. IEEE, 2020, pp. 246–249.

[8] Shashank Yadav, Suvam Patra, Chetan Arora, and Subhashis Banerjee, "Deep cnn with color lines model for unmarked road segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 585–589.

[9] Thiago Rateke and Aldo von Wangenheim, "Road surface detection and differentiation considering surface damages," *Autonomous Robots*, pp. 1–14, 2021.

[10] Vincenzo Vinciguerra, Emilio Ambra, et al., "Ppg/ecg multisite combo system based on sipm technology," in *Convegno Nazionale Sensori*. Springer, 2018, pp. 353–360.

[11] Francesco Rundo, Sabrina Conoci, Alessandro Ortis, and Sebastiano Battiato, "An advanced bio-inspired photoplethysmography (ppg) and ecg pattern recognition system for medical assessment," *Sensors*, vol. 18, no. 2, pp. 405, 2018.

[12] Francesca Trenta, Sabrina Conoci, Francesco Rundo, and Sebastiano Battiato, "Advanced motion-tracking system with multi-layers deep learning framework for innovative car-driver drowsiness monitoring," in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. IEEE, 2019, pp. 1–5.

[13] Francesco Rundo, Sergio Rinella, et al., "An innovative deep learning algorithm for drowsiness detection from eeg signal," *Computation*, vol. 7, no. 1, pp. 13, 2019.

[14] Francesco Rundo, Sabrina Conoci, Sebastiano Battiato, et al., "Innovative saliency based deep driving scene understanding system for automatic safety assessment in next-generation cars," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive*. IEEE, 2020, pp. 1–6.

[15] Francesco Rundo et al., "Advanced 1d temporal deep dilated convolutional embedded perceptual system for fast car-driver drowsiness monitoring," in *2020 AEIT International Conference of Electrical and Electronic Technologies for Automotive*. IEEE, 2020, pp. 1–6.

[16] Francesco Rundo, Concetto Spampinato, and Sabrina Conoci, "Ad-hoc shallow neural network to learn hyper filtered photoplethysmographic (ppg) signal for efficient car-driver drowsiness monitoring," *Electronics*, vol. 8, no. 8, pp. 890, 2019.

[17] Hyeonjeong Lee, Jaewon Lee, and Miyoung Shin, "Using wearable ecg/ppg sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," *Electronics*, vol. 8, no. 2, pp. 192, 2019.

[18] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[19] Zilong Huang, Xinggang Wang, Lichao Huang, et al., "Ccnet: Criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 603–612.

[20] Thiago Rateke, Karla Aparecida Justen, and Aldo von Wangenheim, "Road surface classification with images captured from low-cost camera-road traversing knowledge (rtk) dataset," *Revista de Informática Teórica e Aplicada*, vol. 26, no. 3, pp. 50–64, 2019.