Data and Metadata. 2023; 2:154
doi: 10.56294/dm2023154

ORIGINAL

Pyramid Scene Parsing Network for Driver Distraction Classification Pyramid Scene Parsing Network for Driver Distraction Classification

Red piramidal de análisis sintáctico de escenas para la clasificación de distracciones al volante

Abdelhak Khadraoui[1], Elmoukhtar Zemmouri[1]

[1]Moulay Ismail University, ENSAM, Meknes, Morocco.

Cite as: Khadraoui A, Zemmouri E. Pyramid Scene Parsing Network for Driver Distraction Classification. Data and Metadata. 2023;2:154. https://doi.org/10.56294/dm2023154

Submitted: 02-08-2023 **Revised:** 23-10-2023 **Accepted:** 29-12-2023 **Published:** 30-12-2023

Editor: Prof. Dr. Javier González Argote
Guest Editor: Yousef Farhaoui

Note: Paper presented at the International Conference on Artificial Intelligence and Smart Environments (ICAISE'2023).

ABSTRACT

In recent years, there has been a persistent increase in the number of road accidents worldwide. The US National Highway Traffic Safety Administration reports that distracted driving is responsible for approximately 45 percent of road accidents. In this study, we tackle the challenge of automating the detection and classification of driver distraction, along with the monitoring of risky driving behavior. Our proposed solution is based on the Pyramid Scene Parsing Network (PSPNet), which is a semantic segmentation model equipped with a pyramid parsing module. This module leverages global context information through context aggregation from different regions. We introduce a lightweight model for driver distraction classification, where the final predictions benefit from the combination of both local and global cues. For model training, we utilized the publicly available StateFarm Distracted Driver Detection Dataset. Additionally, we propose optimization techniques for classification to enhance the model's performance.

Keywords: Driver Distraction Detection; Pyramid Scene Parsing Network; Pspnet; Statefarm's Dataset; Convolutional Neural Networks.

RESUMEN

En los últimos años se ha producido un aumento persistente del número de accidentes de tráfico en todo el mundo. La Administración Nacional de Seguridad Vial de EE.UU. informa de que la conducción distraída es responsable de aproximadamente el 45 % de los accidentes de tráfico. En este estudio, abordamos el reto de automatizar la detección y clasificación de la distracción del conductor, junto con la monitorización del comportamiento de riesgo al volante. Nuestra solución propuesta se basa en la Pyramid Scene Parsing Network (PSPNet), que es un modelo de segmentación semántica equipado con un módulo de análisis sintáctico piramidal. Este módulo aprovecha la información de contexto global mediante la agregación de contextos de diferentes regiones. Presentamos un modelo ligero para la clasificación de distracciones al volante, en el que las predicciones finales se benefician de la combinación de indicios locales y globales. Para el entrenamiento del modelo, utilizamos el conjunto de datos de detección de conductores distraídos de StateFarm. Además, proponemos técnicas de optimización de la clasificación para mejorar el rendimiento del modelo.

Palabras clave: Detección de Distracción del Conductor; Red de Análisis de Escena Piramidal; Pspnet; Conjunto de Datos de Statefarm; Redes Neuronales Convolucionales.

INTRODUCTION

According to recent research conducted by the Moroccan National Agency for Road Safety, distracted driving was a contributing factor in 3,005 road fatalities and more than 84 585 injuries in Morocco in the year 2020. Regrettably, this issue seems to be worsening year after year.[1,2,3] Distracted driving, as defined by Strayer et al.[4] encompasses any activity that diverts a driver's attention away from the road, such as texting, eating, conversing with passengers, or adjusting the stereo. In light of this, the objective of our research is to develop and implement a model for the detection and classification of distracted driving in smart cars, leveraging semantic segmentation techniques[5] and convolutional neural networks (CNNs).[4] To identify and categorize driver distraction from visual cues, we explored various models, including convolutional neural networks (CNNs).[4] Building upon the state-of-the-art findings in this field, we devised a simplified model based on PSPNet, which yielded promising results.[6,7]

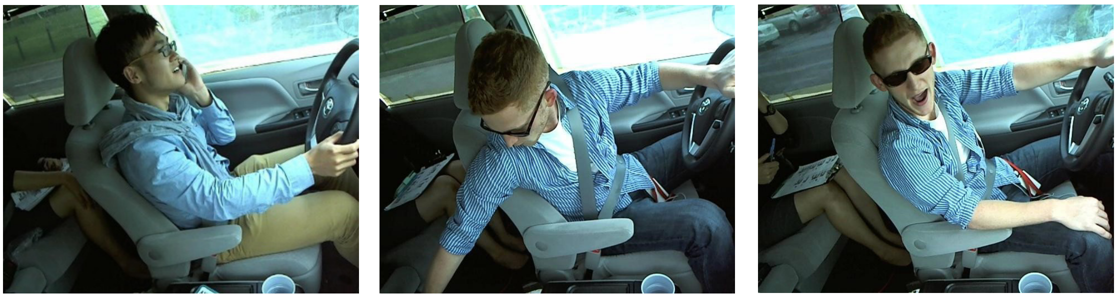
Related Work

Distracted driving can generally be classified into four distinct forms, as outlined by Strayer et al.[8] cognitive, visual, manual, and auditory distractions. When a driver becomes distracted, they divert their focus and actions away from driving-related tasks, engaging in non-driving activities. Some activities inherently involve multiple forms of distraction. For instance, using a cell phone for calls or texting can encompass all four forms of distractions.

Manual Distraction: This occurs when a driver's hands are taken off the steering wheel, impacting their ability to control the vehicle (as depicted in figure 1a). Common instances include eating, drinking, smoking, or retrieving items from a purse or wallet.

Visual Distraction: In this scenario, the driver's attention shifts to looking at a device instead of the road, which is one of the most prevalent distractions (as depicted in figure 1b). Examples encompass glancing at a GPS device, focusing on the entertainment center, observing a passenger.

Cognitive Distraction: Cognitive distractions emerge when the driver's focus is drawn away from driving by interpreting information from a device (as illustrated in Figure 1c). Common instances include listening to a podcast, engaging in conversations through hands-free devices, conversing with other passengers. Auditory Distraction because noise distracts the driver.[3]



a) Manual Distraction (b) Visual Distraction (c) Cognitive Distraction

Figure 1. StateFarm dataset[7] illustrating driver distraction

The table 1 shows some examples of distraction actions and their mapping to distraction types.

Table 1. Assignment to common distraction actions and distraction types

Activity	Location	Distractions
Using Phone	Within the car	Cognitive, Auditory, Manual, Visual
Eat, Drink	Within the car	Visual, Physical
Looking advertisement	Outside vehicle	Visual, Cognitive
Listening music	Within the car	Auditory, Cognitive

METHODS

Pyramid scene parsing network (PSPNet)

PSPNet, as introduced in Zhao et al.[9]'s work, utilizes a pretrained CNN[6] and employs the dilated network technique to extract feature maps from input images. The final feature map size is reduced to 1/8 of the input image dimensions. We leverage the pyramid pooling module to aggregate contextual information on top of this feature map. The pooling kernels cover various portions of the image, including the entire, half, and smaller

https://doi.org/10.56294/dm2023154

areas, thanks to our four-level pyramid structure. These aggregated features are then combined to form a holistic representation. In the final step, we fuse this combined data with the original feature map and apply a convolution layer to produce the ultimate prediction map.

The task of understanding visual scenes necessitates semantic image segmentation, as highlighted in Chen et al.[1]'s work. Semantic segmentation aims to classify each pixel in the input image, effectively performing pixel-level object segmentation.[2] This technique finds applications in diverse fields, such as autonomous driving, robotics, medical image analysis, video surveillance, and more. Consequently, it becomes imperative to enhance the accuracy and precision of semantic image segmentation both in theoretical research and practical implementation. This study primarily introduces the PSPNet, a scene analysis model based on pyramid synthesis,[9] along with a parameter optimization approach tailored to the PSPNet model, leveraging GPU distributed computing for improved efficiency.

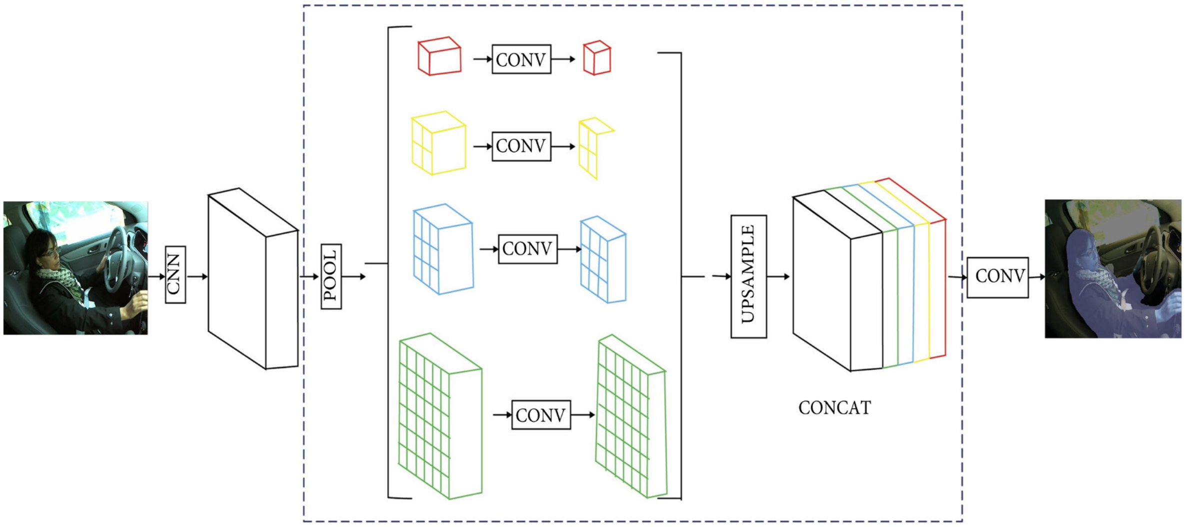


Figure 2. Diagram of semantic segmentation of images used in PSPNet Model

In figure 2, we offer an overview of our proposed PSPNet. Our approach commences with an input image (a) and initially employs a Convolutional Neural Network (CNN) to extract the feature map from the final convolutional layer (b). Following this, we apply a pyramid parsing module to collect a wide array of subregion representations. Subsequently, we utilize upsampling and concatenation layers to construct the final feature representation (c), which encapsulates both local and global context information. To conclude, this representation is input into a convolutional layer to generate the ultimate per-pixel prediction (d).

We term this module as the 'pyramid pooling module,' which is meticulously crafted to establish a global scene context based on the final-layer feature map of the deep neural network, as vividly depicted in part (c) of figure 2.

The pyramid pooling module adeptly amalgamates features across four distinct pyramid scales. The coarsest level, highlighted in red, engages in global pooling to yield a single-bin output. The subsequent pyramid level further subdivides the feature map into discrete sub-regions, creating pooled representations for various spatial locations.

Proposed Method

The method we propose is composed of the following phases:

Semantic Segmentation Utilizing the Pyramid Scene Parsing Network (PSPNet) model, we delve into a critical domain of computer vision known as semantic segmentation. This aspect is fundamental for addressing various scene interpretation challenges. Semantic segmentation involves the prediction of the category associated with each pixel or, more precisely, determining the category to which a given pixel belongs. By precisely delineating the object region to which a pixel pertains, our goal is to enhance the accuracy of pixel categorization.

Method: We employed three evaluation metrics—precision, recall, and F1 score-to assess and analyze the effectiveness of our image segmentation proposal based on PSPNet, drawing from our experimental results. In our implementation, we utilized the multi-scale parallel convolutional neural network model with PSPNet. To tackle the challenges posed by the complex and dynamic nature of the images, diverse driver positions, and the risk of overfitting or disrupting the parameter structure, we employed small-sample transfer learning as a means to constrain the parameter learning process.

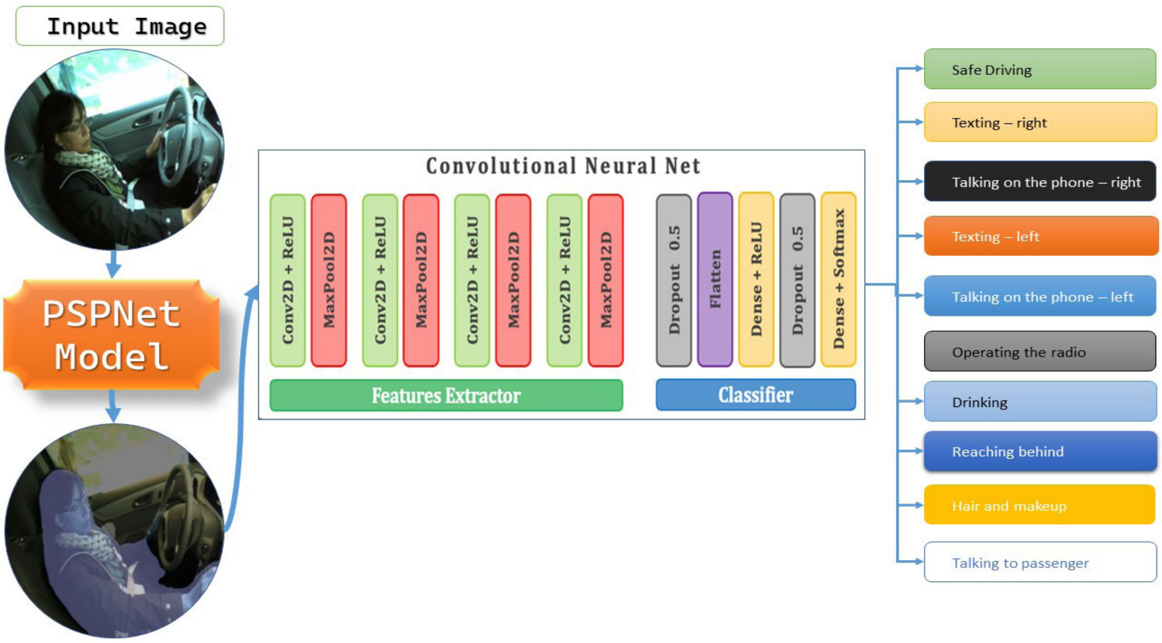


Figure 3. The pipeline of our proposed method for driver distraction classification. Pyramid Scene Parsing Network for Driver Distraction Classification

Classification: For the classification of the driver's pose (and thus her/his distraction), we used The convolutional neural network (CNN). Figure 3 The diagram below describes the pipeline of the proposed method. After segmentation of the database images with PSPNET Model, the proposed convolutional neural network (CNN) is used as the classification model.

RESULTS

Dataset

The dataset we used to train and test the models is the StateFarm's distraction detection dataset.[7] Table 2 present the 10 classes of the dataset and the number of images for each class.

Table 2. StateFarm distraction-detection dataset and total images in the class

Class	Driver state/action Images	Class	Class Driver state/action Images
C0	Safe driving/2489	C1	Texting - right/2267
C2	Talking on the phone - right/2317	C3	Texting - left/2346
C4	Talking on the phone - left/2326	C5	Operating the radio/2312
C6	Drinking/2325	C7	Reaching behind/2002
C8	Hair and makeup/1911	C9	Talking to passenger/2129

Implementation Details

Python is used to implement the suggested preprocessing and classification pipeline. A pre-trained PSPNET Model was employed for the pose estimation step. The Keras library on the Tensorflow backend was used to implement the CNN baseline model. 80 percent of the dataset was used to train all classifiers (including CNN), and the remaining 20 percent was used to test them.

Table 3. Classification model on the test set, in termes of accuracy, macro average precision, recall and F1-score. 10 classes schema

Model	Input	Accuracy	Precision	Recall	F1-Score
CNN	Segmented images	98,43	98,44	98,39	98,40

We compared the performance of the proposed method detection and classification for driver distraction using semantic segmentation. To evaluate the performance of different classifiers, we used four performance metrics that are: accuracy, macro average precision, macro average recall, and macro average F1Score.

CONCLUSION

In this paper, we proposed a method for driver distraction detection and classification. Our method introduces a comprehensive library for semantic segmentation for well-known model as PSPNet. The classification was conducted using CNN. We have demonstrated the efficacy and robustness of these models.

REFERENCES

1. Chen, S., Song, Y., Su, J., Fang, Y., Shen, L., Mi, Z., Su, B.: Segmentation of field grape bunches via an improved pyramid scene parsing network. International Journal of Agricultural and Biological Engineering 14(6), 185–194 (Dec 2021). https://doi.org/10.25165/ijabe.v14i6.6903

2. Cheng, B., Chen, L.C., Wei, Y., Zhu, Y., Huang, Z., Xiong, J., Huang, T.S., Hwu, W.M., Shi, H.: SPGNet: Semantic Prediction Guidance for Scene Parsing. pp. 5218–5228 (2019), https://openaccess.thecvf.com/content_ICCV_2019/html/Cheng_SPGNet_Semantic_Prediction_Guidance_for_Scene_Parsing_ICCV_2019_paper.html

3. Ersal, T., Fuller, H.J.A., Tsimhoni, O., Stein, J.L., Fathy, H.K.: Model-Based Analysis and Classification of Driver Distraction Under Secondary Tasks. IEEE Transactions on Intelligent Transportation Systems 11(3), 692–701 (Sep 2010). https://doi.org/10.1109/TITS.2010.2049741, conference Name: IEEE Transactions on Intelligent Transportation Systems

4. Romero-Carazas R. Prompt lawyer: a challenge in the face of the integration of artificial intelligence and law. Gamification and Augmented Reality 2023;1:7-7. https://doi.org/10.56294/gr20237.

5. Gupta, D.: Image Segmentation Keras : Implementation of Segnet, FCN, UNet, PSPNet and other models in Keras (Jul 2023). https://doi.org/10.48550/arXiv.2307.13215, http://arxiv.org/abs/2307.13215, arXiv:2307.13215 [cs]

6. Long, X., Zhang, W., Zhao, B.: PSPNet-SLAM: A Semantic SLAM Detect Dynamic Object by Pyramid Scene Parsing Network. IEEE Access 8, 214685–214695 (2020). https://doi.org/10.1109/ACCESS.2020.3041038, conference Name: IEEE Access

7. Shelhamer, E., Long, J., Darrell, T.: Fully Convolutional Networks for Semantic Segmentation (May 2016). https://doi.org/10.48550/arXiv.1605.06211, http://arxiv.org/abs/1605.06211, arXiv:1605.06211 [cs] version: 1

8. Gonzalez-Argote J. Analyzing the Trends and Impact of Health Policy Research: A Bibliometric Study. Health Leadership and Quality of Life 2023;2:28-28. https://doi.org/10.56294/hl202328.

9. State, F.: State Farm Distracted Driver Detection, https://kaggle.com/competitions/state-farm-distracted-driver-detection

10. Strayer, D.L., Turrill, J., Cooper, J.M., Coleman, J.R., Medeiros-Ward, N., Biondi, F.: Assessing Cognitive Distraction in the Automobile. Hum Factors 57(8), 1300– 1324 (Dec 2015). https://doi.org/10.1177/0018720815575149, https://doi.org/10.1177/0018720815575149, publisher: SAGE Publications Inc

11. Gonzalez-Argote D, Gonzalez-Argote J, Machuca-Contreras F. Blockchain in the health sector: a systematic literature review of success cases. Gamification and Augmented Reality 2023;1:6-6. https://doi.org/10.56294/gr20236.

12. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid Scene Parsing Network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6230– 6239. IEEE, Honolulu, HI (Jul 2017). https://doi.org/10.1109/CVPR.2017.660

FINANCING

The authors did not receive financing for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Abdelhak Khadraoui, Elmoukhtar Zemmouri.
Research: Abdelhak Khadraoui, Elmoukhtar Zemmouri.
Drafting - original draft: Abdelhak Khadraoui, Elmoukhtar Zemmouri.
Writing - proofreading and editing: Abdelhak Khadraoui, Elmoukhtar Zemmouri.