IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Vehicle Lane-Change Prediction on Highways Using Efficient Environment Representation and Deep Learning

R. IZQUIERDO[1], A. QUINTANAR[1] (Graduate Student, IEEE), J. LORENZO[1], I. GARCÍA-DAZA[1], I. PARRA[1], D. FERNÁNDEZ-LLORCA[1,2] (Senior, IEEE), and M. A. SOTELO[1] (Fellow, IEEE).

[1]Computer Engineering Department, Universidad de Alcalá, Alcalá de Henares, Spain
[2]European Commission, Joint Research Center, Seville, Spain.

Corresponding author: R. Izquierdo (e-mail: ruben.izquierdo@uah.es).

This work was supported in part by the Spanish Ministry of Science and Innovation under Grant DPI2017-90035-R and in part by the Community Region of Madrid under Grant S2018/EMT-4362 SEGVAUTO 4.0-CM.

ABSTRACT This paper introduces a novel method of lane-change and lane-keeping detection and prediction of surrounding vehicles based on Convolutional Neural Network (CNN) classification approach. Context, interaction, vehicle trajectories, and scene appearance are efficiently combined into a single RGB image that is fed as input for the classification model. Several state-of-the-art classification-CNN models of varying complexity are evaluated to find out the most suitable one in terms of anticipation and prediction. The model has been trained and evaluated using the PREVENTION dataset, a specific dataset oriented to vehicle maneuver and trajectory prediction. The proposed model can be trained and used to detect lane changes as soon as they are observed, and to predict them before the lane change maneuver is initiated. Concurrently, a study on human performance in predicting lane-change maneuvers using visual inputs has been conducted, so as to establish a solid benchmark for comparison. The empirical study reveals that humans are able to detect the 83.9% of lane changes on average 1.66 seconds in advance. The proposed automated maneuver detection model increases anticipation by 0.43 seconds and accuracy by 2.5% compared to human results, while the maneuver prediction model increases anticipation by 1.03 seconds with an accuracy decrease of only 0.5%.

INDEX TERMS Automated highways, CNNs, Deep Learning, Intelligent vehicles, Interaction-based, Lane-change prediction, Prediction algorithms, Surrounding Vehicles.

I. INTRODUCTION

AUTONOMOUS VEHICLES (AVs) are positioned as an essential player building tomorrow's automotive industry paradigm. They can be applied to address challenging problems originated from the transportation sector. AVs can drive efficiently by removing irrational human motivations from the decision-making process increasing fuel efficiency and reducing greenhouse gas emissions. AVs should operate also in a safer way, sensing the environment precisely and actuating accordingly within a fraction of a second. Moreover, AVs can take advantage of communications between themselves and the infrastructure using *Vehicle to Vehicle (V2V)* or *Vehicle to Infrastructure (V2I)* communication protocols, being aware of the road conditions virtually everywhere, outperforming sensor ranges and human perception.

Nowadays, commercial vehicles have reached automation level 3 according to SAE International specifications [16], becoming part of our lives. At some point, they will replace human-driven vehicles, yet meanwhile, AVs will share the road with human-driven vehicles, where different behaviors and interactions will arise between them. AVs can share their trajectories and operate in a coordinated mode, increasing fuel efficiency and safety. However, human-driven vehicles cannot share their trajectories or intentions, as they are self-generated at the execution time. In this scenario, AVs need to deal with uncertainties relative to human-driven vehicles while planning their trajectories. Predictions become a critical ability to understand how other traffic agents will act, especially if they do not communicate their intentions. Humans also make predictions and unconsciously apply them in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

TABLE 1. Overview of Public Datasets Used for Maneuver Prediction. *Ego* and *Top* labels are used to denote the in-vehicle and top-view acquisition points of view, respectively. The sensor setup is marked using a tick mark (✓) or an asterisk symbol (∗). The tick mark indicates that sensor data are available. The asterisk symbol represents that the sensor data are not available but have been used to compute some sort of information, such as trajectories.

Dataset	Release date	Length	Rate	View	Image	LiDAR	Radar	GNSS-RTK	Events	Lanes	Trajectories
NGSIM I80 [1]	Dec 2006	3 seq @ 15 min	10 Hz	Top	✓					✓	✓
NGSIM HW101 [2]	Jan 2007	3 seq @ 15 min	10 Hz	Top	✓					✓	✓
KITTI [3], [4]	June 2013	50 seq	10 Hz	Ego	✓	✓		RTK			✓
Oxford RobotCar [5]	Nov 2016	100 seq @ 2 h	33/10 Hz	Ego	✓	✓		GPS			
PKU [6]	Jan 2017	97 min	10 Hz	Ego		∗		GPS			✓
LISA-A [7]	Sept 2017	4 seq @ 100 sec	30/10 Hz	Ego	✓	✓	✓	GPS		✓	✓
ApolloScape [8]	Mar 2018	100 min	2 Hz	Ego	✓	✓		GPS			✓
BDD100K [9]	May 2018	100K @ 1 frame	30 Hz*	Ego	✓			GPS		✓	
Honda 3D Dataset (H3D) [10], [11]	June 2018	160 seq	2 Hz	Ego		✓		GPS	†		✓
HighD [12]	Oct 2018	16 h	25 Hz	Top	∗				†	✓	✓
Argoverse Motion Forecasting [13]	June 2019	320K seq @ 5 sec	10 Hz	Ego		∗		GPS		✓	✓
Waymo [14]	Aug 2019	1000 seq @ 20 sec	10 Hz	Ego	✓	✓		RTK			✓
PREVENTION [15]	Oct 2019	11 seq @ 30 min	10 Hz	Ego	✓	✓	✓	RTK	✓	✓	✓

their decision-making process. Experienced and alert human drivers usually anticipate lane changes of surrounding vehicles by using mainly visual information. This ability enables them to anticipate possible dangerous situations and to react appropriately, increasing safety and comfort. Considering a scenario where both AVs and human-driven vehicles share the road, deploying advanced prediction systems on AVs becomes a critical element to improve safety, fuel efficiency, and traffic flow.

Computation capabilities (e.g., GPUs) and CNN models have reached performance levels that enable complex image processing in real-time applications. The possibility to understand images and video sequences faster and better than humans brings the opportunity to predict the evolution of traffic scenes by using real-world images. AVs can take advantage of these algorithms to predict and anticipate critical situations. CNN models have proved to outperform human performance in some visual tasks. For example, in [17] the performance of professional annotators was analyzed over ImageNet samples [18], resulting 1.6% better than GoogleNet [19], the best-performing model at that moment. Currently, the EfficientNet-L2 model [20] outperforms professional annotators by 3.3% when classifying images. This better performance of deep learning-based models compared to the performance of humans in other contexts remains to be discovered.

This work presents a novel lane-change and lane-keeping detection and prediction system using state-of-the-art CNN-based classifiers and a new efficient representation of the environment that seamlessly encodes spatial and temporal information related to appearance, context, vehicle trajectories, and interactions. The PREVENTION dataset was used to develop and evaluate the model. In addition, an empirical study was conducted to assess the ability of humans to predict lane changes based on visual information, which is the main (but not the only) source of information used when driving. In this way, an approximate comparison between human capabilities and the proposed automatic system to predict lane changes of surrounding vehicles can be established. We performed a comprehensive experimental evaluation, including comparison with different CNN architectures, and assessing the relationship between anticipation and accuracy.

Following the introduction in section I, section II reviews in depth the most relevant works and datasets that address the maneuver prediction problem. Section III describes the developed maneuver detection and prediction models. Section IV presents an empirical study conducted to evaluate the human ability to anticipate lane changes in highway scenarios. The results of the proposed algorithm are presented and discussed in section V. Finally, conclusions and future work are detailed in section VI.

II. STATE OF THE ART

Machine learning approaches, including deep learning, rely on two main elements: data and models. In this section, both datasets that may be used to deploy maneuver detection or prediction models, and the different approaches developed to detect or predict this type of events are reviewed.

A. DATASETS

The datasets analyzed in this section are summarized in Table 1 analyzing three aspects: the acquisition point of view, sensor setup, and availability of data and manual annotations. Regarding the acquisition point of view, static recording systems from a top-view perspective, such as NGSIM [1], [2] and HighD [12] have many advantages over in-vehicle recording systems. They are unaffected by occlusions and provide a complete and commonly more precise understanding of the scene. However, they cannot be directly applied to onboard applications. Datasets recorded from onboard sensors have the advantage of being directly deployed, cutting

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

TABLE 2. Summary of Lane Change Prediction State Of the Art Works.

Work			Input				Prediction
Authors	**Year**	**Dataset**	**Kinematics**	**Context**	**Interaction**	**Model**	**Target**
Kasper et al. [21]	2012	Own	Ego + surr	✓	3x3 Grid	BN	Surr
Graf et al. [22]	2013	Own	Ego + 2 veh	✓	3 Veh model	CBR	Surr
Kumar et al. [23]	2013	Own	Surr	✓	-	SVM	Surr
Schlechtriemen et al. [24]	2014	Own	Ego + surr	✓	Rel. speed	GMM	Surr
Yoon et al. [25]	2016	NGSIM	Single	✓	-	ANN	Single
Bahram et al. [26]	2016	Own	Ego + surr	✓	Game Theory	BN	Surr
Lee et al. [27]	2017	Own	Ego + surr	✓	BV Grid	CNN	Surr
Deo et al. [28] [29]	2018	NGSIM	Center + surr	✓	3X2 Grid	LSTM+CSP	Center
Patel et al. [30]	2018	Own	Center + surr	✓	3X2 Grid	SRNN	Center
Li et al. [31]	2019	NGSIM	Center + surr	✓	3X2 Grid	DBN	Center
Kruger et al. [32]	2019	Own	Ego + surr	✓	3X2 Grid	GPNN	Center

the gap between development and deployment phases. On the other hand, observations are affected by occlusions and the quality of the data is usually lower.

If we have a glance at the sensors used to build these datasets it is possible to identify three main types: camera, *Light Detection and Ranging (LiDAR)*, and radar. The most common are cameras, followed by LiDARs, and lastly, radars. NGSIM and Berkeley Deep Drive (BDD100K) [9] datasets are only based on image systems. Others, such as H3D and PKU [6], [10], [11] are based on LiDAR. Some of them use a combination of camera and LiDAR solutions like KITTI, Oxford RobotCar, ApolloScape, and Waymo [3]–[5], [8], [14]. Amazingly, only the LISA-A [7] and the *PREdiction of VEhicle iNTentION (PREVENTION)* [15] datasets provide radar detections, which are one of the best choices among expensive LiDARs or complex stereo camera systems for object detection, highlighting its robustness. Modern cars are currently equipped with radar sensors as an essential element for proactive security systems, such as *Automatic Emergency Braking (AEB)*, *Collision Avoidance System (CAS)*, and *Adaptive Cruise Control (ACC)*. Datasets recorded from a mobile platform are usually equipped with a *Global Positioning System (GPS)* system for global positioning tasks, often complemented with an *Inertial Measurement Unit (IMU)*. On the opposite side, datasets recorded from a static point of view, i.e., HighD and the NGSIM, are referred to a static road reference system, so global positioning is not needed.

These datasets were developed to fill a gap in a particular field or research topic. The most valuable part of a dataset is the metadata or annotations generated by experts to measure or label specific circumstances. Lane level information is necessary for a precise scene understanding, lanes create dependencies between vehicles, especially in highway scenarios. KITTI, PKU, ApolloScape, H3D, and Waymo datasets do not provide lane-level information. NGSIM and HighD datasets were recorded in straight stretches of highways, and the trajectories are intrinsically referred to the lane reference system. The Argoverse dataset provides a detailed map of the recording area. The PREVENTION dataset offers on-road lane detections to generate a virtual configuration of the road structure. The H3D, HighD, and the PREVENTION datasets are the only ones with event labels. H3D labeled the ego-vehicle events and their motivation. HighD dataset creators announced that maneuver classification annotations would be available soon. In this case, lane-change maneuvers can be potentially computed by combining lane information and trajectories. However, the beginning of the lane change is much more relevant and difficult to define. The PREVENTION dataset includes manual annotations of the lane-change maneuvers, establishing their limits. As far as we are concerned, the PREVENTION dataset is the only one that enables the development of algorithms for detection or prediction of surrounding vehicles' maneuvers from an onboard point of view providing visual information and precise lane change labels at a sufficient frame rate.

B. LANE-CHANGE DETECTION/PREDICTION

Maneuver detection, recognition, or prediction are based on three basic features: motion, context, and interaction. A study conducted to analyze the most relevant features to predict lane changes [24] concludes that the lateral offset, the lateral speed w.r.t the lane axis, and the relative speed to the preceding vehicle are the three most relevant ones. These three variables are a combination of kinematic variables (position and speed), context (lane structure), and interaction (relative speeds). Keeping this fact into consideration, we will follow a review of the state of the art analyzing input variables, type of generated outputs, and also the dataset used to develop these models. Table 2 provides a summary of the works reviewed in this analysis and provides extra information relative to the used algorithms.

The most used data in vehicle prediction problems are vehicle state variables, such as position, speed, acceleration, orientation, and yaw rate [22], [23], [33]. These variables define the kinematic and dynamic state of a vehicle, endowing past, present, and future position understanding. The vehicle's width and length, together with its state, define the occupation of the road space. Road parameters, such as lane marking, the number of lanes, lane width, lane curvature,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

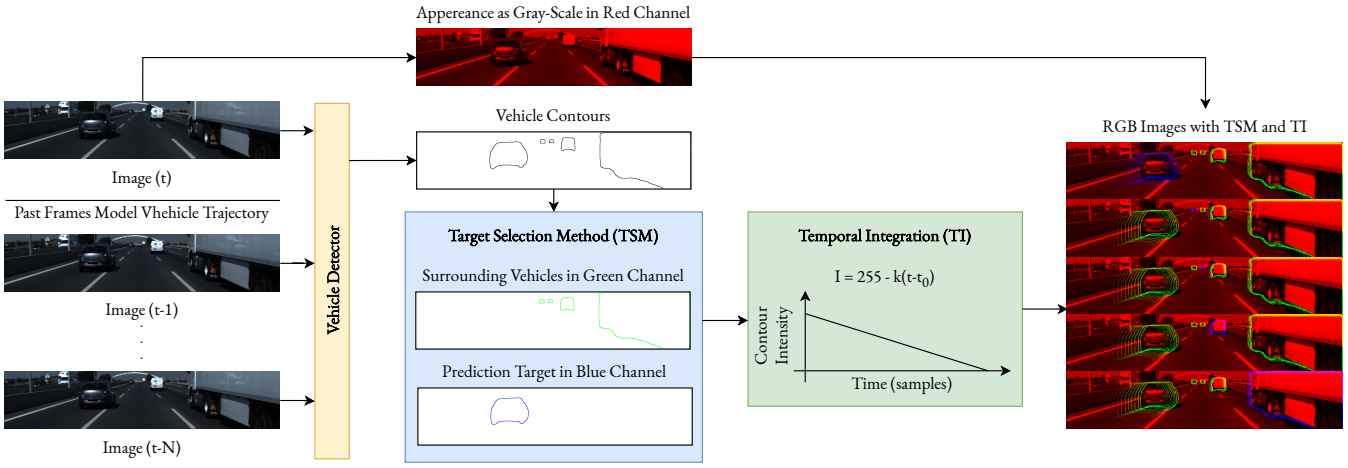


FIGURE 1. Image Encoding Process. The original RGB image is converted into a gray-scale image and stored in the red channel. The TSM extracts vehicle contours from the original RGB image and stores them into the green and blue channels depending on what vehicle is treated as the prediction target. The TIM depicts the contours of the vehicles in each corresponding channel with an intensity value proportional to their age. Finally, the three single-channel images are merged into a new and enriched RGB image.

type of lines, entries, and exits are used to model context. The easiest way is to transform vehicle positions into the lane reference system [25], [26], [31], or include different variables, such as the distance to the lane markings [24]. Maneuvers and trajectories are both caused by and also affect other traffic participants. Consequently, interactions between traffic agents need to be modeled. They can be incorporated in many different ways, from simple ones such as relative speeds [24], to complex scene representations. The most used representation model is a fixed vehicle configuration with a 3×2 or a 3×3 fixed grid representation based on the front and rear vehicles on the prediction targets' lane and the two adjacent lanes [28], [30], [32], [34]. Special consideration is made in [21], where a 3-agent model is proposed, including the ego-vehicle, the prediction target on an adjacent lane, and the preceding vehicle of the prediction target. This configuration is a simplification of the 3×2 or 3×3 configurations for cut-in lane-change maneuvers. Other approaches incorporate interactions indirectly by using appearance information directly from the images, with regions of interest of different sizes, centered on each vehicle [35].

Attending to the classification time they could be considered as detections, before the *Lane Change Event (LCE)*, and as predictions, before the *Lane Change Begining (LCB)*. Some works addressed maneuver predictions, such as [36]–[38] according to this definition. However, their results are obtained using different datasets and provided in terms of classification accuracy instead of lane change anticipation, so that the potential of these approaches cannot be compared.

Emphasis should be made on the prediction target, which can be focused on the ego-vehicle or on a surrounding vehicle for onboard recorded data. For top-view datasets, the prediction subject cannot be defined as an ego-vehicle or as a surrounding vehicle, because that direct relationship does not exist. Various works relative to ego-maneuver predictions have been omitted for reasons of space and their relative technical simplicity. Works based on top-view datasets can be considered in terms of maneuver-prediction complexity at an intermediate position between ego and surrounding vehicle predictions from an onboard point of view.

The datasets used to develop these works are wide, and most of them rely on private datasets. The use of public datasets is limited to the NGSIM and some works based on ego-vehicle predictions on the PKU dataset [39]–[41].

III. SYSTEM DESCRIPTION

This section describes the developed maneuver detection and prediction model, a deep learning-based approach to detect and predict the intended maneuvers of surrounding vehicles at highway scenarios from an onboard point of view.

The maneuver detection or prediction problem faces the following situation: given a scene, a predictive maneuver model must correctly assess all the future maneuvers that have not started yet while a maneuver detection model must classify the currently observed maneuver. Turning towards practical application, the desired behavior of a maneuver-aware system is the combination of both ideas, predicting lane changes as soon as possible while detecting them until their end, warning about possible critical situations.

Using the PREVENTION dataset, images, vehicle contours, and lane-change labels are employed to develop the maneuver prediction system. This deep learning-based approach is tackled from an image classification approach. Given a specific input image, a particular output label is desired for each vehicle. Hereafter, we will refer to the classification term due to the problem's approach, but this is not only limited to the classification of ongoing actions. A future action, not observed at the current moment, can be predicted through the classification of images previous to the beginning of the lane change.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

A. PROBLEM APPROACH

Action detection and prediction must deal with two problems in highway environments. The first one is the recognition of the prediction target. An image can present a scene with several vehicles. Simultaneously, some of them could be performing a lane change, while others would keep in their lane. Thus, different outputs must be possible for the same input image. Fig. 2 shows an image with three vehicles in which one of them is in the process of changing lanes, while the other two are not. In this example, three outputs are expected, specifying each maneuver. To solve this problem we have implemented a *Target Selection Method (TSM)*. The second problem is the highly temporal dependency of lane change maneuvers as we stated in II. Sequences of images have a better chance to detect or predict time-based actions, such as lane changes. The easiest solution could be to stack images as a 4D volume, but the problem scales rapidly and becomes computationally infeasible for training devices, and also for executing in real-world testing. This problem has been solved with a *Temporal Integration Method (TIM)*. Fig. 1 describes graphically the image encoding process that includes the TSM, the TIM, and the context codification. This process relies on vehicle contours and the original RGB image as input to generates one enriched image for each vehicle as output.

1) Target Selection Method

The TSM, using the contours of the vehicles provided in the dataset, draws the shape of the prediction target in a separate single-channel image (blue channel) and the surrounding vehicles, which are considered as interactive elements, in another single-channel image (green channel). Fig. 1 represents the whole image encoding process, where the TSM is represented inside the blue box. This mechanism generates a pair of single-channel images for each vehicle in the scene, simplifying the multi-vehicle prediction problem. This problem also arose in the user test described in section IV when users need to be focused on a specific vehicle. The mechanism used to focus the attention of users is based on the bounding box of the prediction target.

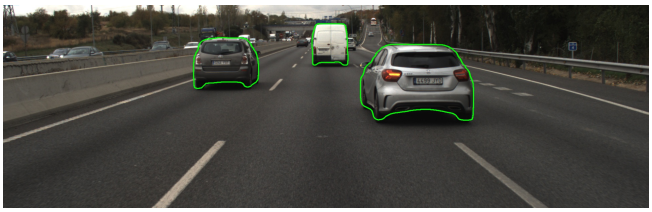


FIGURE 2. Original PREVENTION image with representation of the vehicle's contours.

2) Temporal Integration Method

The TIM creates a temporal representation of the vehicle trajectories. The contour of each vehicle is drawn using an intensity value proportional to the age of its position, i.e., remote (older) positions in time shall display a reduced intensity value. This method allows to represent up to 255 different poses of vehicles in a single-channel 8-bit image. The shape of vehicles can be represented in two different ways: using a contour line or a filled contour. The filled contour representation has an important disadvantage compared to the contour line representation. When the contours become bigger, newer representations can partially or totally overlap the older ones and eventually vanish the motion pattern. In contrast, contour lines minimize the overlapped area, and the information persists in the representation.

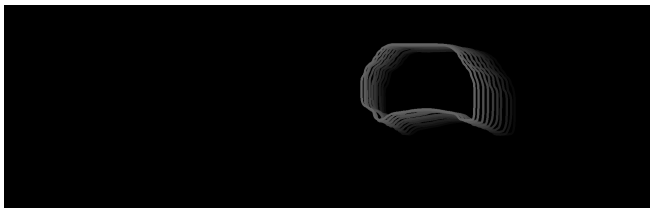


FIGURE 3. Example of TIM. Different gray-scale values represent different vehicle positions at different time instants.

Fig. 3 shows an example of the TIM applied to a single vehicle trajectory with 10 poses using the full 8-bit span. The TIM is applied to each image generated in the TSM stage. This step creates the motion pattern of all the vehicles in each corresponding image, as a prediction target or as an interactive element. The TIM is represented by the green rectangle in the image encoding process (see Fig. 1).

3) Context Codification

The TSM and the TIM generate a pair of single-channel images that solves the two problems stated before. However, context understood as lane markings and road configuration is necessary for a correct scene understanding and also to improve detections or predictions. Due to the nature of the image classification approach used, context could be easily included as the original image. To preserve the original input data size the original RGB image is converted to a gray-scale single-channel image (red channel).

At this point, three single-channel images that integrate context, target selection, vehicle motion, and interactions are generated. They are combined into an enriched RGB image, as it can be observed at the end of the image encoding process in Fig. 1. Each image is ready to be labeled with the desired output action for each corresponding selected target.

The contours needed to generate the enriched images are directly available in the dataset but the generation of this information for real-world testing would take 65 ms on an RTX2080Ti GPU. Additionally, and after this computation time, 1.24 ms are needed to create each enriched image, using an i7-7700K CPU.

B. INPUT DATA LABELING

According to the image classification approach, the expected inputs are images, which are generated as explained in sec-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

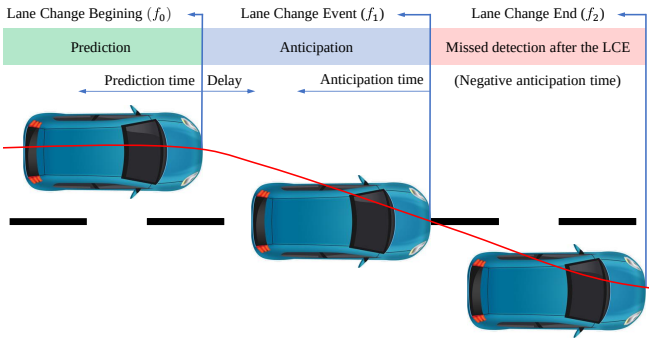


FIGURE 4. Lane change time references.

tion III-A, and the expected outputs are categories, in our case the intended maneuver of each prediction target.

The maneuvers considered in this problem are simplified into three categories: *Lane Keeping (LK)*, *Left Lane Change (LLC)*, and *Right Lane Change (RLC)*. The PREVENTION dataset provides manually annotated labels for each recorded *Lane Change (LC)* maneuver. For our interest, lane-change annotations can be described by the *ID* of the involved vehicle, the type of lane change, that can be LLC and RLC, and the lane-change temporal limits. Fig. 4 shows a graphic description of the temporal limits of a lane change maneuver.

The LCB (f_0) is characterized by the first the activation of the turn signal or the beginning of the lateral displacement. The LCE (f_1) is denoted when the middle of the rear part of the vehicle is at the divisor line. Finally, the end of the maneuver (f_2) is defined by the end of the lateral displacement over the destination lane. The LK label is defined by the absence of an LC maneuver and has no temporal limits. The vehicle ID and the temporal lane-change limits endow the labeling of each input image into each corresponding category.

As commented before, maneuver detection is directly related to the classification of ongoing maneuvers. Using the temporal limits of the lane changes the desired anticipation period, defined as t_p, can be added previously to each maneuver to implement the prediction of oncoming maneuvers.

C. MODELS

The models used to learn driving patterns to detect and predict LC and LK maneuvers are presented in Table 3. These models are specifically designed to extract high but also low-level information from input images to finally classify them. The images generated from the PREVENTION dataset are 1920×600 RGB images. These images were re-scaled to match the size constraints of the used models, which is 224×224. The output layer was replaced with a three-class softmax layer. The training hyperparameters used were: optimizer Adam, mini batch 64, epochs 2, shuffle every epoch, initial learning rate 10^{-4}, weight decay every epoch by 10, momentum 0.9, and L2 regularization 10^{-4}. The weighted cross entropy loss function was used to deal with the imbalanced number of LC and LK samples [42], using a class weight proportionally inverse to the number of samples in each category.

TABLE 3. CNN models used for maneuver classification. Classification performance on ImageNet. Computation time obtained with RTX2080Ti.

Name	Conv layers	top1-acc	top5-acc	ms / it
ResNet101	101	78.5%	93.9%	11.17
ResNet50	50	78.4%	94.1%	5.63
ShuffleNet	50	70.9%	89.8%	6.37
GoogleNet	22	–	–	7.24
ResNet18	18	59.4%	81.3%	2.37
SqueezeNet	18	57.5%	80.3%	2.71
AlexNet	8	63.3%	84.6%	1.19

IV. HUMAN BASELINE FOR MANEUVER PREDICTION

This section introduces the *User Prediction Challenge*, a call to evaluate the ability of humans to detect and predict lane changes using scenes recorded in the PREVENTION dataset. This study has the goal to set a baseline for comparison.

This section is structured as follows: subsection IV-A describes briefly the structure of the study, including a description of the used sequences. Subsection IV-B outlines the selection of participants and provides demographic information. Finally, subsection IV-C presents findings and results derived from the study.

A. METHODOLOGY

The study has been conducted using sequences extracted from the PREVENTION dataset. Some of the LC maneuvers were kept out of this study for different reasons, such as consecutive lane changes. The LK maneuvers were manually selected to show similar situations as those represented in the LC sets. These situations are commonly overtaking and car following with or without a posterior lane change.

Each user test is composed of a total of 30 sequences, 10 LK, 10 LLC, and 10 RLC. Sequences are extracted randomly from each subset and displayed in random order. A random period between 5 to 10 seconds is added previously to the LC maneuver to provide some anticipation gap and variability. This information is omitted to the study's subjects to avoid assumptions. On average, each sequence takes 15 seconds. The complete user test takes no longer than 10 minutes, including video displaying and user interactions.

1) Interface

The test interface has been developed using a QT application. This interface allows us to generate a random set of sequences, display them, and record the user's responses. Sequences are video fragments taken from the front camera and displayed at their natural frame rate.

LK maneuvers do not expect any input from the user side. If this happens, the sequence is stored as a correct detected LK maneuver at the end of the video clip. If the user stops the video clip at some point, the sequence is considered a misclassification, and the type of lane change predicted

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

by the user is stored. For LC maneuvers it is expected to interrupt the video clip at the moment the user detects or predicts the lane change. If this happens, the frame when the user interrupted the video (f_u) is stored together with the predicted or observed sense of the lane change (left or right). If the video clip ends without any input from the user side the maneuver is then classified as an LK maneuver.

B. PARTICIPANTS

Participants were recruited among the Engineering School, from students to teachers and other research staff, as well as family, friends, and colleagues. Thus, more significant variability is achieved in terms of age, occupation, and participants' driving experience. A total of 72 people did the test in two weeks. They provide some demographic information by filling up a small questionnaire. Then they perform the trial evaluating a total of 2160 sequences, 720 LK, and 1440 LC maneuvers.

Subjects were asked to fill a short form before doing the user test. This form has the goal to record some demographic information that could be related to their prediction performance. Each user was assigned to an ID to ensure anonymity. Users were asked with the following form:

- ID, age and gender.
- Occupation: study / work / both / none.
- Has driving license: yes / no
  - Driving experience: $\leq$ 1 yr. / 1-2 yr. / $>$ 2 yrs.
  - Driving frequency: daily / weekly / occasionally.
  - Driving areas: urban / highway.

The form is divided into two parts. The first part collects some personal and demographic information. The second part records driving skills and habits. The main features can be observed in Fig. 5. The anticipation of driving behaviors can be closely related to driving skills according to [43].

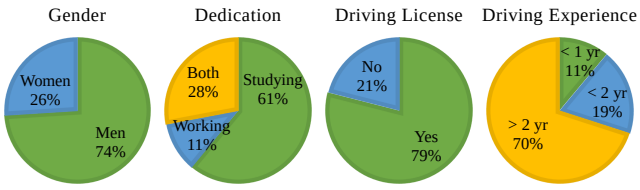


FIGURE 5. Demographic and driving skills distribution.

C. RESULTS

This subsection reviews the data generated by test participants. Detection and prediction results are provided at maneuver level and temporal-wise.

1) Classification Results

How precise humans are detecting and predicting lane changes can be evaluated through accuracy, precision, and recall. These values are computed according to (1), where TP is the true positive, FP the false positive, FN the false negative, and N is the total number of samples.

$$Acc = \frac{TP}{N} \quad Pre = \frac{TP}{TP+FP} \quad Rec = \frac{TP}{TP+FN} \tag{1}$$

2) Maneuver Detection

The maneuver detection criterion states that LC maneuvers detected after the LCE ($f_u \geq f_1$) are considered late detections and account as LK samples. The confusion matrix for the detection criteria is provided in Table 4. According to this, 83.9% of the maneuvers are correctly detected.

TABLE 4. Confusion Matrix of evaluated maneuvers for the detection criteria.

	Classified Class			
Target Class	*LLC*	*LK*	*RLC*	**Recall %**
LLC	624	47	49	86.7
LK	73	584	63	81.1
RLC	37	79	604	83.9
Precision %	85.0	82.2	84.4	83.9

Temporal results can be measured as maneuver anticipation with respect to the LCE ($f_u - f_1$), and as maneuver delay with respect to the LCB ($f_u - f_0$). Fig. 6 and 7 show the average anticipation and delay of each user in a sorted way, respectively. All the values in Fig. 6 are negative due to late detections are considered as LK maneuvers. The mean users' anticipation is -1.66 seconds. Negative values in Fig. 7 represent predictions. Note that only four users achieved a negative delay. The mean users' delay is 1.08 seconds.

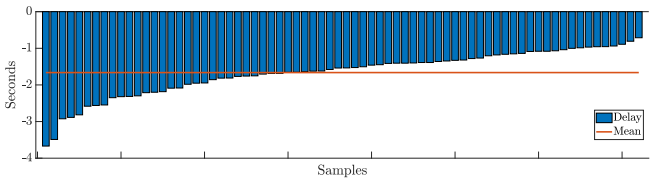


FIGURE 6. Average user anticipation ($f_u - f_1$).

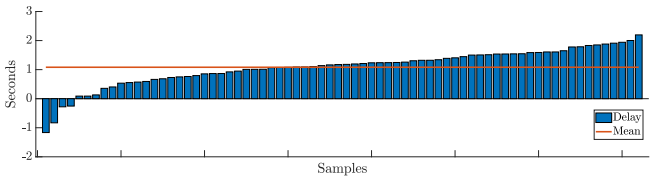


FIGURE 7. Average user delay ($f_u - f_0$).

3) Maneuver Prediction

Lane changes stated after its beginning cannot be considered as predictions. Following this definition, LC maneuvers classified after the LCB ($f_u \geq f_0$) are considered as late predictions and account for LK maneuvers. Table 5 shows the confusion matrix for the maneuver prediction criteria. It can be observed that only 15.6% of LLC and 10.6% of RLC are predicted, which is a dramatic performance reduction compared with the detection performance (86.7% and 83.9%).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

TABLE 5. Confusion Matrix of evaluated maneuvers for the prediction criteria.

	Classified Class			
Target Class	*LLC*	*LK*	*RLC*	**Recall %**
LLC	112	559	49	15.6
LK	73	584	63	81.1
RLC	37	607	76	10.6
Precision %	50.4	33.4	40.4	32.7

4) Accuracy vs Delay

Time-based decisions can become more accurate as much as the decision is delayed. Fig. 8 shows user accuracy versus user delay. Data has been fitted to a 1^{st} order polynomial. The equation parameters reveal that the average accuracy is close to 80% for a delay value equal to zero. However, the prediction criteria establish a delay equal to zero and the obtained accuracy is as low as 32.7%. This suggests that this model overestimates the average *human's performance*.

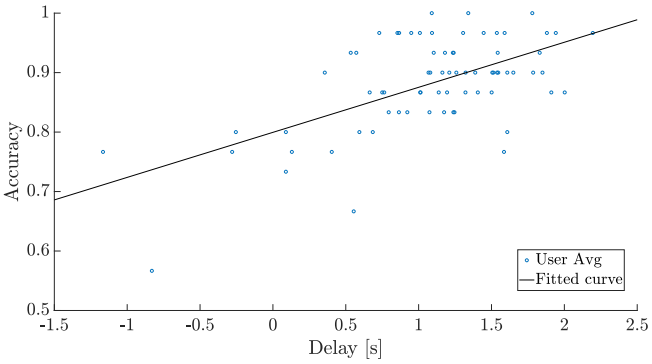


FIGURE 8. Accuracy vs. Delay. $r = 0.076x + 0.799$.

V. RESULTS

This section presents the results generated by training and evaluation of the classification models presented in Table 3. Models were trained for two purposes: detect and predict lane-change maneuvers. To do so, two labeling strategies were used by modifying the parameter t_p. For the detection goal we used $t_p = 0$. For the prediction goal we used $t_p = 10$. This enables up to 10 samples of prediction. The results of both methods will be compared with the *User Prediction Challenge*, described in section IV.

A. MANEUVER DETECTION

This subsection presents maneuver detection results with a parameter $t_p = 0$. As a classification approach, image classification results evaluate how many samples are correctly categorized without temporal implications or assumptions. This result is machine-learning oriented, and it is directly generated after the models' training. Table 6 provides the classification performance of each model, classifying single samples as isolated elements. It can be observed that ResNet50 achieved the best results, reaching 86.9% accuracy.

Single-sample results were integrated into maneuvers by using a Markov model base on the probabilities provided by the CNN models and the a priori probabilities of each event and each transition. The procedure used to classify consecutive temporal-integrated outputs into maneuvers is the same as the described in section IV. Additionally, to evaluate all the LC maneuvers commonly, their length has been normalized to 1 to prevent weighting effects between longer and shorter maneuvers.

TABLE 6. Image-level classification results. Models trained to detect ongoing maneuvers ($t_p = 0$).

		AlexNet	SqueezeNet	GoogleNet	ResNet18	ResNet50	ResNet101
	Accuracy	73.8	76.8	77.8	85.5	**86.9**	82.7
LK	**Precision**	74.9	77.2	79.8	93.0	93.4	88.6
	Recall	93.3	94.3	93.0	90.0	91.1	91.7
LLC	**Precision**	66.8	74.3	67.4	54.3	57.9	57.9
	Recall	36.4	40.7	43.0	62.2	65.3	46.7
RLC	**Precision**	72.4	76.1	73.2	63.2	68.8	64.1
	Recall	45.9	51.0	51.0	70.7	74.0	62.1

Table 7 presents average numeric results for each trained model. Anticipation is provided in seconds before the LCE and relative to the maneuver's length as a percentage. The *Area Under the Curve (AUC)* gathers in a single number the ability to detect as soon as possible all the maneuvers, including those which are not detected. Finally, accuracy is included in Table 7 to introduce the binomial anticipation versus accuracy. Accuracy for LC and LK set is provided. These two accuracy values are referred to the original PREVENTION training set, *All* is the overall accuracy for all the LC and LK maneuvers. The *Avg* set refers to the set used in the *User Prediction Challenge* in section IV.

Fig. 9 graphically represents the performance of the trained models detecting only LC maneuvers. Each maneuver is sorted based on the detection time. LC maneuvers that were detected after the LCE are placed on the left side of the graph, with an equivalent detection frame to f_1. The early detections are located on the right side of the chart, where higher levels of anticipation are achieved. The AUC value in Table 7 is extracted from this representation.

TABLE 7. Maneuver-level detection results ($t_p = 0$).

	Anticipation		**AUC**	**Accuracy**			
	[s]	[%]		LC	LK	All	Avg
AlexNet	2.33	78.9	0.72	87.9	65.4	68.0	80.9
GoogleNet	2.40	81.0	0.75	88.8	68.4	70.7	81.7
SqueezeNet	2.43	83.3	0.78	90.9	66.6	69.5	82.9
ResNet18	2.16	74.1	0.62	80.0	85.1	84.5	81.6
ResNet50	2.09	72.6	0.66	87.3	84.5	84.8	86.4
ResNet101	2.28	77.3	0.66	79.7	79.0	79.1	80.9

Referring to Table 7 it is possible to observe that the simplest models (AlexNet, GoogleNet, and SqueezeNet) anticipate more than ResNet models in general (2.43 seconds for SqueezeNet and 2.09 seconds for ResNet50). However, basic models achieve higher accuracy for LC maneuvers,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

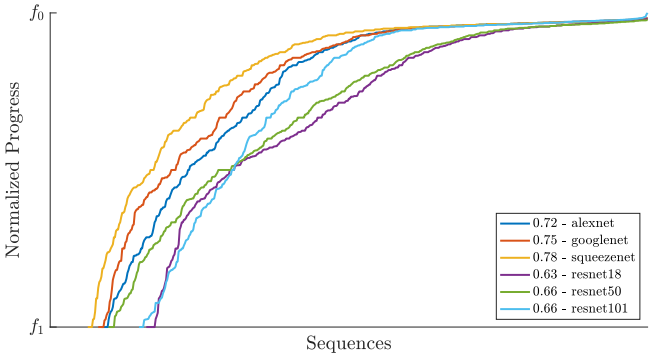


FIGURE 9. Representation of normalized detections of lane-change maneuvers. AUC value showed for each model.

while performing worse in detecting LK maneuvers (90.9% vs 66.6%, SqueezeNet Table 7). On the other hand, ResNet models achieve quiet similar accuracy for LC and LK maneuvers (87.3% vs 84.5%, ResNet50 Table 7). This effect is boosted by the imbalance number of maneuvers in the training set. For the balanced set under the column *Avg*, the number of LK is similar to the LC maneuvers and the ResNet50 model overcomes all the other models. This behavior together with the anticipation and the single-samples classification results suggests that the simplest models are more reactive to small variations, while ResNet models are more robust.

Comparing the human's and model's ability to assess maneuvers correctly, humans reached 83.9% accuracy (see Table 4) and 1.66 seconds of anticipation. AlexNet, GoogleNet, SqueezeNet, ResNet18, and ResNet101 models reached accuracy levels classifying maneuvers below the human ability but higher anticipation periods. Model ResNet50 overcomes both human accuracy and anticipation in 2.5% and 0.43 seconds, respectively.

B. MANEUVER PREDICTION

This subsection presents maneuver prediction results, using a parameter $t_p = 10$. This means that 10 samples previous to every lane-change maneuver have been labeled as lane-change samples. The LK maneuvers were also extended 10 samples from the beginning. The behavior expected from the predictive models is to classify lane-change maneuvers both before they have started and while they are being carried out.

Following the same structure presented in subsection V-A results are analyzed as single-sample results. Table 8 shows the performance of each CNN model trained for the classification of current and future maneuver state. It can be observed that there are no significant differences between ResNet models. Compared with the simplest models, ResNet's performance is from 10 to 12 points higher.

Single-sample results were integrated into maneuvers according to the same procedure described for the maneuver detection in subsection V-A.

Table 9 presents average numeric results for each trained model at maneuver-level, similarly as Table 7. If a maneuver

TABLE 8. Image-level classification results. Models trained to predict maneuvers ($t_p = 10$).

		AlexNet	SqueezeNet	GoogleNet	ResNet18	ResNet50	ResNet101
	Accuracy	73.1	75.6	76.0	**84.2**	**84.4**	**84.1**
LK	**Precision**	75.6	76.3	78.1	92.5	92.3	90.1
	Recall	91.8	93.8	92.5	89.0	89.3	90.9
LLC	**Precision**	58.8	71.8	61.9	49.7	51.4	58.2
	Recall	34.7	39.5	40.7	59.0	59.0	53.5
RLC	**Precision**	67.8	74.0	73.8	59.6	60.7	65.0
	Recall	43.6	48.4	47.5	67.7	68.2	65.0

is correctly predicted its anticipation can be higher than 100%, as the length considered to normalize the maneuver does not include the prediction period. Prediction is presented in two formats: as a percentage of the number of predicted maneuvers, and as a time value, which represents the average prediction time for those predicted. The AUC value is composed as the addition of the ordinary detection AUC (with a maximum value of 1) and the prediction AUC (with a maximum value of 1).

Fig. 10 depicts the performance of the trained models detecting and predicting LC maneuvers. The representation is the same used in Fig. 9, but predicted maneuvers are represented above f_0, which is the LCB. Two areas can be observed in this chart, one from f_1 to f_0 (detection AUC) and the other from f_0 to f_p (prediction AUC).

TABLE 9. Maneuver-level prediction results ($t_p = 10$).

	Ant.		**Pred.**		**AUC**	**Accuracy**			
	[s]	[%]	[s]	[%]		LC	LK	All	Avg
AlexNet	2.97	105.3	0.77	66.1	1.23	86.0	66.9	69.1	79.6
GoogleNet	3.01	106.8	0.75	65.6	1.29	90.4	67.5	70.2	82.8
SqueezeNet	3.11	110.5	0.77	70.2	1.36	90.7	67.0	69.8	82.8
ResNet18	2.58	90.3	0.69	45.3	0.95	82.7	84.8	84.5	83.4
ResNet50	2.58	90.9	0.71	45.7	0.96	82.2	84.5	84.2	83.0
ResNet101	2.69	94.9	0.72	49.5	1.03	83.5	83.3	83.3	83.4

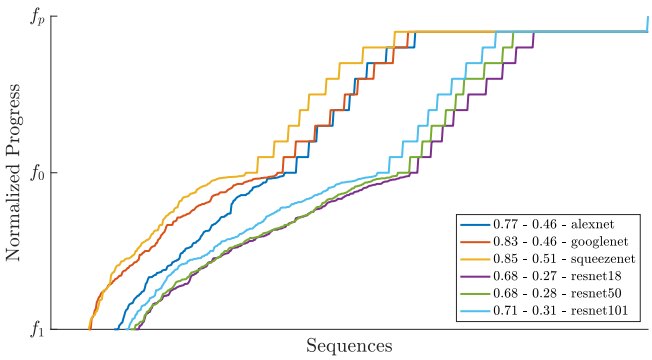


FIGURE 10. Representation of normalized predictions of lane-change maneuvers. Detection area from f_1 to f_0. Prediction area from f_0 to f_p. Detection and prediction AUC showed for each model.

The same effects as for the detection approach were ob-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

served for the prediction approach. Looking at Table 7 it can be observed that the simplest models anticipate more than ResNet models (3.11 seconds for SqueezeNet and 2.69 seconds for ResNet101). The percentage of predicted maneuvers is higher for the simplest models again, (70.2% for SqueezeNet versus 49.5% for ResNet101) and also the prediction time (0.77 seconds for SqueezeNet and 0.72 for ResNet101). The simplest models achieve higher accuracy for LC maneuvers but they are worst at detecting LK maneuvers (90.7% vs 67.0%, SqueezeNet). On the other hand, ResNet models achieve a quite similar accuracy for LC and LK maneuvers (83.5% vs 83.3%, ResNet101).

Comparing the human's and model's ability to assess maneuvers correctly, humans reached 83.5% accuracy with 1.66 seconds of anticipation on average. The simplest models' accuracy is on average 0.7% below the human performance. However, SqueezeNet model has average anticipation of 3.11 seconds, which increases human anticipation by 1.45 seconds. ResNet models have reached a performance comparable to human accuracy, but with higher anticipation periods. ResNet101 almost matches human's accuracy (83.4%) and increases the anticipation time by 1.03 seconds.

C. ANTICIPATION VS ACCURACY

There is a clear relationship: the better anticipation or prediction, the lower the accuracy is. Both features are compared one by one and together with human performance in Fig. 11. The regression model used to fit human performance is represented along with the models trained to detect and predict maneuvers depicted with a $+$ and a $\times$ symbol, respectively. It is easy to understand which models perform better by observing this representation. The more top right the model, the better its global performance is. It can be observed that all the trained models are located above the human's performance line. This means that all the models have higher accuracy or anticipation than average human performance. We can highlight ResNet50 due to its higher accuracy, reasonable anticipation, and consistency for both detection and prediction of maneuvers.

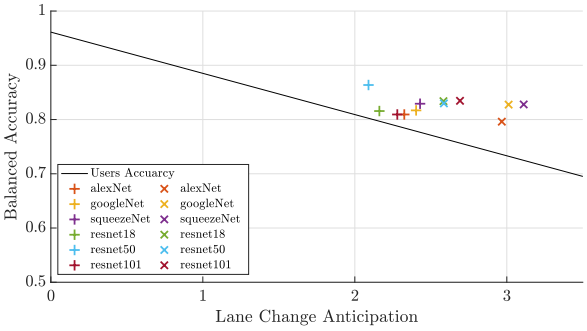


FIGURE 11. Models performance. Accuracy vs Anticipation. Detection models represented with $+$ symbols and prediction models represented with $\times$ symbols. The solid black line represents the average human performance.

VI. CONCLUSION AND FUTURE WORK

This work presents a *Convolutional Neural Network (CNN)*-based model to detect and predict lane-change and lane-keeping maneuvers in highway scenarios from an onboard point of view using deep learning models.

Context information, vehicle interaction, motion histories, and a target selection method are efficiently encoded in an enriched image to detect and predict lane changes in highway scenarios from an onboard point of view by using a CNN classification model. This kind of representation has the advantage of being virtually unlimited in terms of the number of vehicles in the scene and the number of past vehicle representations. In contrast with 4D image inputs, the data size does not scale with the number of temporal instances. Unlike the state-of-the-art approaches, this novel image-based proposal is not limited to a fixed number of interaction vehicles or a fixed vehicle distribution and it is not distance-based. Besides, none of the existing works employs the image appearance to predict or classify maneuvers.

The User Prediction Challenge showed that, surprisingly, humans cannot predict lane-change maneuvers regularly, at least with the set of sequences used in this experiment. This means that PREVENTION sequences are challenging even for humans. Users' delay in predicting lane-change maneuvers has been evaluated. On average, they are detected 1.08 seconds after the lane change has started and 1.66 seconds before the middle point of the vehicle reaches the dividing line between the lanes.

The developed system has proven that it outperforms human's accuracy by 2.5% and anticipation by 0.43 seconds for the models trained for maneuver detections and human's accuracy by -0.5% and anticipation by 1.03 seconds for the models trained for maneuver prediction. The social study focuses the user's attention on a single target marked with a red rectangle. However, real driving scenarios require to be focused on all the vehicles simultaneously, reducing the user's reaction capacity. Actual human reaction times can be expected to be even higher than those observed in this experiment.

A. FUTURE WORK

Based on results and conclusions derived from this work, several research lines can be followed to improve the performance of the system or either take advantage of this system in other applications. The developed prediction model could be improved by exploring the use of visual attention modules in the network architecture. This mechanism has proven to increase the performance in multi-object problems. It could be interesting to exploit this approach to predict maneuvers at intersections from a static and extrinsic point of view, such as infrastructure cameras. This prediction model could help to manage the traffic at controlled intersections, avoiding the need for V2X communications and dealing with non-automated vehicles. The deployment of the predictive model in an AV running at real time would allow its use as input for an ACC or a CAS system.

REFERENCES

[1] J. Halkias and J. Colyar, https://www.fhwa.dot.gov/publications/research/operations/06137/06137.pdf, December 2006, NGSIM - Interstate 80 Freeway Dataset [Online; accessed Jan. 9 2019].

[2] J. Colyar and J. Halkias, "NGSIM - US Highway 101 Dataset," Jan 2007, https://www.fhwa.dot.gov/publications/research/operations/07030/07030.pdf [Online; accessed Jan. 9 2019].

[3] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets Robotics: The KITTI Dataset," International Journal of Robotics Research (IJRR), 2013.

[4] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite," in Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[5] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," The International Journal of Robotics Research (IJRR), vol. 36, no. 1, pp. 3–15, 2017. [Online]. Available: http://dx.doi.org/10.1177/0278364916679498

[6] U. of Peking, "University of Peking database (PKU)," http://poss.pku.edu.cn/download/.

[7] A. Rangesh, K. Yuen, R. K. Satzoda, R. N. Rajaram, P. Gunaratne, and M. M. Trivedi, "A multimodal, full-surround vehicular testbed for naturalistic studies and benchmarking: Design, calibration and deployment," arXiv preprint arXiv:1709.07502, 2017.

[8] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The ApolloScape Dataset for Autonomous Driving," arXiv preprint arXiv:1803.06184, 2018.

[9] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling," 2018.

[10] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward Driving Scene Understanding: A Dataset for Learning Driver Behavior and Causal Reasoning," in Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[11] A. Patil, S. Malla, H. Gang, and Y.-T. Chen, "The H3D Dataset for Full-Surround 3D Multi-Object Detection and Tracking in Crowded Urban Scenes," in International Conference on Robotics and Automation, 2019.

[12] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems," in 2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC), 2018.

[13] M.-F. Chang, J. W. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3D Tracking and Forecasting with Rich Maps," in Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[14] "Waymo Open Dataset: An autonomous driving dataset," 2019.

[15] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. Sotelo, "The PREVENTION dataset: a novel benchmark for PREdiction of VEhicles iNTentIONs," in 2019 IEEE Intelligent Transportation Systems Conference (ITSC). IEEE, 2019, pp. 3114–3121.

[16] S. Taxonomy, "Definitions for terms related to driving automation systems for on-road motor vehicles (J3016)," Technical report, Society for Automotive Engineering, Tech. Rep., 2016.

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein et al., "ImageNet Large Scale Visual Recognition Challenge," International journal of computer vision, vol. 115, no. 3, pp. 211–252, 2015.

[18] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in CVPR09, 2009.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," in Computer Vision and Pattern Recognition (CVPR), 2015. [Online]. Available: http://arxiv.org/abs/1409.4842

[20] H. Pham, Z. Dai, Q. Xie, M.-T. Luong, and Q. V. Le, "Meta pseudo labels," arXiv preprint arXiv:2003.10580, 2020.

[21] D. Kasper, G. Weidl, T. Dang, G. Breuel, A. Tamke, A. Wedel, and W. Rosenstiel, "Object-Oriented Bayesian Networks for Detection of Lane Change Maneuvers," IEEE Intelligent Transportation Systems Magazine, vol. 4, no. 3, 2012.

[22] R. Graf, H. Deusch, M. Fritzsche, and K. Dietmayer, "A Learning Concept for Behavior Prediction in Traffic Situations," in IEEE Intelligent Vehicle Symposium (IVS), 2013, pp. 672–677.

[23] P. Kumar, M. Perrollaz, S. Lefevre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in 2013 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2013, pp. 797–802.

[24] J. Schlechtriemen, A. Wedel, J. Hillenbrand, G. Breuel, and K. D. Kuhnert, "A lane change detection approach using feature ranking with maximized predictive power," in 2014 IEEE Intelligent Vehicles Symposium Proceedings, June 2014, pp. 108–114.

[25] S. Yoon and D. Kum, "The multilayer perceptron approach to lateral motion prediction of surrounding vehicles for autonomous vehicles," in 2016 IEEE Intelligent Vehicles Symposium (IV), June 2016, pp. 1307–1312.

[26] M. Bahram, C. Hubmann, A. Lawitzky, M. Aeberhard, and D. Wollherr, "A Combined Model- and Learning-Based Framework for Interaction-Aware Maneuver Prediction," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, pp. 1538–1550, June 2016.

[27] D. Lee, Y. P. Kwon, S. McMains, and J. K. Hedrick, "Convolution neural network-based lane change intention prediction of surrounding vehicles for ACC," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Oct 2017, pp. 1–6.

[28] N. Deo and M. M. Trivedi, "Multi-Modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs," in IEEE Intelligent Vehicle Symposium (IVS), 2018, pp. 1179–1184.

[29] ——, "Convolutional social pooling for vehicle trajectory prediction," in IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR), 2018, pp. 1468–1476.

[30] S. Patel, B. Griffin, K. Kusano, and J. J. Corso, "Predicting Future Lane Changes of Other Highway Vehicles using RNN-based Deep Models," 2018.

[31] J. Li, B. Dai, X. Li, X. Xu, and D. Liu, "A Dynamic Bayesian Network for Vehicle Maneuver Prediction in Highway Driving Scenarios: Framework and Verification," Electronics, vol. 8, no. 40, 2019.

[32] M. Kruger, A. S. Novo, T. Nattermann, and T. Bertram, "Probabilistic Lane Change Prediction using Gaussian Process Neural Networks," in IEEE 22th International Conference on Intelligent Transportation Systems (ITSC), 2019, pp. 3651–3656.

[33] A. Benterki, M. Boukhnifer, V. Judalet, and C. Maaoui, "Artificial Intelligence for Vehicle Behavior Anticipation: Hybrid Approach Based on Maneuver Classification and Trajectory Prediction," IEEE Access, vol. 8, pp. 56 992–57 002, 2020.

[34] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," IEEE Access, vol. 7, pp. 38 287–38 296, 2019.

[35] M. Biparva, D. Fernández-Llorca, R. Izquierdo, and J. K. Tsotsos, "Video action recognition for lane-change classification and prediction of surrounding vehicles," arXiv preprint arXiv:2101.05043, 2020.

[36] J. Wiest, M. Höffken, U. Kresel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models," in 2012 IEEE Intelligent Vehicles Symposium, 2012.

[37] E. Yurtsever, Y. Liu, J. Lambert, C. Miyajima, E. Takeuchi, K. Takeda, and J. H. L. Hansen, "Risky Action Recognition in Lane Change Video Clips using Deep Spatiotemporal Networks with Segmentation Mask Transfer," in 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019.

[38] Z. Wei, C. Wang, P. Hao, and M. J. Barth, "Vision-Based Lane-Changing Behavior Detection Using Deep Residual Neural Network," in 2019 IEEE Intelligent Transportation Systems Conference (ITSC), 2019.

[39] H. Zhao, C. Wang, Y. Lin, F. Guillemard, S. Geronimi, and F. Aioun, "On-Road Vehicle Trajectory Collection and Scene-Based Lane Change Analysis: Part I," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 1, pp. 192–205, Jan 2017, http://poss.pku.edu.cn/download [Online; accessed Jan. 9 2019].

[40] W. Yao, Q. Zeng, Y. Lin, D. Xu, H. Zhao, F. Guillemard, S. Geronimi, and F. Aioun, "On-Road Vehicle Trajectory Collection and Scene-Based Lane Change Analysis: Part II," IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 1, pp. 206–220, Jan 2017.

[41] R. Izquierdo, I. Parra, J. Munoz-Bulnes, D. Fernández-Llorca, and M. A. Sotelo, "Vehicle trajectory and lane change prediction using ANN and SVM classifiers," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Oct 2017, pp. 1–6.

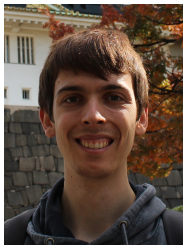
[42] Y. S. Aurelio, G. M. de Almeida, C. L. de Castro, and A. P. Braga, "Learning from imbalanced data sets with weighted cross-entropy function," Neural processing letters, vol. 50, no. 2, pp. 1937–1949, 2019.

[43] S. De Craen, D. A. Twisk, M. P. Hagenzieker, H. Elffers, and K. A. Brookhuis, "Do young novice drivers overestimate their driving skills more than experienced drivers? Different methods lead to different conclusions," Accident Analysis and Prevention, vol. 43, no. 5, pp. 1660–1665, sep 2011.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3106692, IEEE Access

RUBÉN IZQUIERDO received the Bachelor's degree in electronics and industrial automation engineering in 2014, the M. S. in industrial engineering in 2016, and the Ph.D. degree in information and communication technologies 2020 from the Universidad de Alcalá (UAH). He is currently a Postdoc researcher at the Computer Engineering Department, UAH. His research interest is focused on the prediction of vehicle behaviors but also in control algorithms for highly automated and cooperative vehicles. His work has developed a predictive ACC and AES system for cut-in collision avoidance.

ÁLVARO QUINTANAR (Graduate Student Member, IEEE) obtained the Bachelor's Degree in Telecommunications Engineering in 2017 from Universidad de Alcalá (UAH), specializing in Telecommunication Systems, and the Master's Degree in Telecommunications Engineering in 2019 from Universidad de Alcalá (UAH), earning a specialization in Intelligent Transportation Systems. He started his work in the INVETT Research Group in October 2018, where he is currently pursuing the PhD degree in Information and Communications Technologies with the Computer Engineering Department, developing interactive prediction systems that could forecast trajectories and intention of other road users, such as human-driven vehicles and VRUs.



JAVIER LORENZO obtained a Degree in Telematics Engineering in 2015 and his Master's Degree in Telecommunication Engineering, with a specialization in Space and Defense Technologies, in 2017, both from the Universidad de Alcalá (UAH). In the same year, he began his Ph.D. degree in Information and Communications Technologies in the INVETT research group (UAH). His doctoral thesis is focused on anticipating pedestrian crossing behavior through contextual information using mainly image features, using computer vision, and deep learning methods.

IVÁN GARCÍA-DAZA received the MSc and PhD degrees in Telecommunications Engineering from the University of Alcalá (UAH), Madrid (Spain), in 2004 and 2011 respectively. At present he is Assistant Professor at the Computer Engineering Department at the University of Alcalá and member of the INVETT research group since 2007. In this period he has collaborated on more than 20 projects with public and private funding. All the projects are related with computer science techniques applied on Intelligent Transportation System.

IGNACIO PARRA received the M.S. and Ph.D. degrees in telecommunications engineering from the University of Alcalá (UAH), in 2005 and 2010, respectively. He is currently an Associate Professor with the Computer Engineering Department, UAH. His research interests include intelligent transportation systems and computer vision. He received the Master Thesis Award in eSafety from the ADA Lectureship at the Technical University of Madrid, Spain, in 2006.

DAVID FERNÁNDEZ-LLORCA (Senior Member, IEEE) received the Ph.D degree in telecommunication engineering from the University of Alcalá (UAH) in 2008. He is currently Scientific Officer at the European Commission - Joint Research Center. He is also Full Professor with UAH. He has authored over 130 publications and more than 10 patents. He received the IEEE ITSS Young Research Award in 2018 and the IEEE ITSS Outstanding Application Award in 2013. He is Editor-in-Chief of the IET Intelligent Transport Systems. His current research interest includes trustworthy AI for transportation, predictive perception for autonomous vehicles, human-vehicle interaction, end-user oriented autonomous vehicles and assistive intelligent transportation systems.

MIGUEL ÁNGEL SOTELO received the degree in Electrical Engineering in 1996 from the Technical University of Madrid, the Ph.D. degree in Electrical Engineering in 2001 from the University of Alcalá (Alcalá de Henares, Madrid), Spain, and the Master in Business Administration (MBA) from the European Business School in 2008. He is currently a Full Professor at the Department of Computer Engineering of the University of Alcalá. His research interests include Self-driving cars and Predictive Systems. He is author of more than 250 publications in journals, conferences, and book chapters. He has been recipient of the Best Research Award in the domain of Automotive and Vehicle Applications in Spain in 2002 and 2009, and the 3M Foundation Awards in the category of eSafety in 2004 and 2009. Miguel Ángel Sotelo has served as Project Evaluator, Rapporteur, and Reviewer for the European Commission in the field of ICT for Intelligent Vehicles and Cooperative Systems in FP6 and FP7. He is member of the IEEE ITSS Board of Governors and Executive Committee. Miguel Ángel Sotelo served as Editor-in-Chief of the Intelligent Transportation Systems Society Newsletter (2013), Editor-in-Chief of the IEEE Intelligent Transportation Systems Magazine (2014-2016), Associate Editor of IEEE Transactions on Intelligent Transportation Systems (2008-2014), member of the Steering Committee of the IEEE Transactions on Intelligent Vehicles (since 2015), and a member of the Editorial Board of The Open Transportation Journal (2006-2015). He has served as General Chair of the 2012 IEEE Intelligent Vehicles Symposium (IV'2012) that was held in Alcalá de Henares (Spain) in June 2012. He was recipient of the 2010 Outstanding Editorial Service Award for the IEEE Transactions on Intelligent Transportation Systems, the IEEE ITSS Outstanding Application Award in 2013, and the Prize to the Best Team with Full Automation in GCDC 2016. He was President of the IEEE Intelligent Transportation Systems Society (2018-2019).