

Imitation learning based decision-making for autonomous vehicle control at traffic roundabouts

Weichao Wang¹ · Lei Jiang¹ · Shiran Lin¹ · Hui Fang¹ · Qinggang Meng¹

Received: 10 March 2021 / Revised: 31 December 2021 / Accepted: 14 January 2022 / Published online: 4 May 2022 © The Author(s) 2022

Abstract

The essential of developing an advanced driving assistance system is to learn humanlike decisions to enhance driving safety. When controlling a vehicle, joining roundabouts smoothly and timely is a challenging task even for human drivers. In this paper, we propose a novel imitation learning based decision making framework to provide recommendations to join roundabouts. Our proposed approach takes observations from a monocular camera mounted on vehicle as input and use deep policy networks to provide decisions when is the best timing to enter a roundabout. The domain expert guided learning framework can not only improve the decision-making but also speed up the convergence of the deep policy networks. We evaluate the proposed framework by comparing with state-of-the-art supervised learning methods, including conventional supervised learning methods, such as SVM and kNN, and deep learning based methods. The experimental results demonstrate that the imitation learning-based decision making framework, which ourperforms supervised learning methods, can be applied in driving assistance system to facilitate better decision-making when approaching roundabouts.

Keywords Advanced driving assistance system \cdot Intelligent roundabout-join task \cdot Imitation learning

1 Introduction

Autonomous vehicle (AV) and advanced driver assistance systems (ADAS) development is playing a key role in contemporary intelligence transportation systems. An autonomous vehicle captures environmental data via sensor techniques to navigate the vehicle without human interventions [2]. As highlighted in [44], AVs can not only carry out basic manipulations, such as acceleration, deceleration, braking, forward and backward movement, turning

Weichao Wang w.wang3@lboro.ac.uk

Hui Fang H.Fang@lboro.ac.uk

¹ Computer Science Department, Loughborough University, Epinal Way, Loughborough, UK

and other conventional vehicles functions, but also accomplish high-level tasks, such as mission planning, path planning, intelligent obstacle avoidance and all human-like behaviors. Although many AV manufacturers have made significant progress on AV development, e.g. Google self-driving car in the U.S. [21], VisLab's BRAiVE in Italy [9] and Jaguar Land Rover in the U.K. [26], it is still a great challenge for AV to make decisions under complex environments, e.g. a busy urban environment with multiple junctions [37] or with numerous objects moving in various directions [3].

Traffic roundabout is a looping junction where road traffic is restricted to go in one direction around a central island with priority given to the coming vehicles that have already entered the roundabout [29]. The roundabout systems in the U.K. have resulted in many accidents due to human drivers' misjudgements of the speed, distance or intention of approaching vehicles in the roundabout [14]. In addition, there are many different types of roundabouts, e.g., mini-roundabouts, signalized roundabouts and non-signalized roundabouts [37]. Therefore, it is challenging to provide intelligent recommendations to join a roundabout system without making any hassles to the entire system.

In recent years, artificial intelligence and machine learning methods have been widely applied to make decisions at complex junctions [14]. Qi et al. [40] uses convolutional neural networks (CNN) to detect vehicles so that an AV could make decisions based on the environmental contexts. In [15], behavioral rule-based model is built to take vehicle angles, speeds and diameters of crossroads into consideration to deal with issues happening at crossroads. In [44], an adaptive tactical behavior planner (ATBP) is proposed to simulate human-like motion behaviours at non-signalized roundabouts by analysing individual driver's historical navigation patterns. In [16], Gritschneder et al. design a reinforment learning framework to generate optimal actions via a multiple-layer perceptron neural network based on the observations obtained from GPS system to reflect the position and motion of other nearest vehicles. Imitation learning (IL) has become one of the most popular learning frameworks due to its advantages of leveraging domain expert knowledge [23]. An IL model shares similar idea of reinforcement learning but avoid the randomized control trials mechanism in reinforcement learning framework when optimizing control actions. It is more suitable to the tasks which cannot afford the costs incurred by the random trials. Furthermore, it can speed up the training process of the control policy deep models comparing to conventional supervised learning models. Thus, many IL systems, such as [19, 28, 38, 48], have been proposed to control AVs in many on-road and off-road tasks.

In this paper, we propose an IL-based decision-making system to provide intelligent recommendations to join a roundabout timely and smartly. In specific, a deep learning based IL system is trained to learn how human drivers to manipulate vehicles based on observations of other vehicles in roundabouts. In addition, we investigate how different backbone architectures, such as VGG-16 and ResNet-18 make impact on the learning performance. The novelty of our paper is highlighted as: (1) we propose an imitation learning-based decisionmaking system (ILBDM) to join roundabouts timely and safely. To our best knowledge, this is the first system to provide guidance for drivers at roundabout by exploiting imitation learning method. (2) we provide a new roundabout-entering dataset for AV research. As data is the main driving force for new deep learning-based algorithm development, our work has paved way to solve a difficult high level control task; and (3) we evaluate the proposed ILBDM system comprehensively to prove the superior performance of imitation learning method over supervised learning methods for a sequence of decision-making task.

This paper is organized as follows: Section 2 describes the related work which includes intelligent transportation system, neural computational models for autonomous vehicle decision-making, and autonomous vehicle decision-making in roundabout applications. In

Section 3, the proposed ILBDM system is explained in details. After presenting the overall IL framework, we provide the technical details of extracting observations from the driving environment, including car detection, motion feature extraction and backbone network architectures, and how we set up the reward schemes to train the system. In Section 4, we evaluate the performance of the proposed framework under different backbone network architectures. Furthermore, the experimental results demonstrate that the proposed method outperforms the systems with supervised learning algorithms. In Section 5, a conclusion is drawn and ideas for future research are discussed.

2 Related Work

2.1 Intelligence transportation System

Intelligent transportation system (ITS) has become the key to reduce the negative impact from traffic congestions and pollutions that are the most serious contemporary issues caused by the rapid urbanisation development [13, 17]. In [13], it summarises that ITS can solve these issues by using (1) routing optimisation, (2) intelligent traffic light control, and (3) decentralised multi-agent communications. Routing optimisation is an active research field of ITS. Many optimisation algorithms, such as genetic algorithm [5], ant colony algorithm [1] and particle swarm optimisation algorithm [10], have been proposed to make optimal path planning for vehicle navigation. Intelligent traffic light control system provides another solution to reduce traffic congestions. Chen et al. [10] proposes a real time traffic light control algorithm that adjusts both the sequence and length of traffic lights by using several traffic factors which include traffic volume, waiting time, and traffic density. Vallati et al. [47] designs a PDDL+ encoding planning module to optimise the traffic light control for solving those traffic congestions which are caused by unexpected accidental events. In [25], two intelligent traffic light control schemes are used in fog computing to deal with resisting malicious vehicles and single-point failure.

In addition to these ITS systems, advanced vehicle control systems have become an emerging technology to make contributions to solve these traffic problems as well as to enhance the driving safety.

2.2 Neural computational autonomous vehicle control

Control policy neural networks have been widely proposed in autonomous vehicle control systems since the work of [39]. In [39], A three-layer back-propagation neural network named ALVINN is proposed to take road images as input and produce travel directions as output. In [8], a deep learning network called PilotNet is used to estimate steering angles by extracting and finding salient objects from visual perceptional input data. Reinforcement learning (RL) approaches are deployed in many AV systems in recent years, e.g., [12, 53, 54]. Wolf et al. [54] proposes a deep Q-network (DQN) policy network to steer vehicle in a simulated driving environment. In [12], several deep reinforcement learning methods, including DQN, Deep Deterministic Actor Critic and Deep Attention Reinforcement Learning, are trained to control a vehicle on the Open-source Racing Car Simulator (Torcs) to demonstrate the feasibility of using RL framework for AV control tasks. In [53], An RL model predictive control neural network is trained to control a vehicle to run on an elliptical dirt track at the Georgia Tech Autonomoous Racing Facility. Although RL based methods

do not need any labeled training data, most of them have to be trained in a simulated environment to reduce the costs incurred by the exploration steps in the RL framework.

Imitation learning (IL) is an appealing deep learning framework to learn a policy network guided by human domain expert to speed up the network convergence as well as enforce strong constraints on the mapping space between input observations and output actions. In [59], Zeng et al. use LIDAR data and high definition maps to find trajectories that minimize predefined losses. [43] proposes to combine imitative model with goal-directed planning to outperform directing IL methods. In [7], a model named ChauffeurNet is trained by taking the advantages of both human expert's guided data and synthesized perturbations of the expert's driving data. In [11], Codevilla et al. assume that both perceptual input and driver intention are required to make optimal decisions. Therefore, a conditional imitation learning based model is proposed to consider dirver intention in the decision-making process.

2.3 Autonomous vehicle decision-making at roundabouts

As one of the most difficult decision-making tasks, vehicle control at roundabout has raised significant attentions during this decade [18, 34, 35, 37, 50, 51]. In [18], low-level texture features and motion features are extracted from monocular video sequences to detect and track moving vehicles in roundabouts. The method is tested on BRAiVE AV/ADAS system and achieve a good accuracy performance with a real-time processing speed. In [35], a panoramic stereo-vision based system is designed to detect upcoming vehicles and calculate the time-to-contact that defines the estimated time of potential collision with the egovehicle. In [37], Okumura et al. propose an action planning method for AVs to merge into a roundabout. In this work, four learning inputs (approaching car speed, difference in heading between the vehicle and the road, the distance from the vehicle to the merge point, and distance from the vehicle to the nearest branch point) can support AVs to make the right "enter", "wait" and "merge" decisions. In [49], grid-based image processing approach (GBIPA) is proposed to characterize traffic situations that can be used for machine learning algorithms to learn the roundabout joining criteria. Approaching car features (Position, direction and speed) can be extracted by proposed GBIPA as learning inputs, and the trained classifiers using the proposed GBIPA approach is evaluated on test videos captured at roundabouts, where the SVM yields the best performance with a 90.28% classification accuracy. In [50], Wang et al. designed a human-like decision-making system at mini-roundabouts based on both of front view and side view cameras. In addition, [51] extends some of previous works in [49] and proposes a multi-grid-based image processing approach using multiple cameras (MGC), it can deal with two issues: 1) the autonomous vehicles' can swiftly change the position/orientation when reaching a roundabout, and 2) The driver's views and behaviors can also be varied. Proposed MGC include different size of grid to boost the accuracy and to protect the autonomous vehicle when entering a roundabout.

3 Proposed ILBDM system

In this proposed work, we design an IL based decision-making algorithm to facilitate intelligent decisions to enter roundabouts. Considering that the vehicle control at roundabouts can be formulated as a sequence of decisions, IL method is more suitable to the task comparing to the convention supervised learning methods. In particular, the system learns a neural computational model by feeding human expert data to make strong constraints when searching the solution space to update the deep policy network. It differs to our previous work, i.e. [49–51] on two folders: first, this IL based model learns to maximize the expected rewards when taking an action at a timestamp whilst our previous work makes an IID assumption of the control actions at individual timestamps by using supervised learning methods. Secondly, we investigate whether deep policy backbone networks outperform the conventional decision models, such as SVM and kNN classifiers, for the roundabout decision-making task.

The proposed ILBDM system is a fast and reliable imitation learning-based approach. In particular, we deploy the Deep Q-Learning from Demonstrations (DQfD) method [22] as our IL system. Although there are labelling data from domain experts as guidance which is similar to the supervised learning framework, the IL highlights that the decisioin-making is a continuous process as the decisions made in the past can influence the decisions made in the future. Therefore, a well-learned function can map states to actions that could maximise the expected discounted rewards over the entire decision-making process. Following the assumption in the reinforcement learning framework, a Markov decision process is formulated for the task for IL learning. Here, a tuple (*S*, *A*, *R*, *T*, γ) consists of a set of states S, a set of actions A, a reward function R(s, a), a transition function T(s, a, s') = P(s'|s, a), and a discount factor γ . A policy network π is learned to provide recommendations on actions by maximizing cumulated discounted rewards which can be expressed as a function $Q^{\pi}(s, a)$:

$$Q^{\pi}(s,a) = E[R(s,a) + \gamma \sum_{s'} P(s'|s,a) \max_{a'} Q^{\pi}(s',a')]$$
(1)

Here, $Q^{\pi}(s, a)$ represents the expected cumulated discounted rewards, R(s, a) represents the immediate reward when taking an action a at state s, s' represents the state at the next timestamp and $Q^{\pi}(s', a')$ is the expected maximum reward if taking action a' at state s'.

The overview of the proposed framework is illustrated in Fig. 1. A monocular camera system mounted in front of our vehicle is used to capture video sequence data from the driving environment. The raw data are fed into a pre-processing pipeline to extract efficient observed states from the environment for the decision-making at roundabouts. This forms the state space S. At each timestamp, an action $a \in A$ is made by a deep policy network to maximize the expected cumulated rewards in the driving sequence guided by



Fig. 1 Main framework of the proposed ILBDM system. A decision of whether the AV should wait (0) or go (1) can be made at each time stamp

the action made by an expert driver. The training data is a set of observation-action pairs $D = \{\langle o_i, a_i \rangle\}_{i=1}^N$ generated by the expert driver in ILBDM system. The goal of ILBDM is to learn a policy that imitates an expert policy π given demonstrations from that expert driver π_E . A demonstration is defined as a sequence of state-action pairs that result from a policy interacting with the environment d={ s1, a1, s2, a2, ...}.

Regarding to the loss function, the proposed system learns a policy by minimizing the Huber loss function [46]) over the set of demonstrations with respect to the policy. The Huber loss is a loss function used in robust regression which is less sensitive to outliers in data than the squared error loss. It is described as the following equation:

$$L_{\delta}(y_t, Q_t) = \begin{cases} \frac{1}{2}(y_t - Q_t)^2 & \text{for } |y_t - Q_t| \le \delta\\ \delta |y_t - Q_t| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$
(2)

where y_t is the target output defined in target network:

$$y_t = r_t + \max_a Q^{target}(s_{t+1}, a).$$
 (3)

Here $Q_t = Q(s_t, a_t)$ represent the Q value from the deep policy network and δ is the control parameter which can be tuned in Eq. 2. The IL training network minimizes the loss until the model converges.

3.1 Perceptional observations

Effective observed states extraction improves the reliability of an intelligent decisionmaking system as it makes the system insensitive to noise signals from the complicated driving environment. There are two modules for the perceptional observation extraction from driving sequences. These include a vehicle detection module which uses the Faster R-CNN network [42] to dectect vechicles at roundabouts and a motion extraction module to extract their movement features based on an optical flow algorithm in [32]. Examples of extracting observations from driving sequence is illustrated in Fig. 2. Here, the detected ROIs are used as filters so that vehicle movements based on the optical flow algorithm can be extracted as the input of the decision-making DL policy networks.

Vehicle detection module is one of the key modules for the roundabout entering decision-making process. In our work, the faster R-CNN method originally proposed in [42] is adapted to detect the vehicle regions of interest. The faster R-CNN is a two-stage CNN based detection method which includes a Region Proposal Network (RPN) for proposals selection and a classifier to verify the objects from these candidates. The RPN uses



Fig. 2 Observations stage of proposed ILBDM system. Images in the first row shows Vehicle Detection results, images in the second row visualize the optical flow results from their corresponding frames (hue represents optical flow direction and saturation shows optical flow magnitude followed [32]) and images in the third row shows the masked optical flow as the observations

the first 13 covolutional layers of VGG-16 network to generate feature maps and two three-layer regressors to locate the anchor boxes which have high object scores as object proposals. Following the selection stage, these proposals are further verified by a classifier to decide whether there is a vehicle in each candidate box. In our work, a pre-trained model downloaded from [56] is used to detect vehicles at roundabouts.

After experimental comparisons of several state-of-the-art methods, the faster R-CNN is selected for the vehicle detection module in our system as it is the most effective method to process our collected data in terms of both accuracy and processing speed. The comparison with the methods including single shot detection (SSD) [31], inception [57] and mask R-CNN [60] which are shown in Table 1. In particular, 1000 frames from different weather conditions and various types of roundabouts extracted from random-selected 20 sequences are tested by using the four algorithms. It shows that the precision of the Faster R-CNN approach is 2.5% better than Mask R-CNN, 8.52% better than the Inception, and 23.55% better than SSD. For the detection timing per image, the Faster R-CNN approach spends 0.12 s on the detection of per image, which outperforms 0.03s faster than Mask R-CNN, 1.06s faster than Inception. Although detection time of SSD is the best in all the algorithms, the false negative (FN) rate of the detection is far from satisfation.

It is important to reduce the FN to a minimal level when considering that any missing detection of vehicles could be more risky comparing to the cases of false detection (FP). Therefore, we re-set parameters in the faster R-CNN method to ensure a minimal FN rate is achieved. As presented in Table 1, we can achieve 14 FN when we accept the FP number to 74 for designing our vehicle detection module. Because the frame rate is 30 frames per second, the false negative number is acceptable as there is averagely about one vehicle missed in every 100 frames. Although there are more false detection (FP) of vehicles in the sequences as illustrated in Fig. 3 (a-c), they bring little impact on the final decision-making as the filtered optical flow features are used as the perceptional observations and the movements in these false detection regions are not significant (illustrated in Fig. 3 (d-f)).

Motion extraction is the second core module for the perceptional observation extraction. As is mentioned in [24], approaching vehicle velocities are the most important feature when a decision is made to enter a roundabout. In our system, the optical flow algorithm in [32] is deployed to extract features for representing the vehicle movements. Due to its accuracy and robustness, this method has been widely used in many motion-based applications, e.g., [52, 55]. Figure 2 illustrates the estimated optical flows when our vehicle approaches a roundabout. Here, we use a color-map scheme to visualize the optical flow based on both its magnitude and its direction. The blue hue indicates the main direction of the optical flow is to the left while the red hue indicates the optical flow at the correponding pixels is to the

False ne	egatives, R.rate: Recall rate, P: Precision, TPI: Time per image)	
lable I	ILBDM venicle detection results (IP: True Positives, IN: True negative	es, FP: Faise positives, FIN:

Methods	TP	TN	FP	FN	R.rate(%)	P (%)	TPI (Avg)
Inception	609	201	81	109	84.81	88.26	1.18
SSD	476	229	174	121	79.73	73.23	0.08
Mask R-CNN	633	212	39	116	84.51	94.20	0.15
Faster R-CNN	752	135	33	80	90.38	95.80	0.12
Detect. Module	763	149	74	14	98.20	91.16	0.12



Fig. 3 (a,b,c) Raw data for vehicle detection. (d,e,f) Estimated flow field

right. It demonstrates that the movement feature can be an efficient representation for the control task. Due to the complicated environment at roundabouts, movements from irrelevant objects could easily distract the decision-making to enter the roundabout. Therefore, the ROIs from the vehicle detection module are set as masks to filter the optical flow feature which is illustrated in the third row of Fig. 2.

3.2 Decision making policy network backbones

Many DL based backbone networks have been developped for various learning tasks during this decade, e.g., AlexNet [4], GoogleNet [6], VGG family (including VGG-16 and VGG-19) [33], ResNet family (including ResNet-18, ResNet-34, ResNet-50 and ResNet-101) [45], and DenseNet [58]. Considering that the deep policy network for decision-making in our proposed system requires to output decisions with acceptable processing speed, we select three backbone networks as the candidate policy networks. These include a simple CNN, VGG-16 and ResNet-18. The architectures of these backbones are illustrated in Table 2

The CNN architecture is the default backbone used in DQfD. This architecture is concise and efficient for non-linear mappings, thus deploying in many classification and controlling tasks, e.g. [30] and [22]. As illustrated in Table 2, the network has three convolution layers followed by average pooling and ReLU as its activation functions. The first convolution layer contains 6 kenels with kernel size 5×5 , the second convolution layer contains 16 kernels with kernel size 5×5 and the third convolution layer contains 120 kernels with kernel size 5×5 .

VGG-16 is a popular convolution neural network model designed by Zimmermann et al. in [60]. As illustrated in Table 2, VGG-16 adopts a deeper network structure, which has 9 convolution layers. Max pooling is used in the network to make it easier to capture changes in images, bring greater local information differences, and describe edge textures better. It achieves a great trade-off between precision and process speed to perform as a backbone architecture for a real-time system.

ResNet architecture is playing a dominant role in many recent vision classification and control tasks [20]. The concept of residual learning can effectively reduce the impact of disappeared gradient issue as well as focus on learning detailed patterns. In our work, the ResNet-18 is used as one of the backbone architectures due to its suitability for the fast decision-making task.

Table 2 Dv	eep learning	network architec	tures										
CNN	(224*224 RGB Image)	conv 6, 5*5		average poo		conv 16, 5*5		average pool		conv 120, 5*5		full con- nection	softmax
VGG-16	(224*224 RGB Image)	conv 64, 3*3	maxpool	conv 128, 3*3	maxpool	conv 256, 3*3	maxpool	conv 512, 3*3	maxpool	conv 512, 3*3	maxpool	full con- nection	softmax
		conv 64, 3*3		conv 128, 3*3		conv 256, 3*3		conv 512, 3*3		conv 512, 3*3		full con- nection	
						conv 256, 3*3		conv 512, 3*3		conv 512, 3*3		full con- nection	
Resent-18	(224*224 RGB Image)	conv 64, 7*7		maxpool		(conv 64, 3*	3)*2 (conv 1	28, 3*3)*2 (co	onv 256, 3*3	2*	conv 512, 3*3	full con- nection	softmax
						(conv 64, 3*	3)*2 (conv 1	28, 3*3)*2 (co	onv 256, 3*3)	*2	conv 512, 3*3		

3.3 Decision making reward scheme

A reward scheme can be used to learn different driving strategies by setting rewards to encourage preferred behaviors. For example, setting larger rewards for the "Go" action when the demonstration provides "Go" guidance would learn a more aggressive driving behavior. While setting larger rewards for the correct "Wait" action can lead to a more cautious driving behavior. For the ILBDM system, the reward scheme works as a part of the Eq. 1, R(s, a), The return from a state is defined as the sum of discounted future reward at time t:

$$R_t = \sum_{i=t}^T \gamma^{(i-t)} r(s_i, a_i) \tag{4}$$

where T is the time-step when AV approaches a roundabout, with a discounting factor $\gamma \in [0, 1]$. Note that the γ is set to 0.8 in our experiments. For the work, we adopt a balanced reward scheme to train the system. A positive reward of 1 is provided at each step if the AV/ADAS's action is consistent with the human expert driver before entering the roundabout, i.e., true positive and true negative (currect prediction for "Go" and "Wait"), and 0 for inconsistent decision (false "Go", and false "Wait").

4 Experimental results

In this section, we present the experimtal results to demonstrate the performance of the ILBDM system. It includes the experimental settings, the reward and loss convergence during the training iterations, the decision making results, and the comparison with the benchmarking methods under the supervised learning framework, which include SVM, kNN, and three deep learning based classifiers. The DL based classifiers deploy the same backbone networks in our system.

4.1 Experimental settings

All the videos for this study are real-life driving recordings produced by a camera fixed on the right window of an ego vehicle in order to provide the road condition on the right side of the car. Video captured in that setting demonstrated the usual view of the drivers in a roundabout in the UK, where priority was given to approaching vehicles from right directions [27]. Nextbase 312GW cameras were used due to its quality reputation and its wide application in traffic experiments [49–51]. Nearly 50 different roundabouts across the Leicestershire, UK were filmed in 18 months, from October 2016 to April 2018. The time frames were 9 am to 11 am and 3 pm to 6 pm. The morning time normally provides satisfactory quality video in a natural daylight condition. The afternoon time provides busy traffic in the peak hours that maximized the intricacy in the roundabouts.

The experiments were run on a computer with an Intel Core i7-7700 CPU operating at 2.80 GHz and GTX1060 graphics card in order to evaluate the ILBDM performance. Tensorflow deep learning framework is adopted in this paper [41]. For video data collection, Images with 1920*1080 pixels and a frame rate of 30fps is taken from 130 videos when AV/ADAS approaches a roundabout. Video recorder is the main sensor used in this experiment. Furthermore, the data were splitted into training and test datasets. The training dataset contains 16,380 images (10415 for a wait before entering a roundabout, 5965 for go to roundabout) and the testing dataset contains 1800 images (1000 for a wait before entering a

			1	1 :		e	1	
VN	RN	SN	PSN	NSN	TR.P	TR.N	TE.P	TE.N
130	50	18180	6765	11415	5965	10415	577	965

 Table 3
 Learning sample statistics VN: The number of videos, RN: The number of roundabouts, SN: The number of samples, PSN: The number of positive samples, NSN: The number of negative samples

TR.P: Training positive sample, TR.N: Training negative sample, TE.P: Test positive, TE.N: Test Negative

roundabout, 800 for entering a roundabout) were built. The detailed training data statistics are shown in Table 3.

The benchmarking algorithms include both traditional machine learning techniques and DL based supervised learning classifiers. Support vector machine (SVM) and k-Nearest Neighbor [36] are the two classical conventional ML methods in the comparison. Here, the SVM classifier is an RBF SVM with $\gamma = 0.5$. Regarding to the kNN method, the k closest matching examples from the training dataset are retrieved by comparing the Euclidean distance of features in the feature space to make decisions for the test image. The k value is set to 5 in our experiment. Furthermore, we compare with the supervised learning based DL classifiers which deploy the same backbone networks to demonstrate the advantage of IL based framework.

4.2 Model training

The models are trained by using different learning algorithms: 1) traditional machinelearning based approach, i.e. RBF SVM. Here the other convention ML method, i.e. kNN, has not training stage as it uses a retrieval way to make decisions, 2) deep learning-based supervised classifiers (CNN, VGG-16 and Resnet-18), and 3) ILBDM learning based system (DQfD with different CNN policy networks, i.e., CNN, VGG-16 and Resnet-18). In supervised learning methods (machine-learning based approaches and deep learning-based networks), one can easily track the performance of a model during training by evaluating it on the training and validation sets. Fig. 4 shows the convengence of the accuracy on the training set and loss from DQfD with the three backbone newtorks respectively. The inputs of image size is 224*224, the learning rate is 0.0005, and epoch is 20. The convergence of the rewards and losses during the training of ILBDM with different CNN policy networks are illustrated in Fig. 4. It shows that all the backbone networks can achieve convergence after around 200k-300k interations. The timings of the training process of the methods are shown in Table 4. The timing for training the supervised learning DL methods with the three backbone networks are 1.35, 3.25 and 4.35 hours respectively. For the ILBDM system, the timing for the three backbones are 2.15, 6.45 and 7.25 hours respectively. It proved that the convergence of the imitation learning based methods are slower comparing to the supervised learning methods. However, the inference timings of networks in the ILBDM system are similar to the other classifiers.

4.3 Comparison results

We evaluate the proposed ILBDM system by comparing with conventional ML methods first. The comparison results are shown in Table 4. We use SVM and kNN as the classifiers to process the same observations extracted from images. The accuracy rates of SVM and kNN are 76.23% and 81.03% respectively. In addition, we also test the same data with our



Fig. 4 DQfD-CNN policy-based reward and loss curve, first row is from CNN, second row is from VGG-16, Third row is from Resnet-18

previous work named GBIPA-SC-NR in [49]. The GBIPA-SC-NR is a grid-based decisionmaking algorithm. After extracting measurements from grids divided evenly on an image, three conventional classifiers, including SVM, kNN and multi-layer perceptional (MLP) artificial neural netowork (ANN) are used to classify the data into decisions. The accuracy rates are 87.62%, 77.62%, and 81.49% respectively. The reason that the GBIPA-SC-NR ourperforms the same classifiers used on the observation data is because the dimension of

Deringer

Approach	Methods	Acc (%)	TT (Hours)	IT (Second)
Supervised learning	SVM	76.23	1.25	0.9467
	SVM	87.63	1.45	0.8234
GBIPA-SC-NR	kNN	77.62	N/A	1.1125
	ANN	81.49	1.35	1.2908
	CNN	87.36	1.35	0.1279
Deep Learning	VGG-16	92.57	3.25	0.1363
	ResNet-18	83.08	4.35	0.1233
	DQfD-CNN	96.21	2.15	0.1035
Proposed ILBDM	DQfD-VGG-16	93.32	6.45	0.1206
	DQfD-ResNet-18	89.56	7.25	0.1136

Table 4 Proposed ILBDM learning result (Acc:Accuracy, TT: Training Time, IT: Inference Time)

the feature space is much lower in the GBIPA-SC-NR which reduce the impact of the curse of dimensionality. For the proposed ILBDM system, the accuracy rates are 96.21%, 93.32% and 89.56% respectively. The accuracy demonstrates that the overall performance of the proposed system is significantly better comparing to the convetional ML methods.

Furthermore, we compare the results from the proposed system with the supervised learning based methods by using the same backbone networks. The accuracy rates are 87.36%, 92.57% and 83.08% for CNN, VGG-16 and ResNet-18 respectively. Although there are variations in the results, all of the networks in the ILBDM system outperforms the networks under the supervised learning framework. This demonstrates that the effectiveness of IL framework for the roundabout joining task. From a theoretical point of view, supervised learning models learn non-linear mapping functions to project observations captured by the vision sensor to decisions. However, they do not consider contextual temporal information when making decisions. In contrast, imitation learning methods implicitly learn the temporal contextual features as their models treats the outputs as a sequence of actions. This characteristic makes imitation learning more suitable for this decision-making task.

In addition, Table 4 shows the AV/ADAS decision timing based on four groups of learning approaches. It is illustrated that the decision timing for AV/ADAS to enter a roundabout from deep learning methods and proposed ILBDM system are faster than the traditional machine learning algorithms and GBIPA-SC-NR. The fastest decision time is based on DQfD-CNN in proposed ILBDM class with the number of 0.1035(s) which is 1.1873 (s) faster than ANN in GBIPA-SC-NR class. Therefore, Table 4 illustrates the proposed ILBDM approach provides remarkable performance by considering both the decision accuracy and inference timing.

5 Discussion

According to the literatures, it is found that deep imitation learning based method combines the advantages from both the supervised learning and reinforcement learning based frameworks. Therefore, in this work, we propose an imitation learning based system -ILBDM and prove that it outperforms all the supervised learning methods to accomplish the decision-making task when joining the roundabouts. The positive impacts from our work can be summarized in four folders: first, high-quality data were collected for the experiments. In particular, a significant amount of real-world data containing roughly 50 roundabouts were recorded in different time frames at different days. The data reflect the real traffic conditions, thus increasing the possibility of applying the techniques in reality. It is noticed that this is the first large real world dataset for solving this challenging task. Although the data in [34] contain 50 different roundabouts which is comparable to our work, they were generated by using a driving simulator; secondly, the proposed ILBDM can effectively make decisions, thus showing the capability of applying in real-world. The accuracy rate of the proposed system based on the DQfD-CNN achieves 96.21% which are significantly better than the other state-of-the-art algorithms; thirdly, the proposed ILBDM can work with cars moving in different speed situations. ILBDM provides vehicle detection and optical flow modules to determine the approaching car's speed and positions. It means that the speed and distance of the oncoming cars can be tracked, measured and calculated as effective observational states; fourthly, proposed ILBDM system improvs our previous work of the grid based method, GBIPA-SC-NR. Compared with GBIPA-SC-NR, both of the accuracy and inference timing are improved significantly.

Real-time processing is vital for an autonomous decision-making model. In our work, the total execution time for planning from one frame is 0.43 seconds (0.2 seconds for optical flow extraction, 0.12 seconds for car detection and 0.11 seconds for the action network). As the purpose of the work is not developing a fully autonomous decision-making model to replace human driver but a decision augmentation tool to facilitate safe behavior of human driver, we believe the performance is acceptable. While, in our future work, we will further investigate to replace the optical flow estimation module by 3D-CNN and speed up the processing to achieve real-time performance which could potentially serve as a fully autonomous driving system. Furthermore, since the ILBDM can learn from individuals' driving styles and behaviors, the system has potential to model different types of human-like decisions. In the future work, we will collect more training data based on different driver styles so that driver's behaviors in reality can be learned and simulated.

6 Conclusion

In this paper, we present an imitation learning based decision making system named ILBDM for an AV/ADAS to make the most suitable decisions to join roundabouts timely and safely. The ILBDM system have an effective observation extraction pipeline which include the vehicle detection based on the Faster R-CNN and motion feature extraction from optical flow. It trains deep policy networks based on several popular backbone networks, inlcuding CNN, VGG-16 and ResNet-18 to recommend actions to maximize the cumulative returns from a sequence of decision-makings. The learned network in the proposed ILBDM system were evaluated on 130 videos from real world. The results demonstrates that the proposed ILBDM system can be applied to effectively help AV/ADAS make the most suitable decisions when approaching a roundabout. Furthermore, it is believed that the proposed framework has potentials to be adapted and deployed in other high-level autonomous vehicle control tasks when collecting corresponding data.

Declarations

Competing interests All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version. This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Abdulkader MMS, Gajpal Y, ElMekkawy TY (2015) Hybridized ant colony algorithm for the multi compartment vehicle routing problem. Appl Soft Comput 37:196–203
- 2. Adler B, Xiao J, Zhang J (2014) Autonomous exploration of urban environments using unmanned aerial vehicles. J Field Robot 31(6):912–939
- Aeberhard M, Rauch S, Bahram M, Tanzmeister G, Thomas J, Pilat Y, Homm F, Huber W, Kaempchen N (2015) Experience, results and lessons learned from automated driving on germany's highways. IEEE Intell Transp Syst Mag 7(1):42–57
- Alom MdZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MstS, Van Esesn BC, Awwal AAS, Asari VK (2018) The history began from alexnet:, A comprehensive survey on deep learning approaches. arXiv:1803.01164
- Anggodo YP, Ariyani AK, Ardi MK, Mahmudy WF (2017) Optimization of multi-trip vehicle routing problem with time windows using genetic algorithm. J Environ Eng Sustain Technol 3(2):92–97
- Ballester P, Araujo RM (2016) On the performance of googlenet and alexnet applied to sketches. In: Thirtieth AAAI Conference on Artificial Intelligence
- 7. Bansal M, Krizhevsky A, Ogale A (2018) Chauffeurnet:, Learning to drive by imitating the best and synthesizing the worst. arXiv:1812.03079
- Bojarski M, Yeres P, Choromanska A, Choromanski K, Firner B, Jackel L, Muller U (2017) Explaining how a deep neural network trained with end-to-end learning steers a car. arXiv:1704.07911
- Brummelen JV, O'Brien Ma, Gruyer D, Najjaran H (2018) Autonomous vehicle perception: the technology of today and tomorrow. Transportation research part C:, emerging technologies 89:384–406
- Chen AL, Yang GK, Wu ZM (2006) Hybrid discrete particle swarm optimization algorithm for capacitated vehicle routing problem. J Zhejiang Univ Sci A 7(4):607–614
- Codevilla F, Miiller M, López A, Koltun V, Dosovitskiy A (2018) End-to-end driving via conditional imitation learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp 1–9. IEEE
- EL Sallab A, Abdou M, Perot E, Yogamani S (2017) Deep reinforcement learning framework for autonomous driving. Electronic Imaging 2017(19):70–76
- El Hamdani S, Benamar N (2017) A comprehensive study of intelligent transportation system architectures for road congestion avoidance. In: International Symposium on Ubiquitous Networking, pp 95–106. Springer
- García Cuenca L, Puertas E, Fernandez Andrés J, Aliane N (2019) Autonomous driving in roundabout maneuvers using reinforcement learning with q-learning. Electronics 8(12):1536
- García Cuenca L, Sanchez-Soriano J, Puertas E, Fernandez Andrés J, Aliane N (2019) Machine learning techniques for undertaking roundabouts in autonomous driving. Sensors 19(10):2386

- Gritschneder F, Hatzelmann P, Thom M, Kunz F, Dietmayer K (2016) Adaptive learning based on guided exploration for decision making at roundabouts. In: 2016 IEEE Intelligent Vehicles Symposium (IV), pp 433–440. IEEE
- Guerrero-Ibáñez J, Zeadally S, Contreras-Castillo J (2018) Sensor technologies for intelligent transportation systems. Sensors 18(4):1212
- Hassannejad H, Medici P, Cardarelli E, Cerri P (2015) Detection of moving objects in roundabouts based on a monocular system. Expert Syst Appl 42(9):4167–4176
- Hawke J, Shen S, Gurau C, Sharma S, Reda D, Nikolov N, Mazur P, Micklethwaite S, Griffiths N, Shah A et al (2019) Urban driving with conditional imitation learning. arXiv:1912.00177
- 20. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition
- 21. Hecht J (2018) Lidar for self-driving cars. Opt Photonics News 29(1):26-33
- 22. Hester T, Vecerik M, Pietquin O, Lanctot M, Schaul T, Piot B, Horgan D, Quan J, Sendonaris A, Osband I et al (2018) Deep q-learning from demonstrations. In: Thirty-Second AAAI Conference on Artificial Intelligence
- Hussein A, Gaber MM, Elyan E, Jayne C (2017) Imitation learning: a survey of learning methods. ACM Computing Surveys (CSUR) 50(2):1–35
- Indu S, Gupta M, Bhattacharyya A (2011) Vehicle tracking and speed estimation using optical flow method. Int J Eng Sci Technol 3(1):429–434
- Jiangtao Li J, Zhang L, Dai F, Zhang Y, Meng X, Shen J (2018) Secure intelligent traffic light control using fog computing. Future Gener Comput Syst 78:817–824
- Jones M, Bontrager P, Paszkowicz S, Wheller P (2018) System and method for configuring autonomous vehicle responses based on a driver profile, August 21. US Patent 10,054,944
- Jurewicz C, Sobhani A, Chau P, Woolley J, Brodie C (2017) Understanding and improving safe system intersection performance. Safe System performance on Intersections, Austroads APR556-17 https://austroads.com.au/publications/road-design/ap-r556-17
- Kebria PM, Khosravi A, Salaken SM, Nahavandi S (2019) Deep imitation learning for autonomous vehicles based on convolutional neural networks. IEEE/CAA Journal of Automatica Sinica 7(1):82–95
- Kennedy JV, House C, Ride NM (2008) The uk standards for roundabouts and mini-roundabouts. In: National roundabout conference, TRB, Kansas City, Missouri, USA, pp 18–21
- Lin S, Cai L, Lin X, Ji R (2016) Masked face detection via a modified lenet. Neurocomputing 218:197– 202
- Liu Wei, Anguelov Dragomir, Erhan Dumitru, Szegedy Christian, Reed Scott, Cheng-Yang Fu, Berg AlexanderC (2016) Ssd: Single shot multibox detector
- 32. Liu C et al (2009) Beyond pixels: exploring new representations and applications for motion analysis. PhD thesis Massachusetts Institute of Technology
- 33. Masood Sarfaraz, Rai Abhinav, Aggarwal Aakash, Doja MohammadNajmud, Ahmad Musheer (2018) Detecting distraction of drivers using convolutional neural network Pattern Recognition Letters
- Muffert M, Milbich T, Pfeiffer D, Franke U (2012) May i enter the roundabout? a time-to-contact computation based on stereo-vision. In: 2012 IEEE Intelligent Vehicles Symposium, pp 565–570. IEEE
- Muffert Maximilian, Pfeiffer David, Franke Uwe (2013) A stereo-vision based object tracking approach at roundabouts. IEEE Intell Transp Syst Mag 5(2):22–32
- Noi PT, Kappas M (2018) Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery. Sensors 18(1):18
- Okumura B, James MR, Kanzawa Y, Derry M, Sakai K, Nishi T, Prokhorov D (2016) Challenges in perception and decision making for intelligent automotive vehicles: a case study. IEEE Trans Intell Veh 1(1):20–32
- Pan Y, Cheng CA, Saigol K, Lee K, Yan X, Theodorou E, Boots B (2017) Agile autonomous driving using end-to-end deep imitation learning. arXiv:1709.07174
- Pomerleau DA (1989) Alvinn: An autonomous land vehicle in a neural network. In: Advances in neural information processing systems, pp 305–313
- 40. Qi Wang, Gao Junyu, Yuan (2017) Embedding structured contour and location prior in siamesed fully convolutional networks for road detection. IEEE Trans Intell Transp Syst 19(1):230–241
- 41. Rampasek L, Goldenberg A (2016) Tensorflow: Biology's gateway to deep learning? Cell systems 2(1):12–14
- 42. Ren S, He K, Girshick R, Sun J (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. In: Advances in neural information processing systems, pp 91–99
- Rhinehart C, McAllister R, Levine S (2018) Deep imitative models for flexible inference, planning, and control. arXiv:1810.06544
- 44. Rodrigues M, McGordon A, Gest G, Marco J (2018) Autonomous navigation in interaction-based environments—a case of non-signalized roundabouts. IEEE Trans Intell Veh 3(4):425–438

- 45. Sun L (2016) Resnet on tiny imagenet. Submitted on 14
- 46. Sun S, Shetty A, Gurunath N, Bhirangi R (2019) Improving dqn and trpo with hierarchical metacontrollers
- 47. Vallati M, Magazzeni D, De Schutter B, Chrpa L, McCluskey TL (2016) Efficient macroscopic urban traffic models for reducing congestion: a pddl+ planning approach. In: Thirtieth AAAI Conference on Artificial Intelligence
- Wang T, Chang DE (2019) Improved reinforcement learning through imitation learning pretraining towards image-based autonomous driving. arXiv:1907.06838
- Wang W, Meng Q, Chung PWH (2018) Camera based decision making at roundabouts for autonomous vehicles. In: 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), pp 1460–1465 IEEE
- 50. Wang W, Nguyen QA, Chung PWH, Meng Q (2018) Multi-cameras based decision making at miniroundabouts for autonomous vehicles. Poster Papers, pp 75
- Wang W, Nguyen QA, Ma W, Wei J, Chung PWH, Meng Q (2019) Multi-grid based decision making at roundabout for autonomous vehicles. In: 2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES), pp 1–6. IEEE
- 52. Williams S, Relton SD, Fang H, Alty J, Qahwaji R, Graham CD, Wong DC (2020) Supervised classification of bradykinesia in parkinson's disease from smartphone videos. Artif Intell Med 110:101966
- 53. Williams G, Wagener N, Goldfain B, Drews P, Rehg JM, Boots B, Theodorou EA (2017) Information theoretic mpc for model-based reinforcement learning. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp 1714–1721. IEEE
- Wolf P, Hubschneider C, Weber M, Bauer A, Härtl J, Dürr F, Marius Zöllner J (2017) Learning how to drive in a real world simulation with deep q-networks. In: 2017 IEEE Intelligent Vehicles Symposium (IV), pp 244–250. IEEE
- 55. Wong DC, Relton SD, Fang H, Qhawaji R, Graham CD, Alty J, Williams S (2019) Supervised classification of bradykinesia for parkinson's disease diagnosis from smartphone videos. In: 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS), pp 32–37. IEEE
- 56. Wu N, Rathod V (2017) Tensorflow detection model zoo
- Yadav N, Binay U (2017) Comparative study of object detection algorithms. International Research Journal of Engineering and Technology (IRJET) 4(11):586–591
- Yi Z, Newsam S (2017) Densenet for dense flow. In: 2017 IEEE international conference on image processing (ICIP), pp 790–794. IEEE
- Zeng W, Luo W, Suo S, Sadat A, Yang B, Casas S, Urtasun R (2019) End-to-end interpretable neural motion planner. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 8660–8669
- Zimmermann RS, Siems JN (2019) Faster training of mask r-cnn by focusing on instance boundaries. Computer Vision and Image Understandin 188:102795

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.