Robust Imitative Planning: Planning from Demonstrations Under Uncertainty

Panagiotis Tigas[1,*] **Angelos Filos**[1,*] **Rowan McAllister**[2], **Nicholas Rhinehart**[2,3]
Sergey Levine[2], **Yarin Gal**[1]

[1]University of Oxford [2]University of California, Berkeley [3]Carnegie Mellon University
{panagiotis.tigkas, angelos.filos}@cs.ox.ac.uk

Abstract

Learning from expert demonstrations is an attractive framework for sequential decision-making in safety-critical domains such as autonomous driving, where trial and error learning has no safety guarantees during training. However, naïve use of imitation learning can fail by extrapolating incorrectly to unfamiliar situations, resulting in arbitrary model outputs and dangerous outcomes. This is especially true for high capacity parametric models such as deep neural networks, for processing high-dimensional observations from cameras or LIDAR. Instead, we model expert behaviour with a model able to capture uncertainty about previously unseen scenarios, as well as inherent stochasticity in expert demonstrations. We propose a framework for planning under epistemic uncertainty and also provide a practical realisation, called *robust imitative planning* (RIP), using an ensemble of deep neural density estimators. We demonstrate online robustness to out-of-training-distribution scenarios on the CARLA autonomous driving simulator, improving over other probabilistic imitation learning models and reducing the total number of hazardous events while improving runtime to real-time using a trajectory library.

1 Introduction

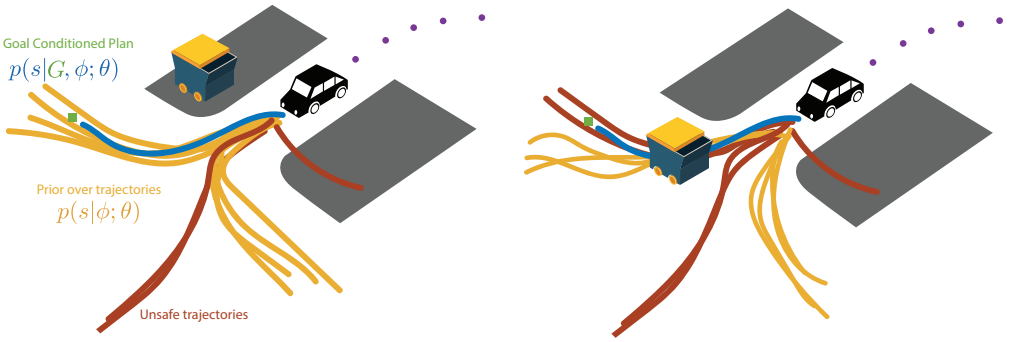
Robustness and safety are critical challenges for mobile robots, especially in the domain of autonomous driving. Learning-based approaches can enable mobile robots and autonomous vehicles to respond intelligently in a wide range of situations but does not by itself resolve the challenges of robustness and safety: a learning-based system may perform well in domains that resemble those it was trained in, but can fail in unpredictable ways in novel situations (i.e. out-of-training-distribution). Generative models can provide a measure of their uncertainty in different situations, but robustness in novel environments requires estimating *epistemic* uncertainty (e.g., "have I been in this state before?"), where conventional density estimation models only capture *aleatoric* uncertainty (e.g., "what's the frequency of times I ended up in this state?").

Despite model-free reinforcement learning's recent successes in video, board games and controlled robotics settings, trial-and-error approaches are either too unsafe to use for safety-critical applications, such as autonomous driving, or the specification of a reward function is as hard as solving the original control problem in the first place. On the other hand, learning to imitate expert behaviour from demonstrations given continuous actions and high-dimensional observations is an attractive tool for control, since a model mimicking expert demonstrations can simply learn to stay in "safe", expert-like parts of the state space and no explicit reward function has to be specified.

However, approaches based on behavioural cloning suffer from state distribution shift (i.e. covariate shift) [35], where high capacity parametric models (e.g. neural networks) usually fail to generalise,

[*]The authors contributed equally.

Machine Learning for Autonomous Driving Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.



(a) Common example (in-distribution) (b) Novel scenario (covariate shift)

Figure 1: Learning from demonstrations can be very effective for learning policies in safety-critical domains. However, in novel scenarios the model's reliability degrades radically, leading to catastrophic outcomes. The imitative model is successful in estimating the density in an in-distribution scene (a) but fails catastrophically when a novel (e.g. trash bin on the street) context is experienced.

and instead extrapolate confidently yet incorrectly, resulting in arbitrary outputs and dangerous outcomes [28], as depicted in Figure 1. Bayesian neural network techniques [17, 5, 14, 23] can be used to estimate epistemic uncertainty for deep neural network models [13]. Nonetheless, standard imitation learning is poorly suited for fully utilising these epistemic uncertainty measures: although we can detect *when* a model trained with imitation learning is uncertain about the best action, this model does not necessarily provide us with a good answer to *what* it should do in that situation (e.g. short of stopping the vehicle). What we require is a model that can not only report on its uncertainty but also provide a mechanism for taking low-risk actions that are likely to recover in uncertain situations.

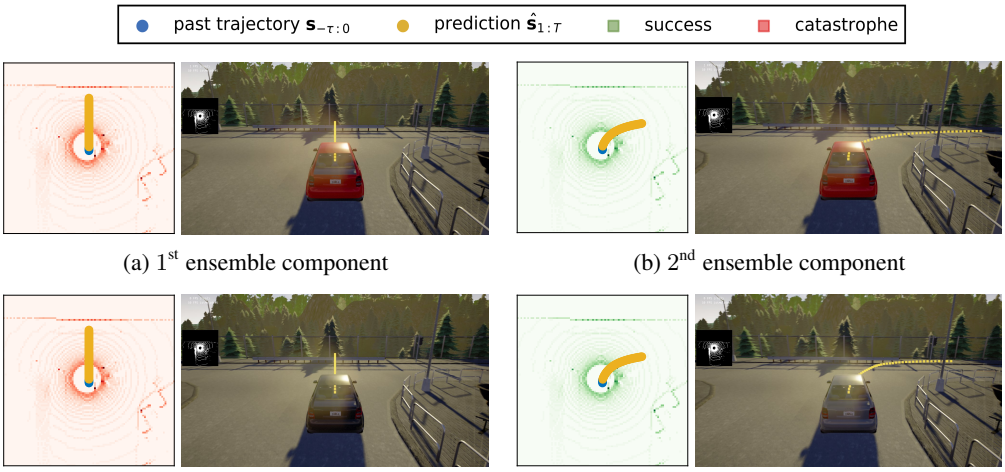
In this work, we build on *deep imitative models*, an approach that combines generative modelling from demonstration data with planning. Deep imitative models [33] are *context-conditioned density estimators* use for plannin. likelihood-based models trained on expert demonstrations that reason about the probability that planned trajectories are expert-like when trying to accomplish new tasks at test time. During planning, candidate trajectories are scored based on their likelihood under the imitative models, and the mode (i.e. most likely trajectory for a given context) is followed. However, the quality of the plans depends highly on the density estimates, which can be unreliable when conditioning on out-of-distribution contexts.

We address this shortcoming by capturing epistemic uncertainty of the density estimatior via deep ensembles' uncertainty [23], to provide for control in novel and unexpected situations. Our framework, which we call robust imitative planning (RIP), uses demonstration data to learn density models over human-like driving, and then estimates its uncertainty *about these densities* using an ensemble of imitative models. When a trajectory that was never seen before is selected, the model's high epistemic uncertainty pushes us away from it. During planning, the disagreement between the most probable trajectories under the ensemble of imitative models is used to inform planning. A unified framework of epistemic uncertainty-aware planning objectives, called *robust imitative planning* (RIP), is proposed that principally integrates all sources of uncertainty in sequential decision making.

Moreover, in order to solve the RIP objectives an efficient search method based on trajectory libraries [26] is used, reducing the planning time by a factor of 400, enabling real-time use of deep imitative models, which used to be prohibitively slow for deployment on real vehicles. Finally, our method outperforms vanilla imitative modelling and behaviour cloning in a variety of novel driving scenarios, in online CARLA simulation setups, with a focus on *out-of-training-distribution* scenes.

2 Planning from Demonstrations

Given expert demonstrations, explicit policies can be trained to imitate the expert, a method often termed behavioural cloning [1]. However, this approach lacks flexibility in deployment, since non-trivial changes to the data collection procedure should be made to allow for goal-conditioning policies. On the other hand, generative models of expert behaviour (e.g., density estimators) can be more



(a) 1st ensemble component

(b) 2nd ensemble component

(c) Optimistic imitative planning

(d) Robust imitative planning (our method)

Figure 2: Qualitative comparison of planning objectives on an out-of-training-distribution example.

flexible during deployment, as they can be used as a proxy to score how likely plans were to come from an expert demonstrator. Goal-conditioning can be trivially done under this framework. In this section, we briefly review prior work [33] that follows this paradigm and then present our robust imitative planning framework in the next section.

2.1 Problem Setup & Notation

Our method and Rhinehart et al. [33] both assume access to a dataset of time-profiled expert trajectories of sequential positions, synced with high-dimensional observations of the corresponding scenes. The high-dimensional observations and partial trajectories are used to perform context-conditioned density estimation of the distribution over future expert trajectories. The model's ability to produce an exact density estimate of arbitrary future trajectories is critical to our approach. This property enables the model to score the "expertness" of any plan of future positions.

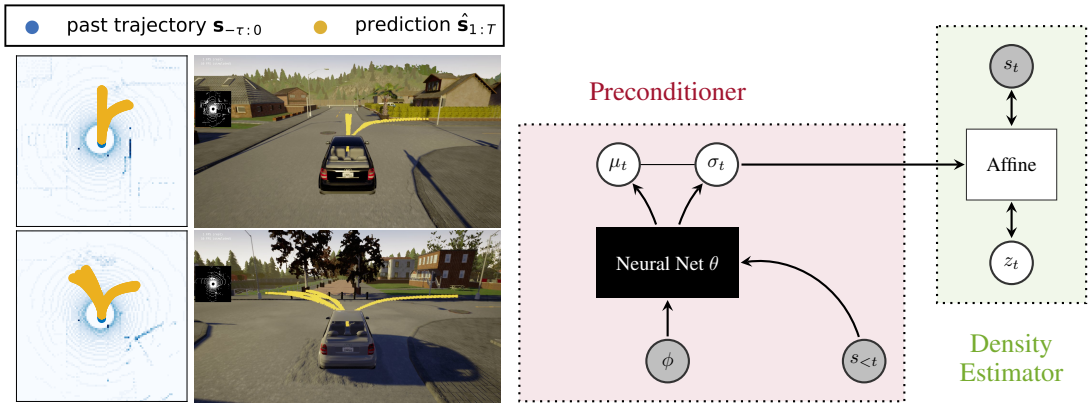
Let $\mathbf{s}_t \in \mathbb{R}^D$ denote the agent's state ($xy-$coordinates) at time step t, and $t = 0$ define present time. Contextual information is given by $\phi \triangleq \{\mathbf{s}_{-\tau:0}, \boldsymbol{\chi}\}$, where τ is the number of past states, and $\boldsymbol{\chi}$ is a LIDAR observation at time $t = 0$. Variables are marked in bold, functions are not bold, and random variables are capitalised. Variables without time subscripts refers to their value at *all* future time steps up to horizon T, e.g. $\mathbf{S} \triangleq \mathbf{S}_{1:T} \in \mathbb{R}^{T\times D}$. The probability density function of a random variable $\mathbf{S}$ is given by $p(\mathbf{S})$, and the corresponding probability density at a specific value $\mathbf{s}$ as $p(\mathbf{s}) \triangleq p(\mathbf{S} = \mathbf{s})$.

2.2 Modelling the Expert Demonstrators

Our method's main requirement is the ability to perform density estimation of the conditional distribution over future expert trajectories. Following Rhinehart et al. [33], we use ReparameteRized Pushforward Policies (R2P2) to implement this distribution, however, *any* density estimation method with exact likelihood inference could be used. R2P2 models an expert driver by fitting a multimodal state-trajectory distributions $q(\mathbf{S}|\phi;\theta)$ to expert trajectory data $\mathcal{D} = \{\phi^i, \mathbf{s}^i\}_{i=1}^N$, drawn from an unknown distribution $\mathbf{s}^i \sim p(\mathbf{S}|\phi;\theta)$. While R2P2 captures stochasticity in expert behaviour, e.g., choosing to turn either left or right at an intersection (see Fig 3a), it *fails to capture epistemic uncertainty in the model's density estimate*. With only a point-estimate of model parameters, the model is oblivious to whether it is currently operating in unfamiliar scenes (outside the distribution of those seen during training).

2.3 Planning with a Model of Expert Drivers

The model $q(\mathbf{s}|\phi;\theta)$, defines the backbone of the control framework used in [33] and our method. Before we describe how our method uses q, we first describe how q is used. Deep imitative models



(a) Samples from generative model, R2P2 [32]. (b) Computation graph.

Figure 3: Multi-modal imitative model $q(S|\phi;\theta)$.

(IM) [33] use q to plan to goal states, using state trajectories that have high likelihood w.r.t. the expert model $q(\mathbf{S}|\phi;\theta)$. The imitative planning objective is the log posterior probability of a state trajectory, conditioned on satisfying some goal $\mathcal{G}$:

$$\mathbf{s}^{\mathcal{G}}_{\text{IM}} \triangleq \underset{\mathbf{s}}{\text{argmax}} \log \underbrace{p(\mathbf{s}|\mathcal{G},\phi;\theta)}_{\text{imitation posterior}} = \underset{\mathbf{s}}{\text{argmax}} \log \underbrace{q(\mathbf{s}|\phi;\theta)}_{\text{imitation prior}} + \log \underbrace{p(\mathcal{G}|\mathbf{s})}_{\text{goal-likelihood}}. \tag{1}$$

The planned trajectory $\mathbf{s}^{\mathcal{G}}_{\text{IM}}$ is the maximum a posteriori probability (MAP) estimate of how an expert would drive to the goal, capturing any inherent *aleatoric* stochasticity of the human behaviour (e.g., multi-modalities).

However, IM only uses a point-estimate of θ, thus $q(\mathbf{s}|\phi;\theta)$ does not quantify model (i.e. *epistemic*) uncertainty. This is especially problematic when estimating what an expert would or would not do in *unfamiliar* scenes. If ϕ is unfamiliar, the expert model $q(\mathbf{s}|\phi;\theta)$ has *undetermined* output since the pre-conditioner network in Figure 3b may fail to generalise to the novel scenes, and thus the plans produced by Eqn. (1) can confidently lead into a crash scenario. Thus, IM cannot assess how reliable its planning is in unfamiliar scenes. Therefore, we need both 1) a model that captures epistemic uncertainty and 2) a planning objective that takes both aleatoric and epistemic uncertainty into account.

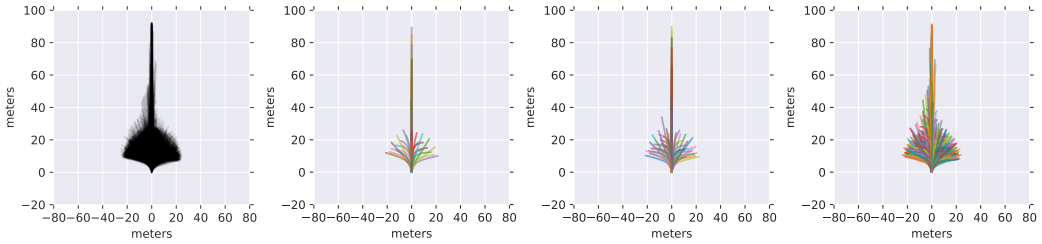
3 Robust Imitative Planning

We place a prior distribution $p(\theta)$ over possible models θ, which induces a distribution over the density models models $q(\mathbf{s};\theta)$. After observing data $\mathcal{D}$, this distribution over density models has a posterior $p(\theta|\mathcal{D})$. Decision-making under the posterior $p(\theta|\mathcal{D})$ can be formulated as optimisation [2] of the generic objective

$$\begin{aligned} s^{\mathcal{G}}_{\text{RIP}} &\triangleq \underset{\mathbf{s}}{\text{argmax}} \overbrace{\underset{\theta\in\text{supp}(p(\theta|\mathcal{D}))}{\square}}^{\text{aggregation operator}} \log \underbrace{p(\mathbf{s}|\mathcal{G},\phi;\theta)}_{\text{imitation posterior}} \\ &= \underset{\mathbf{s}}{\text{argmax}} \left[\underset{\theta\in\text{supp}(p(\theta|\mathcal{D}))}{\square} \log \underbrace{q(\mathbf{s}|\phi;\theta)}_{\text{imitation prior}} \right] + \log \underbrace{p(\mathcal{G}|\mathbf{s})}_{\text{goal-likelihood}} . \end{aligned} \tag{2}$$

where $\square$ is an operator applied on the posterior $p(\theta|\mathcal{D})$.

The original imitative models objective in Eqn. (1) is a particular instance of the more general family of objectives described by Eqn. (2), where operator $\square$ selects a single θ_0 from the posterior and then follows that θ_0 for the whole trajectory. However, as discussed in Section 2, this approach ignores the



(a) Trajectories (b) $K = 64$ (c) $K = 128$ (d) $K = 1024$

Figure 4: Our trajectory library from CARLA's autopilot demonstrations, 4 seconds.

structure of the posterior distribution over models and hence the induced posterior over trajectories conditioned on a context. Therefore it is oblivious to the uncertainty and prone to fail in certain cases. Table 1 enlists the different objectives-operators we examined.

A principled means to capture epistemic uncertainty is with Bayesian inference. However, evaluating $p(\theta|\mathcal{D})$ with exact inference is intractable [29]. Approximate inference methods [17, 5, 14, 18] have been introduced that can efficiently capture epistemic uncertainty. In our implementation, we use ensembles of deep models as an approximation to epistemic uncertainty estimation, as done by Lakshminarayanan et al. [23], Chua et al. [7]. We consider ensembles of K models, using θ_k to refer to the parameters of our k^{th} model q_k.

3.1 Planning with Trajectory Libraries

In the absence of scalable global optimisers, we search the trajectory space in Eqn. (2) by looking only at a trajectory library [26], $\mathcal{T}_{\mathbf{S}}$, a set of fixed, finite trajectories from which we select the one which corresponds to our objective. Hence we optimise the modified objective:

$$\mathbf{s}^{\mathcal{G}}_{\text{RIP}} \approx \underset{\mathbf{s}\in\mathcal{T}_{\mathbf{S}}}{\text{argmax}} \square_{k\in[K]} \log q(\mathbf{s}|\phi;\theta_k) + \log p(\mathcal{G}|\mathbf{s}) \tag{3}$$

Solving for Eqn. (3) results in $\times 400$ improvement in runtime compared to the gradient descent alternative, allowing for real-time deployment of imitative models, which is otherwise orders of magnitude slower.

In this work, we perform K-means clustering of the expert plan's from the training distribution and keep 128 of the centroids, as illustrated in Figure 4.

4 Related Work

Rational decisions under uncertainty are usually formulated as optimisation of expected utility, where the expectation is taken w.r.t. subjective model-uncertainty or process stochasticity [39, 10, 2]. Instead of expected utility, penalties on cost variance can be used for robustness [10]. Similarly, based on this paradigm, we capture both types of uncertainty (i.e. epistemic and aleatoric) and propose objectives

Table 1: Summary of planning objectives: ours methods (grey) and baselines, using shorthand: $q_k = q(\mathbf{s}|\phi;\theta_k)$ for trajectory density (under model k).

Methods	**Aggregation Operator** $\square$	Interpretation
Imitative Models	$\log q_{k=1}$	Sample
Optimistic	$\max_k \log q_k$	Max
Soft Optimistic	$\log \sum_k q_k$	Soft Max
Robust Imitative Planning - Epistemic Uncertainty-Aware (ours)		
Bayes' Optimal	$\sum_k \log q_k$	Model Average
Soft Pessimistic	$-\log \sum_k q_k^{-1}$	Soft Min
Pessimistic	$\min_k \log q_k$	Min

that take them into account. We also implement a baseline in Section 5 that optimises the expected utility (a.k.a. *Bayes' optimal* plan.

Robust optimisation [40, 37, 38, 16, 11, 3] aims to improve on the *worst-case scenarios* in the face of uncertainty - extensively studied in signal processing [21] and control [4].

Imitation learning uses expert supervision to learn desired behaviour [30]. Behaviour Cloning (BC) is a common approach, mapping currently observational inputs to future expert actions [31]. A body of previous work has explored BC for autonomous driving in the CARLA simulator [8, 9, 24, 25, 36, 33]. The importance of risk-averse policies for autonomous driving has been highlighted already by Choi et al. [6], Ghosh et al. [15], Lötjens et al. [27]. In contrast to BC and most imitation learning approaches, the imitative model provides a *posterior distribution over paths*, while most of the prior methods either provide a distribution over a particular event prediction (e.g., collision [20, 27]) or just over actions [19, 22].

5 Experiments

We are interested in learning to drive from finite expert demonstrations, and being robust to out-of-training-distribution scenarios. The goal of our experimental evaluation is to answer the following questions: (1) How the quantification of epistemic uncertainty impacts the out-of-training-distribution performance of imitative models? (2) What is the best way to inform decision-making under uncertainty in terms of robustness and safety? (3) What are the limits of the RIP objective?

5.1 Experimental Setup

Expert Demonstrations. We generate a realistic dataset of expert demonstrations using the CARLA simulator expert driver bot (i.e. autopilot) [12], including other cars and pedestrians, in `Town 1`. Our model uses as context information, ϕ, the past $\tau = 3$ car positions and the LIDAR point-cloud, $\boldsymbol{\chi}$, following the pre-processing by Rhinehart et al. [33, 34]. Our dataset consists of $80,000$ scenes, gathered at 10Hz out of which we used 80% for training, 10% for testing and 10% for validation.

Metrics. To evaluate safety we track off-road events on the goal-conditioned plans generated by our models. The goal is set as the last position of the trajectory followed by the autopilot (ground truth trajectory). To assess if the generated plan was off-road, we used the segmentation maps that are provided by CARLA simulator. The segmentation map of each scene was stored in the dataset which allowed us to evaluate the quality of the plan at test time, without interacting with the simulator.

All methods are tested on the same scenarios on `Town 2-5` and on a fixed number of episodes.

Out-of-distribution Scenes. For testing out-of-distribution scenes, we gathered data from `Town 2-5` which consist of street topologies and obstacles that are significantly different from the training scenes.

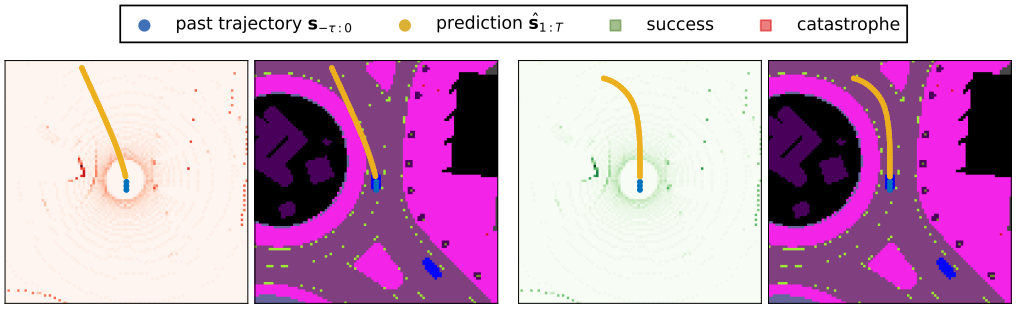
Challenging Scenes. Current approaches [12, 8, 9, 24, 25, 36] perform almost perfect on straight paths in the absence of other vehicles and pedestrians. Therefore, we focus our attention on the more interesting scenarios, like roundabouts in order to test robustness on challenging out-of-distribution scenarios. Note that our training scenes (`Town 1`) do not contain roundabouts.

5.2 Results

Table 2 highlights the benefit of taking epistemic uncertainty into consideration during planning since in out-of-training-distribution, RIP and BOP demonstrate improved performance. The OIP variant leads to catastrophic outcomes, even in in-sample scenarios, suggesting that optimism in the face of uncertainty can be dangerous and hence non-robust. In particular, the in-sample performance of the epistemic uncertainty agnostic IM [33] method is similar to RIP, an observation that reinforces the argument that despite IM's capacity to fit the expert demonstrations, it can confidently extrapolate in novel situations and lead to catastrophes. In Table 2 we also illustrate that in scenes that are significantly *out-of-distribution* (`Roundabouts` column), IM and OIP are outperformed by the RIP and BOP objectives.

Table 2: Quantitative results on in-distribution (`Town 1`) and out-of-distribution (`Town 2-5` and `Roundabout`) scenarios.

	% **Hazards (i.e. off-road)** ↓					
Methods	`Town 1`	`Town 2`	`Town 3`	`Town 4`	`Town 5`	`Roundabouts`
Baselines						
Vanilla Imitative Model [33]	10.61±10.61	11.71±11.71	16.50±16.50	1.18±1.18	14.94±14.94	68.01±68.01
Optimistic	10.03±10.03	10.86±10.86	16.59±16.59	1.03±1.03	11.62±11.62	65.68±65.68
Soft Optimistic	10.11±10.11	11.03±11.03	16.69±16.69	1.14±1.14	11.52±11.52	65.90±65.90
Robust Imitative Planning - Epistemic Uncertainty-Aware (ours)						
Bayes' Optimal	**5.27±5.27**	**5.92±5.92**	**12.04±12.04**	**0.47±0.47**	**6.84±6.84**	**32.11±32.11**
Soft Pessimistic	9.13±9.13	9.83±9.83	15.12±15.12	0.94±0.94	11.83±11.83	**31.44±31.44**
Pessimistic	9.08±9.08	9.94±9.94	15.61±15.61	0.97±0.97	12.03±12.03	**32.01±32.01**



(a) Deep Imitative Model [33] (b) Robust imitative planning (our method)

Figure 5: Qualitative comparison of planning objectives on an out-of-training-distribution example. Left-hand-side of each example is the verhead LIDAR view and right-hand-side is the bird-eye-view of the segmentation map.

To assess the effect of the proposed objective (RIP) we also examine what are the plans our objective proposes in an extreme out-of-distribution roundabout scene (never encountered in the training dataset). In Figure 5 we can see that Deep Imitative Models can suggest plans that are hazardous, however, RIP and BOP objectives, after taking a more pessimistic approach, suggest a plan that is safe. Additionally, in figure.f̃ig:disagreement we examine another scene where the car is on an out-of-distribution scene of `Town 1`. Different assemble models suggest different goal-conditioned plans, illustrating the effect of the epistemic uncertainty. Following the most confident of such models can prove catastrophic, since there are no guarantees in such out-of-distribution scenes, that the high confidence is, in fact, a correct prediction of the model. By taking into consideration the disagreement between the models, RIP and BOP can correctly suggest a plan that satisfies the goal and is safer (Figure 5.d)

6 Conclusion

In this work, we propose different objectives for planning from demonstrations under *uncertainty*. Our framework builds on top of *Deep Imitative Models* [33], whose plans can be risky in scenes that are *out-of-distribution*. We demonstrated examples where *Deep Imitative Models* can fail and how we can reduce hazards on such cases by using an ensemble of density estimators and aggregate operators over the models' outputs, that take into consideration their epistemic uncertainty, in order to safely plan under uncertainty.

References

[1] Christopher G Atkeson and Stefan Schaal. Robot learning from demonstration. In *ICML*, volume 97, pages 12–20. Citeseer, 1997.

[2] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[3] Dimitris Bertsimas, Vishal Gupta, and Nathan Kallus. Data-driven robust optimization. *Mathematical Programming*, 167(2):235–292, 2018.

[4] Shankar P Bhattacharyya and Lee H Keel. Robust control: the parametric approach. In *Advances in Control Education*, pages 49–52. Elsevier, 1995.

[5] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. *arXiv preprint arXiv:1505.05424*, 2015.

[6] Sungjoon Choi, Kyungjae Lee, Sungbin Lim, and Songhwai Oh. Uncertainty-aware learning from demonstration using mixture density networks with sampling-free variance modeling. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6915–6922. IEEE, 2018.

[7] Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In *Advances in Neural Information Processing Systems*, pages 4754–4765, 2018.

[8] Felipe Codevilla, Matthias Miiller, Antonio López, Vladlen Koltun, and Alexey Dosovitskiy. End-to-end driving via conditional imitation learning. In *International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

[9] Felipe Codevilla, Eder Santana, Antonio M López, and Adrien Gaidon. Exploring the limitations of behavior cloning for autonomous driving. *arXiv preprint arXiv:1904.08980*, 2019.

[10] Marc Peter Deisenroth. *Efficient reinforcement learning using Gaussian processes*, volume 9. KIT Scientific Publishing, 2010.

[11] Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations research*, 58(3):595–612, 2010.

[12] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In *Conference on Robot Learning (CoRL)*, pages 1–16, 2017.

[13] Yarin Gal. *Uncertainty in deep learning*. PhD thesis, University of Cambridge, 2016.

[14] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning*, pages 1050–1059, 2016.

[15] Shromona Ghosh, Felix Berkenkamp, Gireeja Ranade, Shaz Qadeer, and Ashish Kapoor. Verifying controllers against adversarial examples with bayesian optimization. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7306–7313. IEEE, 2018.

[16] Joel Goh and Melvyn Sim. Distributionally robust optimization and its tractable approximations. *Operations research*, 58(4-part-1):902–917, 2010.

[17] Alex Graves. Practical variational inference for neural networks. In *Neural Information Processing Systems*, pages 2348–2356, 2011.

[18] José Miguel Hernández-Lobato and Ryan Adams. Probabilistic backpropagation for scalable learning of Bayesian neural networks. In *International Conference on Machine Learning*, pages 1861–1869, 2015.

[19] Wonseok Jeon, Seokin Seo, and Kee-Eung Kim. A bayesian approach to generative adversarial imitation learning. In *Advances in Neural Information Processing Systems*, pages 7429–7439, 2018.

[20] Gregory Kahn, Adam Villaflor, Vitchyr Pong, Pieter Abbeel, and Sergey Levine. Uncertainty-aware reinforcement learning for collision avoidance. *arXiv preprint arXiv:1702.01182*, 2017.

[21] Saleem A Kassam and H Vincent Poor. Robust techniques for signal processing: A survey. *Proceedings of the IEEE*, 73(3):433–481, 1985.

[22] Zachary Kenton, Angelos Filos, Owain Evans, and Yarin Gal. Generalizing from a few environments in safety-critical reinforcement learning. *arXiv preprint arXiv:1907.01475*, 2019.

[23] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *Neural Information Processing Systems*, pages 6402–6413, 2017.

[24] Zhihao Li, Toshiyuki Motoyoshi, Kazuma Sasaki, Tetsuya Ogata, and Shigeki Sugano. Rethinking self-driving: Multi-task knowledge for better generalization and accident explanation ability. *arXiv preprint arXiv:1809.11100*, 2018.

[25] Xiaodan Liang, Tairui Wang, Luona Yang, and Eric Xing. CIRL: Controllable imitative reinforcement learning for vision-based self-driving. *arXiv preprint arXiv:1807.03776*, 2018.

[26] Chenggang Liu and Christopher G Atkeson. Standing balance control using a trajectory library. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3031–3036. Citeseer, 2009.

[27] Björn Lötjens, Michael Everett, and Jonathan P How. Safe reinforcement learning with model uncertainty estimates. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8662–8668. IEEE, 2019.

[28] Rowan McAllister, Yarin Gal, Alex Kendall, Mark Van Der Wilk, Amar Shah, Roberto Cipolla, and Adrian Vivian Weller. Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, 2017.

[29] Radford M Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 2012.

[30] Takayuki Osa, Joni Pajarinen, Gerhard Neumann, J Andrew Bagnell, Pieter Abbeel, Jan Peters, et al. An algorithmic perspective on imitation learning. *Foundations and Trends® in Robotics*, 7(1-2):1–179, 2018.

[31] Dean A Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *Advances in Neural Information Processing Systems (NIPS)*, pages 305–313, 1989.

[32] Nicholas Rhinehart, Kris M Kitani, and Paul Vernaza. R2P2: A reparameterized pushforward policy for diverse, precise generative path forecasting. In *European Conference on Computer Vision*, pages 772–788, 2018.

[33] Nicholas Rhinehart, Rowan McAllister, and Sergey Levine. Deep imitative models for flexible inference, planning, and control. *arXiv preprint arXiv:1810.06544*, 2018.

[34] Nicholas Rhinehart, Rowan McAllister, Kris Kitani, and Sergey Levine. PRECOG: Prediction conditioned on goals in visual multi-agent settings. *International Conference on Computer Vision*, 2019.

[35] Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *International Conference on Artificial Intelligence and Statistics*, pages 627–635, 2011.

[36] Axel Sauer, Nikolay Savinov, and Andreas Geiger. Conditional affordance learning for driving in urban environments. *arXiv preprint arXiv:1806.06498*, 2018.

[37] Herbert E Scarf. A min-max solution of an inventory problem. Technical report, RAND CORP SANTA MONICA CALIF, 1957.

[38] Sergio Verdu and H Poor. On minimax robustness: A general approach and applications. *IEEE Transactions on Information Theory*, 30(2):328–340, 1984.

[39] John Von Neumann and Oskar Morgenstern. *Theory of games and economic behavior*. Princeton University Press, 1953.

[40] Abraham Wald. Contributions to the theory of statistical estimation and testing hypotheses. *The Annals of Mathematical Statistics*, 10(4):299–326, 1939.