**Carnegie Mellon University**

**DISCLOSURE OF INTELLECTUAL PROPERTY**

All information requested in this document must be completed in order to expeditiously process this Intellectual Property Disclosure. Any missing or incomplete information may delay processing your submission. The submission of this Intellectual Property Disclosure does NOT provide any legal patent protection for the subject intellectual property. For additional information, please refer to CTTEC’s website at: www.cmu.edu/cttec.

**1. Title**:

Safety-aware Causal Representation for Trustworthy Reinforcement Learning in Autonomous Driving

1. **Creator(s): By signing this form the undersigned Creator(s) acknowledge and agree that they are bound by Carnegie Mellon University’s Intellectual Property Policy, online at http://www.cmu.edu/policies/documents/IntellProp.html (the “Policy). Therefore, by signing below: (i) if the Policy allows CMU to own this intellectual property and its associated intellectual property rights, you hereby assign to Carnegie Mellon any and all ownership you have in such intellectual property and intellectual property rights; and (ii) if the Policy allows CMU to receive license rights to this intellectual property and its associated intellectual property rights, you hereby grant to CMU any and all such licenses. Electronic or original signatures for all Creators are required.**

**a. Lead Creator:**

Ding Zhao 10/13/23

print or type name signature date

Mechanical Engineering dingzhao@cmu.edu\_\_ 40

department phone current e-mail % of contribution

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nature of contribution to the IP (Briefly explain why this person is a creator)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Nature of contribution to the IP (Briefly explain why this person is a creator)-continued

Faculty\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Carnegie Mellon Employment Status at the time the intellectual property was created (Faculty, Staff, Student, Visitor, Courtesy, etc.)

Full institutional address (if not affiliated with Carnegie Mellon)

Full residential address (street, city, state)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Residence

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Citizenship

**b. Next Creator:**

Haohong Lin 10/13/23

print or type name signature date

Mechanical Engineering (412)-251-9156 [haohongl@andrew.cmu.edu](mailto:haohongl@andrew.cmu.edu) 40

department phone current e-mail % of contribution

Nature of contribution to the IP (Briefly explain why this person is a creator)

FUSION Framework Design, Code Implementation, Experiment Setup on MetaDrive\_\_\_\_\_\_\_\_\_\_\_\_

Nature of contribution to the IP (Briefly explain why this person is a creator)-continued

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Carnegie Mellon Employment Status at the time the intellectual property was created (Faculty, Staff, Student, Visitor, Courtesy, etc.)

Student

Full institutional address (if not affiliated with Carnegie Mellon)

Full residential address (street, city, state)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Residence

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Citizenship

**c. Next Creator:**

Wenhao Ding 10/13/23

print or type name signature date

Mechanical Engineering (412)-xxx-xxxx [wenhaod@andrew.cmu.edu](mailto:wenhaod@andrew.cmu.edu) 5

department phone current e-mail % of contribution

Nature of contribution to the IP (Briefly explain why this person is a creator)

Causal Representation of FUSION framework design\_\_\_\_\_\_\_\_\_\_\_

Nature of contribution to the IP (Briefly explain why this person is a creator)-continued

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Carnegie Mellon Employment Status at the time the intellectual property was created (Faculty, Staff, Student, Visitor, Courtesy, etc.)

Student

Full institutional address (if not affiliated with Carnegie Mellon)

Full residential address (street, city, state)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Residence

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Citizenship

**d. Next Creator:**

Zuxin Liu 10/13/23

print or type name signature date

Mechanical Engineering (412)-616-6786 [zuxinl@andrew.cmu.edu](mailto:zuxinl@andrew.cmu.edu) 5

department phone current e-mail % of contribution

Nature of contribution to the IP (Briefly explain why this person is a creator)

Offline Safe RL framework and baseline agent design\_\_\_\_\_\_ \_\_\_\_\_\_

Nature of contribution to the IP (Briefly explain why this person is a creator)-continued

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Carnegie Mellon Employment Status at the time the intellectual property was created (Faculty, Staff, Student, Visitor, Courtesy, etc.)

Student

Full institutional address (if not affiliated with Carnegie Mellon)

Full residential address (street, city, state)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Residence

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Country of Citizenship

Please list additional inventor(s) and relevant information on an additional sheet.

**3. Description**

**a.** Please provide a short (1-2 paragraph), non-confidential description of your intellectual property in the space below.

We introduce the saFety-aware strUctured Scenario representatION (FUSION), a pioneering methodology conceived to facilitate the learning of an adaptive end-to-end driving policy by leveraging structured scenario information. FUSION capitalizes on the causal relationships between decomposed reward, cost, state, and action space, constructing a framework for structured sequential reasoning under dynamic traffic environments. Empirical evidence attests that FUSION significantly enhances the safety and generalizability of autonomous driving agents, even in the face of challenging and unseen driving scenarios.

**b.** Please attach a detailed description of the intellectual property – do not ‘reinvent the wheel’ (pun intended) - as both unpublished and/or previously published documents are encouraged. Feel free to attach sketches, drawings, photographs and other materials that help illustrate the description. In the attached materials, please address the following questions: How does this technology differ from present technology? What problems does it solve, and what advantages does it possess? What are the present and future uses and applications of this technology? What are the disadvantages or limitations of this technology? Has the technology been tested experimentally? Are experimental data available? Also, please suggest 3-10 keywords that someone in your field or someone in industry might use to search for your technology.

Keywords: Autonomous Driving, Safety, Offline Reinforcement Learning, Generalizability

# 1. Introduction and Background

Learning from Demonstration (LfD) techniques, such as Imitation Learning (IL) and offline Rein- forcement Learning (RL) [[1](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1ksv4uv), [2](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.44sinio), [3](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2jxsxqh), [4](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.z337ya)], have revolutionized end-to-end frameworks in autonomous vehicles. Nonetheless, the safety and generalizability of learning-based driving policies across diverse scenarios remain elusive [[5](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3j2qqm3), [6](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1y810tw), [7](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.4i7ojhp), [8](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2xcytpi)]. These challenges become even more pronounced in intricate contexts involving complex vehicle-to-road and vehicle-to-vehicle interaction. Prior studies [[9](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1ci93xb), [10](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3whwml4)] illustrate that minor domain shifts in road structures or surrounding vehicles can result in catastrophic outcomes, given the high-stakes nature of autonomous driving. While existing research has success- fully applied end-to-end learning-based algorithms to race cars [[11](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2bn6wsx), [12](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.qsh70q), [13](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3as4poj)], urban driving scenarios remain a puzzle. The complexity arises from the fact that urban settings demand robust structural reasoning from context-rich, safety-critical situations [[14](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1pxezwc)]. For instance, *humans* can effortlessly adapt their driving based on static contexts like roadblocks or dynamic contexts such as surrounding traffic, often making intuitive judgments, as illustrated in Figure. [1](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.30j0zll). Such innate structural abstraction capabilities, although second nature to us, pose significant challenges for data-driven approaches in autonomous vehicles. Two pivotal issues emerge: (i) striking a balance between safety and driving efficiency, and (ii) ensuring safety performance in unseen driving contexts.

A diagram of a road safety system

Description automatically generated

Figure 1: Diagram depicting offline-to-online generalization via a modular reasoning framework. The agent learns a causal abstraction from demonstration trajectories and then applies it to different environments online. This abstracted representation enables learning agile agents for unseen scenarios in a zero-shot manner while enhancing safety and efficiency.

Recent LfD advancements in autonomous driving have strived for safety improvements through various means, including agile actions [[1](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1ksv4uv), [3](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2jxsxqh)], object-centric world models [[5](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3j2qqm3), [15](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.49x2ik5)], safety-enhanced scene representation [[6](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1y810tw), [7](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.4i7ojhp)], and structure-aware representation of multi-modal sensory inputs [[16](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2p2csry)]. Moreover, techniques like domain-invariant IL [[17](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.147n2zr)] and hierarchical IL [[18](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3o7alnk)] bolster the generation of a safe and universally applicable causal representation under out-of-distribution (OOD) states. However, a recurring limitation is the presupposition of access to perfect expert demonstrations, which are often unattainable in intricate urban scenarios.

Autonomous driving stands out from other sequential decision-making arenas due to its (i) critical emphasis on safety and (ii) intricate dynamics under safety-sensitive conditions. While humans can discern and act upon pivotal information in multi-modal observations, data-driven agents, such as autonomous vehicles, grapple with comprehending their surroundings structurally in an end-to-end training paradigm.

In this study, we introduce sa**F**ety-aware str**U**ctural **S**cenario representat**ION** (FUSION), which aims to improve the generalizability of safety performance of self-driving cars in unseen scenarios. More concretely, our contributions are summarized as follows:

* We introduce a safety-aware offline reinforcement learning framework that successfully balances the trade-off between efficiency and safety, termed the Causal Ensemble World Model (CEWM).
* We develop a Causal Bisimulation Learning (CBL) paradigm that regularizes the state representation in a compact way, enabling better generalizability towards OOD state inputs during the online deployment stage.
* We provide comprehensive evaluations, comparing our approach with existing baselines in safety- aware imitation learning and offline safe reinforcement learning.

# 2. Methodology

In this section, we zoom in on more details about our proposed FUSION with two important modules: (i) Causal Ensemble World Model (CEWM), and (ii) safety-aware Causal Bisimulation Learning (CBL).

A screen shot of a computer

Description automatically generated

Figure 2: Overview of Safety-aware structural Scenario Representation Framework. The left diagram shows a safety-aware decision transformer that conducts sequential decision-making based on the temporal contexts. The right diagram shows the general form of the graphical model in the CEWM and Policy Learning modules in FUSION, where the connection between different timesteps will be determined by the attention weights in the causal transformer. The nodes at a later timestep depend on their parental nodes in the previous timesteps.

## 2.1 Causal Ensemble World Model Learning

Similar to the settings in [[16](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2p2csry)], in autonomous driving problems, the entire state space can be decomposed into several disjoint subspaces, including the (estimated) ego navigation state, lidar observation, and visual observation, e.g. the birds-eye-view observation that we use for this work, as is shown in Figure [2](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3znysh7).

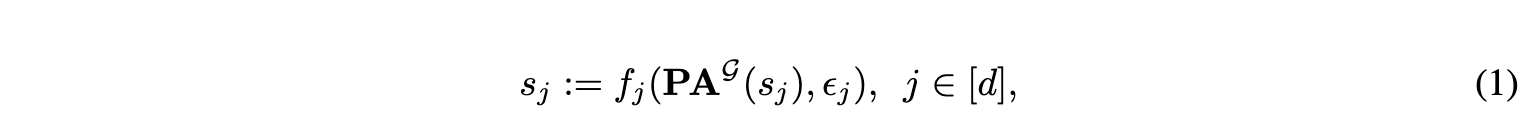
**A math problem with numbers and symbols

Description automatically generated with medium confidence**

**Definition 1** (Factorizable State Space)**.** *The factorizable state space in MDP indicates a disjoint state space decomposition, where S* = *S*1 *∪ S*2 *∪ · · · ∪ SN , N indicates how many disjoint state components we have in certain problems.*

To help the FUSION framework gain better awareness of the structure of the state and action space, we propose the CEWM based on multi-modal observations. Definition [1](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.tyjcwt), along with the reward, cost, and action variables, form the nodes in this world model. To better describe the structural dependency between them, we further design the CEWM according to the following definition of Structured Causal Model (SCM).

**Definition 2.** *An SCM θ* := (*S, E* ) *consists of a collection S of d functions [*[*19*](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.23ckvvd)*],*

**

For general offline RL problems, SCM aims to parameterize the world model between different nodes in the state, action, and preference space, as shown in Figure [2](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3znysh7). Every child node *sj* is determined by its parent node **PA***G* (*sj*) and the exogenous noise variable *ϵj*, which are aggregated by a variable- specific function *fj* with an arbitrary parameterization. More specifically, in our autonomous driving problem, the nodes in the given causal graph *G* are the factorized state space, the action space, and the preference space. The edges between different nodes represent their causal dependency in

spatiotemporal space. Besides capturing the cause-and-effect relationship between the reward, cost, and factorizable state space, the SCM also enjoys a great property in that the child nodes (e.g. the state and reward/cost in subsequent timesteps) are only dependent on their parental node (in the state or action space in previous timesteps), while will be independent of other nodes conditioned on these parent nodes. Such property guarantees a potential for autoregressive generation during the inference time.

Based on this property, we derive the CEWM under the SCM, which can then be decomposed into the following disjoint components, including the reward-to-go model, cost-to-go model, the factorized state-action transition dynamics, and the policy optimization, as is shown below:

A math equations and formulas

Description automatically generated with medium confidence

Therefore, we exert an auxiliary task of trajectory optimization in the optimization process of safety-aware decision transformer to estimate the three components in ([2](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3dy6vkm)), i.e.

A group of math equations

Description automatically generated

This trajectory optimization objective benefits our safety-aware DT with better structural awareness of the trajectory level between the state, action, reward-to-go, and cost-to-go. The design of this safety-aware DT model aligns with the CEWM that we propose in ([2](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3dy6vkm)) because the latter token is generated conditioned on the previous tokens during the inference time in an auto-regressive way. During training time, we supervise the output at each token along the trajectories from the offline demonstration in Algorithm [5](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2dlolyb).

## 2.2. Safety-aware Bisimulation Learning

Though CEWM provides an *explicit* structure to model the causality, learning such a model from offline datasets is non-trivial. The reason is that demonstrations in the mixed-quality dataset have diverse levels of safety due to spurious correlations between actions and states. To avoid getting misled by such spurious correlation, we introduce an additional self-supervised regularization term in an *implicit* way, namely Causal Bisimulation Learning, or CBL.

CBL is originally inspired by [[20](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.ihv636)] in online off-policy RL. The difference between our work and [[20](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.ihv636)] is that we further regularize the FUSION model with safety-aware Bisimulation Learning in an offline RL setting. To give a clearer picture of the proposed CBL, we first define the safety-aware bisimulation relationships, which extends the traditional bisimulation relationships for MDP in [[21](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.32hioqz), [20](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.ihv636)] to achieve better generalizability:

**Definition 3** (Safety-aware Bisimulation Relation)**.** *A safety-aware bisimulation relation U ⊂ S × S*

*is a binary relation which satisfies, ∀*(*s*1*, s*2) *∈ U:*

A math equations on a white background

Description automatically generated

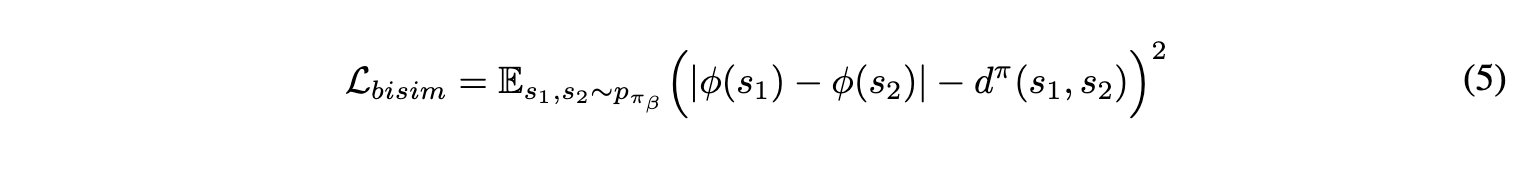
Intuitively, in the Constrained MDP setting, the bisimilarity between two states is not only determined by the step-wise reward and transition dynamics but also by their similarity in the step-wise cost. In practice, the reward, cost, and transition dynamics could hardly match exactly for two different states, therefore, we propose a smooth alternative [[22](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.1hmsyys)] of the safety-aware bisimulation relationship, denoted as Safety-aware Bisimulation Metrics as is shown in Figure [3](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.2s8eyo1):

**Definition 4** (Safety-aware Bisimulation Metrics)**.** *The bisimulation metric dπ* : *S × S →* R+ *is a mapping from the joint state space to a non-negative scalar. According to the definition of the safety-aware bisimulation relationship, the distance is defined as:*

A math equations on a white background

Description automatically generated

We then use the following learning objectives to ensure the state representation aligns with the bisimulation metrics in the latent space:



In the inference time, we greedily exploit the value prediction in the online inference time, as is shown in Algorithm [3](https://docs.google.com/document/d/1zfovl9Hzy9NqjdBs4uphvObpl215FdxW/edit#heading=h.3rdcrjn). Notably, we take the minimum cost-to-go preference and cost prediction, and the maximum reward-to-go preference and reward prediction at each step. This strategy aims to improve the safety and efficiency of FUSION given the preference in the online deployment stage.

A diagram of a transformer encoder

Description automatically generated

Figure 3: Safety-aware bisimulation metrics with the distribution distance in transition dynamics, rewards, and safety cost.

**A black and white text with black text

Description automatically generated**

**A math equations on a white background

Description automatically generated**

# 3. Conclusion

# In this work, we propose FUSION, a trustworthy autonomous driving system with a causality-empowered safe reinforcement learning algorithm in an offline setting. We first design a safety-aware causal transformer termed CEWM to model the causal relationship between the state space, reward value, and cost value at different timesteps. Then we regularize the learned representation in CEWM with a CBL to enforce their compactness via safety-aware bisimulation in an implicit way, then greedily infer the action during online deployment. Exhaustive empirical results show that our method consistently outperforms offline demonstration and several strong baselines in safe IL or offline safe RL under diverse urban autonomous driving scenarios. We also conduct extensive analysis to analyze the benefits of different modules that we design in FUSION and show a comprehensive and interpretable evaluation of FUSION against its variants or other baselines. One potential limitation is that all the experiments are conducted in the MetaDrive simulator since it is more portable than CARLA or other autonomous driving simulators. It would be interesting to extend FUSION's framework to other autonomous vehicle simulators with higher fidelity, as well as the multi-agent RL settings in the near future.

**4. Intellectual Property Protections**

**a.** The Intellectual Property described in this Disclosure includes (Please “X” all that apply):

\_\_\_\_\_\_ A composition of matter (chemical entity)

\_\_\_\_\_\_ A machine

\_\_\_\_\_\_ A method or process

\_\_ X\_\_ Software

\_\_\_\_\_\_ Designs or other copyrightable materials

\_\_\_\_\_\_ Biological materials

\_\_\_\_\_\_ Other; please specify \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

For the questions #4b and 4c, please indicate the date in the format “Month/Day/ Year” (ex. 01/01/17).

**b.** State first date of:

1. Conception \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Sketch or drawing \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Written description \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Completion of working model \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(or operational process)

**c.** State earliest date, actual or expected, of:

1. Disclosure to others (non-CMU employees) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. Printed publication \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. Oral disclosure (e.g. seminars, conferences, etc.); \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Use for profit or Offer for sale. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**d.** If your work has been disclosed to parties outside of Carnegie Mellon, was it made under the provisions of either a Funding Agreement or governed by a Confidentiality Agreement?

Yes.

**e.** Has there beena prior art/patent and/or literature searchrelating to this technology? If yes, please include copies of any resulting documentation.

**f.** Haveany patent applications been submitted for this intellectual property as of this date? If yes, please include copies of any resulting documentation.

# References

1. M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel,

M. Monfort, U. Muller, J. Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.

1. J. Chen, B. Yuan, and M. Tomizuka. Model-free deep reinforcement learning for urban autonomous driving. In *2019 IEEE intelligent transportation systems conference (ITSC)*, pages 2765–2771. IEEE, 2019.
2. Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots. Imitation learning for agile autonomous driving. *The International Journal of Robotics Research*, 39(2-3): 286–302, 2020.
3. J. Chen, S. E. Li, and M. Tomizuka. Interpretable end-to-end urban autonomous driving with latent deep reinforcement learning. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5068–5078, 2021.
4. K. Renz, K. Chitta, O.-B. Mercea, A. Koepke, Z. Akata, and A. Geiger. Plant: Explainable planning transformers via object-level representations. *arXiv preprint arXiv:2210.14222*, 2022.
5. H. Shao, L. Wang, R. Chen, H. Li, and Y. Liu. Safety-enhanced autonomous driving using interpretable sensor fusion transformer. In *Conference on Robot Learning*, pages 726–737. PMLR, 2023.
6. X. Jia, P. Wu, L. Chen, J. Xie, C. He, J. Yan, and H. Li. Think twice before driving: To- wards scalable decoders for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21983–21994, 2023.
7. H. Shao, L. Wang, R. Chen, S. L. Waslander, H. Li, and Y. Liu. Reasonnet: End-to-end driving with temporal and global reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13723–13733, 2023.
8. W. Ding, C. Xu, M. Arief, H. Lin, B. Li, and D. Zhao. A survey on safety-critical driving scenario generation—a methodological perspective. *IEEE Transactions on Intelligent Transportation Systems*, 2023.
9. M. Xu, Z. Liu, P. Huang, W. Ding, Z. Cen, B. Li, and D. Zhao. Trustworthy reinforcement learning against intrinsic vulnerabilities: Robustness, safety, and generalizability. *arXiv preprint arXiv:2209.08025*, 2022.
10. F. Fuchs, Y. Song, E. Kaufmann, D. Scaramuzza, and P. Dürr. Super-human performance in gran turismo sport using deep reinforcement learning. *IEEE Robotics and Automation Letters*, 6(3):4257–4264, 2021.
11. P. R. Wurman, S. Barrett, K. Kawamoto, J. MacGlashan, K. Subramanian, T. J. Walsh, R. Capo- bianco, A. Devlic, F. Eckert, F. Fuchs, et al. Outracing champion gran turismo drivers with deep reinforcement learning. *Nature*, 602(7896):223–228, 2022.
12. D. Shah, K. Stachowicz, A. Bhorkar, I. Kostrikov, and S. Levine. Fastrlap: A system for learning high-speed driving via deep rl and autonomous practicing. In *ICRA2023 Workshop on Pretraining for Robotics (PT4R)*, 2023.
13. A. Mohan, A. Zhang, and M. Lindauer. Structure in reinforcement learning: A survey and open problems. *arXiv preprint arXiv:2306.16021*, 2023.
14. W. Ding, H. Lin, B. Li, and D. Zhao. Generalizing goal-conditioned reinforcement learning with variational causal reasoning. In *Advances in Neural Information Processing Systems*, 2022.
15. K. Lee, Z. Wang, B. Vlahov, H. Brar, and E. A. Theodorou. Ensemble bayesian decision making with redundant deep perceptual control policies. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 831–837. IEEE, 2019.
16. I. Bica, D. Jarrett, and M. van der Schaar. Invariant causal imitation learning for generalizable policies. *Advances in Neural Information Processing Systems*, 34:3952–3964, 2021.
17. R. Akrour, F. Veiga, J. Peters, and G. Neumann. Regularizing reinforcement learning with state abstraction. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 534–539. IEEE, 2018.
18. J. Peters, D. Janzing, and B. Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.
19. A. Zhang, R. McAllister, R. Calandra, Y. Gal, and S. Levine. Learning invariant representations for reinforcement learning without reconstruction. *arXiv preprint arXiv:2006.10742*, 2020.
20. L. Li, T. J. Walsh, and M. L. Littman. Towards a unified theory of state abstraction for mdps. In

*AI&M*, 2006.

1. N. Ferns, P. Panangaden, and D. Precup. Metrics for finite markov decision processes. In *UAI*, volume 4, pages 162–169, 2004.
2. A. Kumar, J. Fu, M. Soh, G. Tucker, and S. Levine. Stabilizing off-policy q-learning via bootstrapping error reduction. *Advances in Neural Information Processing Systems*, 32, 2019.
3. S. Fujimoto, D. Meger, and D. Precup. Off-policy deep reinforcement learning without explo- ration. In *International conference on machine learning*, pages 2052–2062. PMLR, 2019.
4. H. Xu, X. Zhan, and X. Zhu. Constraints penalized q-learning for safe offline reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 8753–8760, 2022.
5. K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer. Ensembledagger: A bayesian approach to safe imitation learning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5041–5048. IEEE, 2019.
6. H. Le, C. Voloshin, and Y. Yue. Batch policy learning under constraints. In *International Conference on Machine Learning*, pages 3703–3712. PMLR, 2019.
7. Z. Liu, Z. Guo, Y. Yao, Z. Cen, W. Yu, T. Zhang, and D. Zhao. Constrained decision transformer for offline safe reinforcement learning. *arXiv preprint arXiv:2302.07351*, 2023.
8. L. Chen, K. Lu, A. Rajeswaran, K. Lee, A. Grover, M. Laskin, P. Abbeel, A. Srinivas, and

I. Mordatch. Decision transformer: Reinforcement learning via sequence modeling. *Advances in neural information processing systems*, 34:15084–15097, 2021.

1. H. Liu, Z. Huang, X. Mo, and C. Lv. Augmenting reinforcement learning with transformer- based scene representation learning for decision-making of autonomous driving. *arXiv preprint arXiv:2208.12263*, 2022.
2. M. Janner, Q. Li, and S. Levine. Offline reinforcement learning as one big sequence modeling problem. *Advances in neural information processing systems*, 34:1273–1286, 2021.
3. A. Loquercio, M. Segu, and D. Scaramuzza. A general framework for uncertainty estimation in deep learning. *IEEE Robotics and Automation Letters*, 5(2):3153–3160, 2020.
4. P. Sermanet, C. Lynch, Y. Chebotar, J. Hsu, E. Jang, S. Schaal, S. Levine, and G. Brain. Time-contrastive networks: Self-supervised learning from video. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 1134–1141. IEEE, 2018.
5. S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta. R3m: A universal visual representation for robot manipulation. In *6th Annual Conference on Robot Learning*.
6. Y. J. Ma, S. Sodhani, D. Jayaraman, O. Bastani, V. Kumar, and A. Zhang. Vip: Towards universal visual reward and representation via value-implicit pre-training. *arXiv preprint arXiv:2210.00030*, 2022.
7. R. S. Sutton, D. Precup, and S. Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.
8. Y. Lee, J. J. Lim, A. Anandkumar, and Y. Zhu. Adversarial skill chaining for long-horizon robot manipulation via terminal state regularization. *arXiv preprint arXiv:2111.07999*, 2021.
9. R. Dadashi, S. Rezaeifar, N. Vieillard, L. Hussenot, O. Pietquin, and M. Geist. Offline reinforcement learning with pseudometric learning. In *International Conference on Machine Learning*, pages 2307–2318. PMLR, 2021.
10. Z. Wang, X. Xiao, Z. Xu, Y. Zhu, and P. Stone. Causal dynamics learning for task-independent state abstraction. *arXiv preprint arXiv:2206.13452*, 2022.

B. Chen, Z. Liu, J. Zhu, M. Xu, W. Ding, L. Li, and D. Zhao. Context-aware safe reinforcement learning for non-stationary environments. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10689–10695. IEEE, 2021.

**g.** Have you been personally involved in any prior patent processes for any other technology?

No

**5. Sponsorship**

**a.**

|  |  |  |
| --- | --- | --- |
| **External Sponsor(s)** | **CMU Oracle #(s)** | **Contract or Grant #(s)** |
| **Ford Motor Company** |  |  |
|  |  |  |
|  |  |  |

(Your department administrator may be of assistance in identifying funding sources used.)

Have external sponsors been informed of the invention? Please state yes or no. \_\_\_\_\_

**b.**  If funded under a grant or contract, please describe this intellectual property as one of the following (Please “X” one):

\_\_\_\_\_\_ Background IP (developed prior to but used in research supported by the grant)

\_\_\_\_\_\_ Background IP Improvement (developed both prior to and during the grant)

\_\_X\_\_\_ Foreground IP (developed only during the grant)

**c. Internal Sponsor** (Department Research Funds, etc.)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**d.** Was this intellectual property developed in collaboration with any other 3rd parties (companies, universities, etc.) or as a part of a research consortium? Please state yes or no, and list collaborators or consortium below:

**6. Use of 3rd Party Resources**

**a.** Have youused any third-party resources in the creation of your technology (i.e, material or equipment from a company or university under a Material Transfer Agreement (MTA) or other formal or informal agreement)?

No

**b.** Have you used any software, libraries, etc. from other internal (e.g., CMU) sources (ex. projects or researchers) in the development of this technology or does the technology otherwise build upon earlier work at CMU?

Yes

**c.** Have you used any Open Source software in the development of this technology? Please list below:

MetaDrive: <https://github.com/metadriverse/metadrive/>

FSRL: <https://github.com/liuzuxin/FSRL>

**7. Commercialization**

**a.** Give youropinion on the current stage of developmentof the technology as it relates to its current marketability (Please “X” the appropriate response.)

\_\_\_\_\_X\_\_\_\_\_\_embryonic (needs substantial work to bring to market)

\_\_\_\_\_\_\_\_\_\_\_\_ partially developed (could be brought to market with nominal investment)

\_\_\_\_\_\_\_\_\_\_\_\_off-the-shelf (could be brought to market with minimal investment)

**b.** Has the technology been disclosed or provided to any industry representatives? Has any industry representative, entrepreneur or funder shown commercial interest in this technology? Do you know (or have reason to believe) there will be commercial interest in this invention? Please specify any application areas and/or products currently on the market that you feel may embody aspects of your technology.

**c.** How long would it take you to verbally transfer the requisite, implicit knowledge to industry peers in order to enable the development and commercialization of the technology without your substantial, further involvement (ex. # months, weeks, years)?

**d.** Please rate how interested you are instarting a firm to commercialize the technology? (Please

“X” one)

\_\_\_\_ 1. Highly interested. I’ve already started to look into it.

\_\_X\_ 2. Somewhat interested. I’d like to know more about it.

\_\_\_\_ 3. I’ve not considered it positively or negatively.

\_\_\_\_ 4. Not particularly interested, but willing to consider it under the right circumstances.

\_\_\_\_ 5. Would not be willing to start a firm.

**8.** Pleaserecommend two professionals from the CMU community who would be best suited to evaluate the technical and commercial merits of this technology:

a. Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Dept.: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ext.: \_\_\_\_\_\_\_\_\_\_\_

Email: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

b. Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Dept.: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Ext.: \_\_\_\_\_\_\_\_\_\_\_

Email: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Please feel free to attach additional material or data

that would provide us with helpful information.

**Email the completed electronic copy of this Invention Disclosure form to:**

[*innovation@cmu.edu*](mailto:innovation@cmu.edu)

*If unable to sign electronically, paper copies may be sent to:*

*Department Administrator*

*Center for Technology Transfer & Enterprise Creation*

*4615 Forbes Avenue, Suite 302*