The Twelfth International Conference
on Learning Representations

Physics-Regulated Deep Reinforcement Learning: Invariant Embeddings

Hongpeng Cao[1], Yanbing Mao[2], Lui Sha[3] and Marco Caccamo[1,4]

[1]Technical University of Munich (TUM), Munich, Germany
[2]Wayne State University, Detroit, MI, USA
[3]University of Illinois Urbana-Champaign (UIUC), Urbana, IL, USA
[4]Munich Institute of Robotics and Machine Intelligence, Munich, Germany

Motivations

Deep Reinforcement Learning (DRL) in Safety-Critical Applications

Autonomous driving [1]

Industry applications [2]

Autonomous Exploration [3]

Autonomous Flight [4]

[1] https://bernardmarr.com/how-tesla-is-using-artificial-intelligence-to-create-the-autonomous-cars-of-the-future/

[2] Liu, Quan, et al. "Deep reinforcement learning-based safe interaction for industrial human-robot collaboration using intrinsic reward function." Advanced Engineering Informatics 49 (2021): 101360.

[3] Lee, Joonho, et al. "Learning quadrupedal locomotion over challenging terrain." Science robotics 5.47 (2020): eabc5986.

[4] https://www.traveldailymedia.com/autonomous-aircraft-market-research/

Motivations

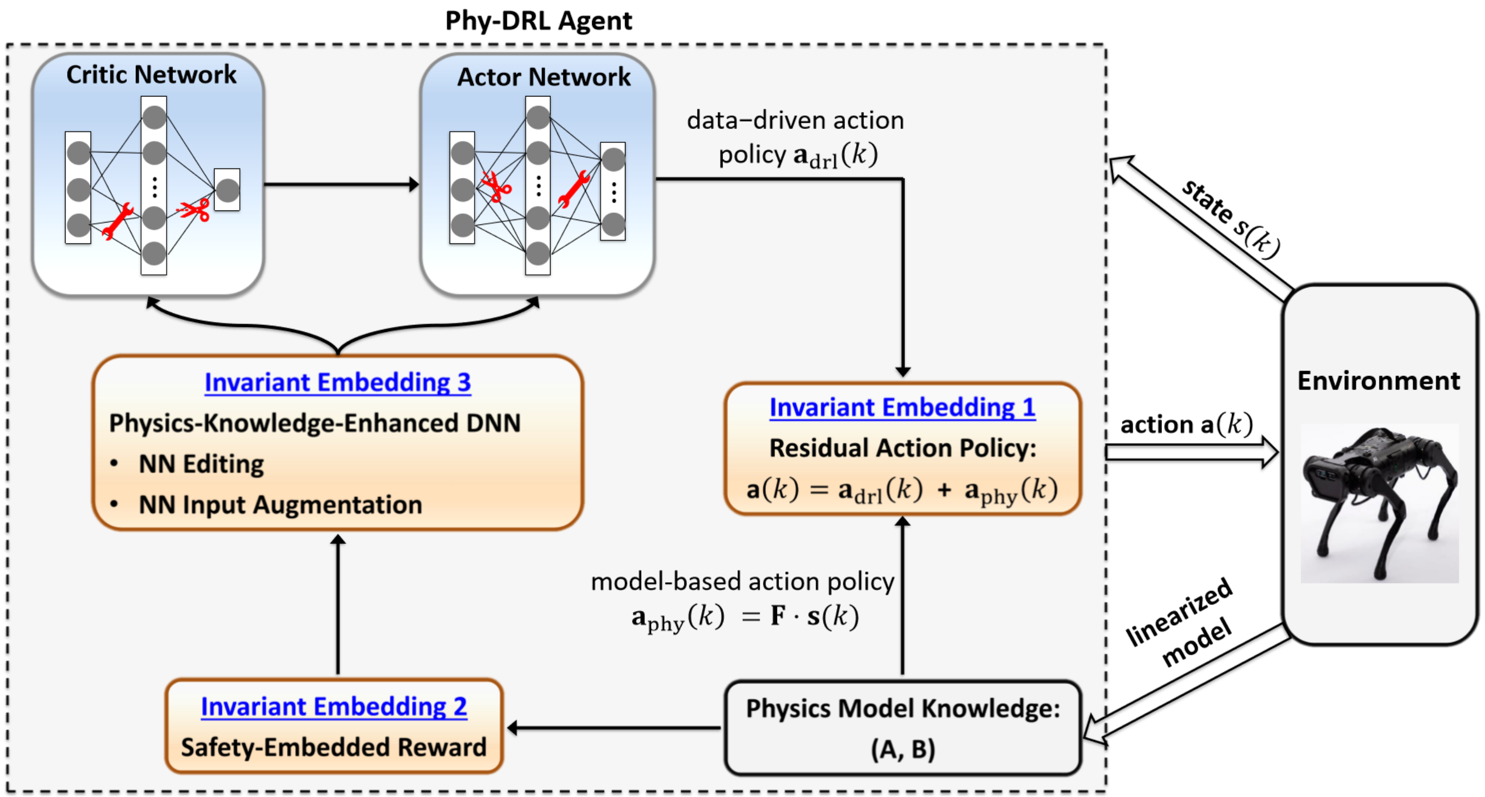
Unsolved problems:

- **Safety and stability**:
  - Hard to verify DNNs due to high dimensionalities and high nonlinearities.
  - Hard to predict the output of DNNs due to the vulnerability to the disturbances.
  - Purely data-driven DNN applied to physical systems can infer relations violating physics laws.
- **Sampling complexity**:
  - High demand of training data.
  - Unsafe explorations.

Can we use the physics knowledge about the system to 'regulate' DRL to make it safer and more reliable in Safety-Critical Applications?

Contributions

Overview:



Phy-DRL:
A Physics-regulated Deep Reinforcement Learning Framework

Invariant Embedding 1

Residual Action Policy:
Integrating data-driven DRL action policy and physics-model-based action policy.

Invariant Embedding 2

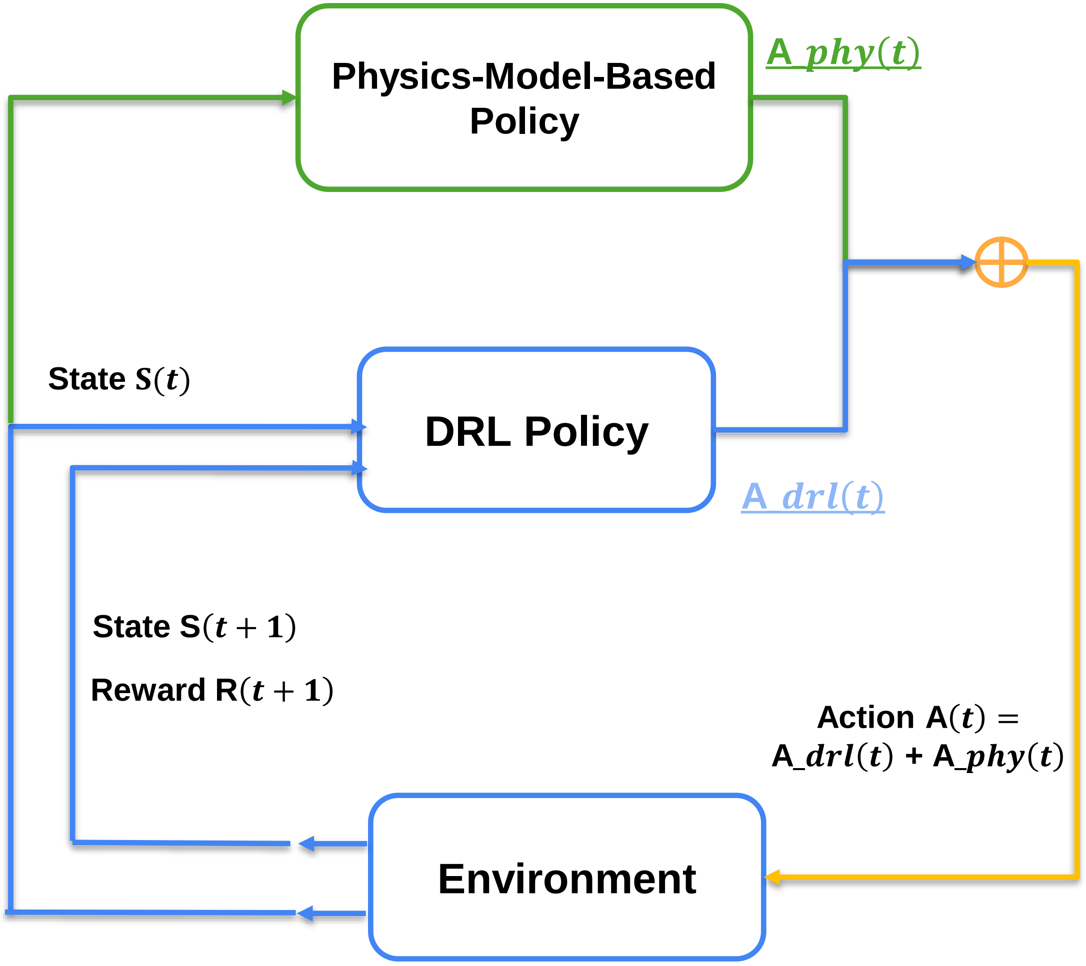
Safety-Embedded Reward:
In conjunction with the Residual Action Policy, empowers the Phy-DRL with a mathematically provable safety guarantee and fast training.

Invariant Embedding 3

Physics-Knowledge-Enhanced Critic and Actor Networks:
Including input augmentation and network editing for guaranteeing strict compliance with available knowledge about the action-value function and action policy.

Residual Action Policy

Integrating data-driven DRL action policy and physics-model-based action policy.



Real plant

$$\mathbf{s}(k+1)=\mathbf{A}\mathbf{s}(k)+\mathbf{B}\mathbf{a}(k)+\mathbf{f}(\mathbf{s}(k),\mathbf{a}(k)), \qquad k\in\mathbb{N}$$

Safety constraints

$$\mathbb{X}\triangleq\left\{\mathbf{s}\in\mathbb{R}^n \,\middle|\, \underline{\mathbf{v}}\leq\mathbf{D}\cdot\mathbf{s}-\mathbf{v}\leq\overline{\mathbf{v}}\right\},$$

Physics-Model-Based Policy

Safety envelope

$$\Omega\triangleq\left\{\mathbf{s}\in\mathbb{R}^n \,\middle|\, \mathbf{s}^\top\mathbf{P}\mathbf{s}\leq 1,\ \mathbf{P}\succ 0\right\}.$$

Obtained by computing feedback Matrix F

$\mathbf{a}_{\text{phy}}(k)=\mathbf{F}\mathbf{s}(k)$, where **F** is obtained via solving LMIs

DRL Policy

Learned by maximizing the expected return

$$\mathbf{a}_{\text{drl}}(k)=\arg\max \mathbf{E}_{\mathbf{s}(k)\sim\rho,\ \mathbf{a}_{\text{drl}}(k)\sim\pi}\left[\sum_{t=k}^{\infty}\gamma^{t-k}\cdot\mathcal{R}\left(\mathbf{s}(t),\mathbf{a}_{\text{drl}}(t)\right)\right]$$

Safety Embedded Reward

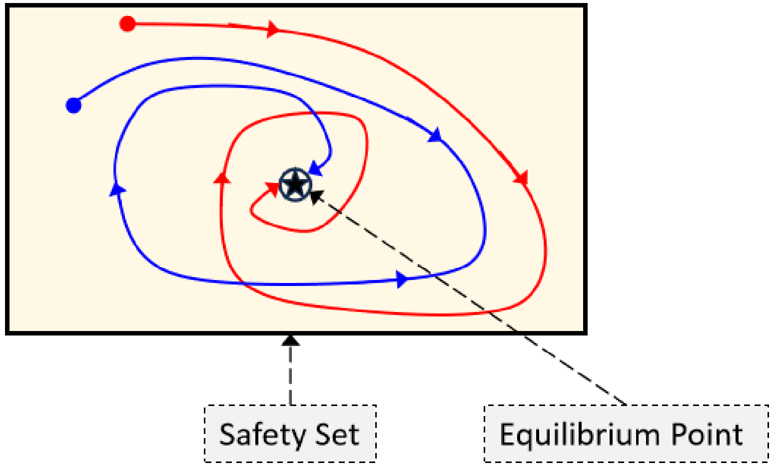
In conjunction with the Residual Action Policy, empowers the Phy-DRL with a mathematically provable safety guarantee and fast training.

Safety-Embedded Reward

$$\mathbf{r(s(k), s(k+1))} = \underbrace{\mathbf{s(k)^T \cdot \overline{A}^T \cdot P \cdot \overline{A} \cdot s(k)}}_{V(\bar{s}(k+1))} - \underbrace{\mathbf{s(k+1)^T \cdot P \cdot s(k+1)}}_{V(s(k+1))}, \text{ where } \overline{A} = A + BF$$

Predicted value using linear model and model-based controller

The value calculated using measured states from the real system



Mathematically provable safety guarantee

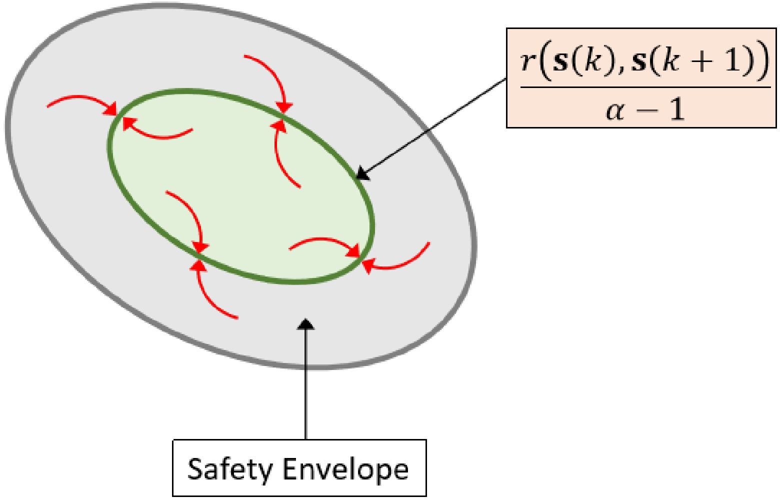
Consider the safety set $\mathbb{X}$, the safety envelope Ω, and the system under control of Phy-DRL. The matrices $\mathbf{F}$ and $\mathbf{P}$ involved in the model-based action policy and the safety-embedded reward are computed according to

$$\mathbf{F} = \mathbf{R} \cdot \mathbf{Q}^{-1}, \quad \mathbf{P} = \mathbf{Q}^{-1},$$

where $\mathbf{R}$ and $\mathbf{Q}^{-1}$ satisfy

$$\begin{bmatrix} \alpha \cdot \mathbf{Q} & \mathbf{Q} \cdot \mathbf{A}^\top + \mathbf{R}^\top \cdot \mathbf{B}^\top \\ \mathbf{A} \cdot \mathbf{Q} + \mathbf{B} \cdot \mathbf{R} & \mathbf{Q} \end{bmatrix} \succ 0, \quad \textit{with a given } \alpha \in (0,1).$$

Given any $\mathbf{s}(1) \in \Omega$, the system state $\mathbf{s}(k) \in \Omega \subseteq \mathbb{X}$ holds $\forall k \in \mathbb{N}$ (i.e., the safety of system (1) is guaranteed), if the sub-reward $r(\mathbf{s}(k), \mathbf{s}(k+1))$ satisfies $r(\mathbf{s}(k), \mathbf{s}(k+1)) \geq \alpha - 1$, $\forall k \in \mathbb{N}$.

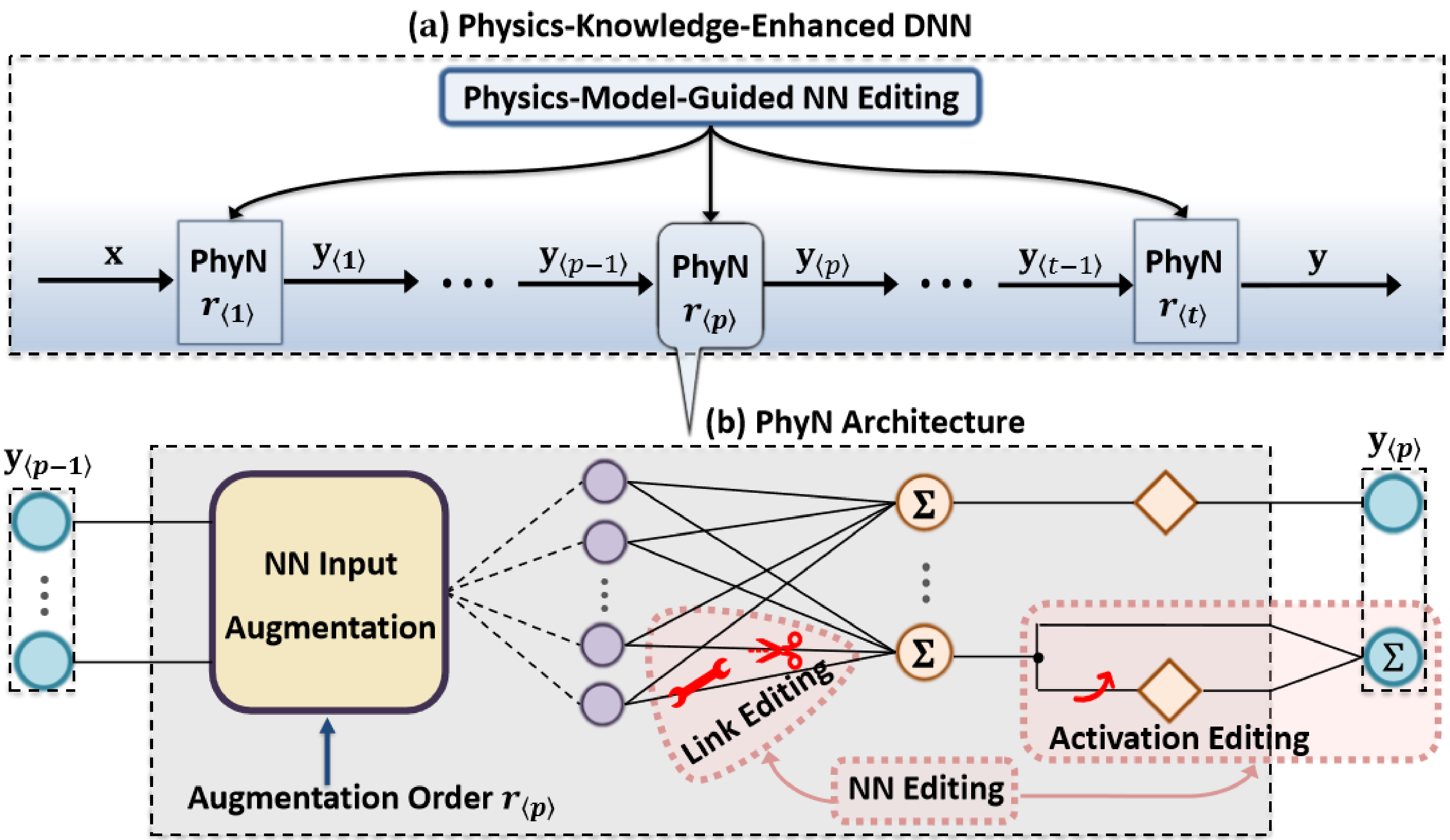


Physics-Knowledge-Enhanced Networks

Including input augmentation and network editing for guaranteeing strict compliance with available knowledge about the action-value function and action policy.

Input Augmentation:
Catching hard to learn quantities

Network Editing:
Ensuring The end-to-end input/output of the actor network strictly complies with available knowledge



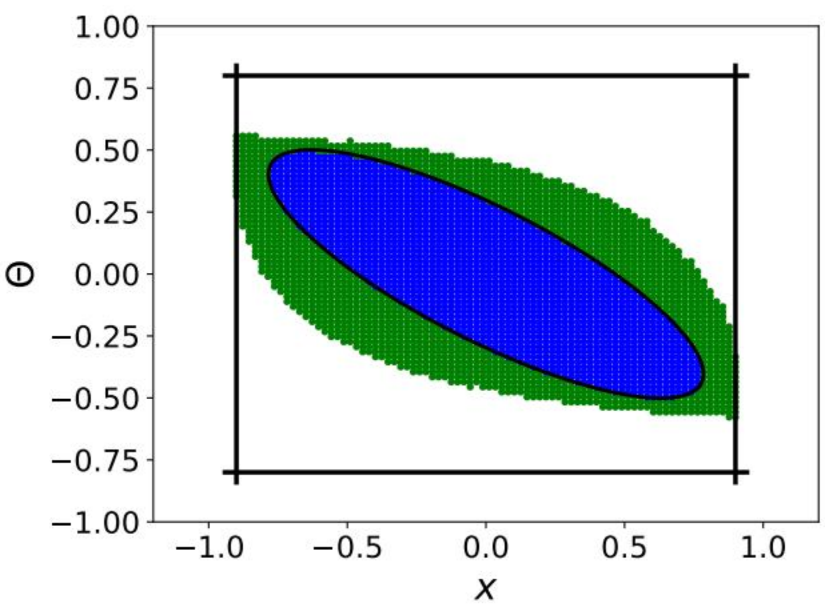
Experimental Results

Monte Carlo Simulation results in a non-linear cart-pole system

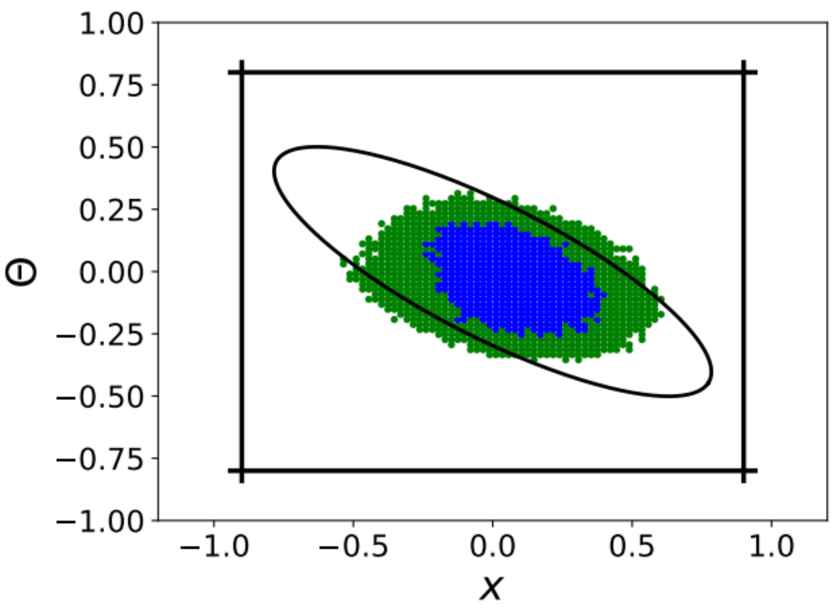
Phy-DRL can render the safety envelope **invariant**, where the others fail.

Blue points **Safe Internal-Envelope (IE) Sample** $\triangleq \widetilde{\mathbf{s}}$: if $\mathbf{s}(1) = \widetilde{\mathbf{s}} \in \Omega$, then $\mathbf{s}(k) \in \Omega, \forall k \in \mathbb{N}$.

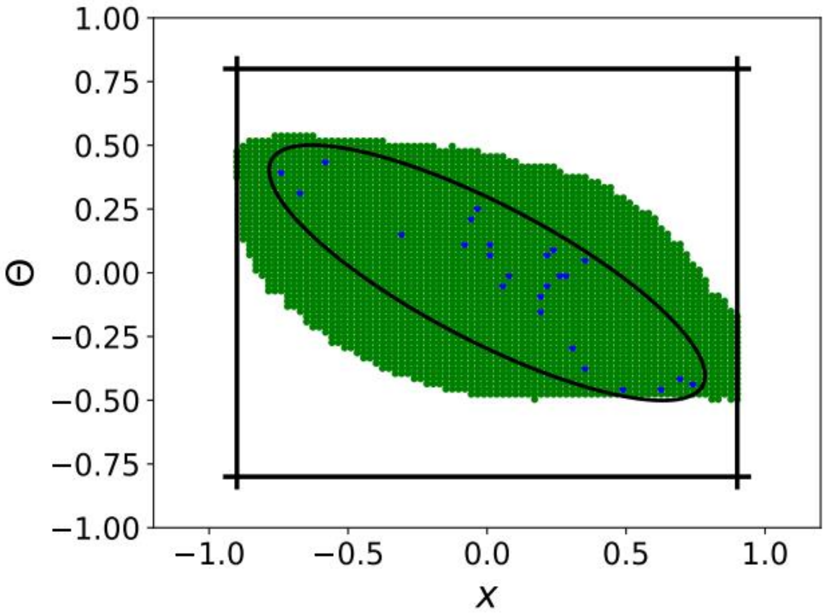
Green points **Safe External-Envelope (EE) Sample** $\triangleq \widetilde{\mathbf{s}}$: if $\mathbf{s}(1) = \widetilde{\mathbf{s}} \in \mathbb{X}$, then $\mathbf{s}(k) \in \mathbb{X} \setminus \Omega, \exists k \in \mathbb{N}$.



Phy-DRL



Linear model based

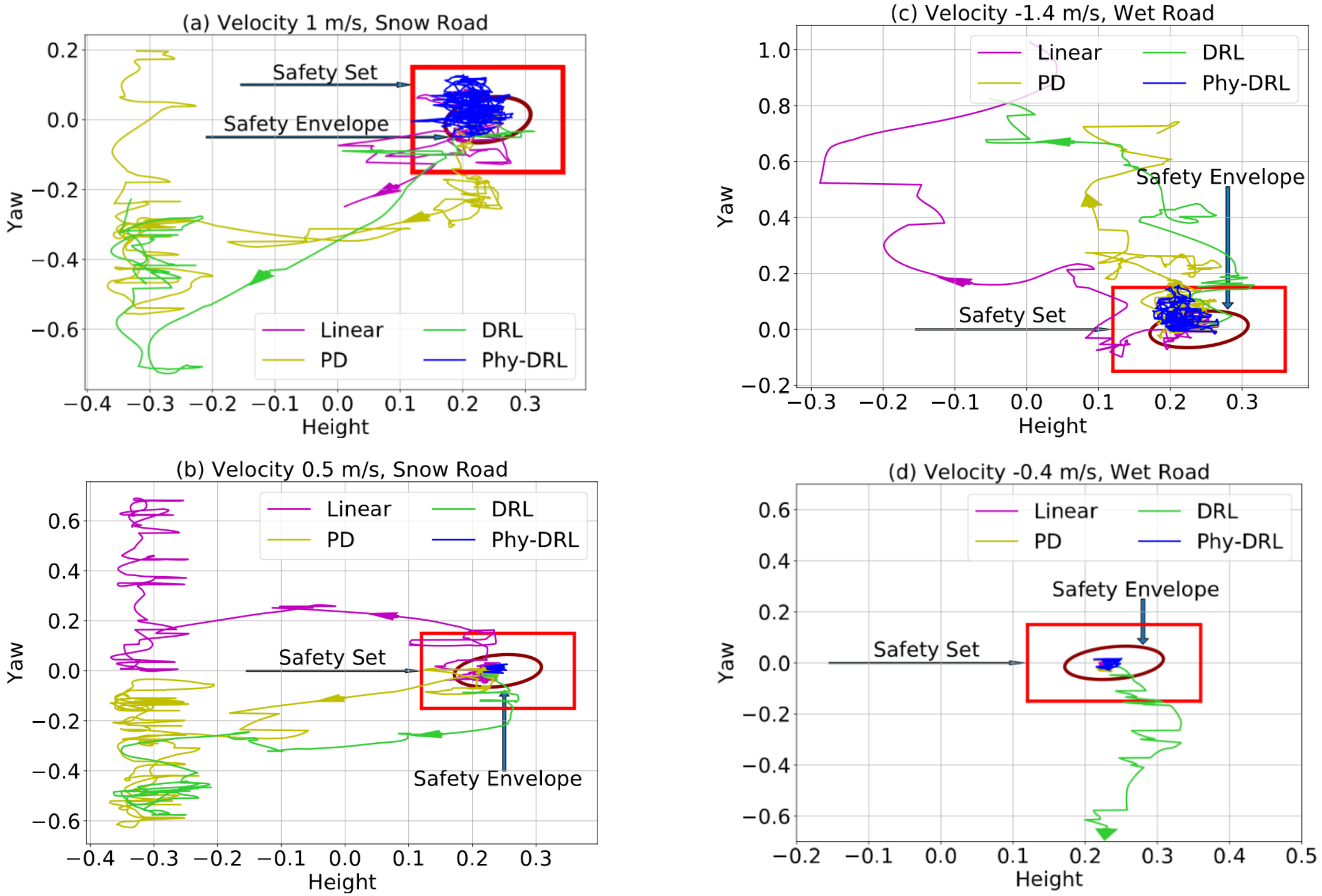


DRL without residual

Experimental Results

Quadruped robot locomotion

Phy-DRL is a more **robust** and **safer** action policy in safe center-gravity management, safe lane tracking and safe velocity regulation test in four testing scenarios.



Conclusions

We proposed a Phy-DRL framework with three invariant embeddings to improve safety assurance for DRL-enabled systems

- ***Residual Action Policy*** :
  - Using model-based controller to catch causality
  - Using data-driven DRL to deal with model mismatch
  - Less data dependencies

- ***Safety-Embedded Reward :***
  - Efficient construction of reward function using **P** matrix
  - Encourage learning a safe and stable policy simultaneously
  - Provide mathematically provable safety guarantees for DRL

- ***Physics-Knowledge-Enhanced Critic and Actor Networks:***
  - Augmenting input using physics model knowledge to catch the hard-to-learn quantities
  - Ensuring The end-to-end input/output of the actor network strictly complies with available knowledge

ICLR

The Twelfth International Conference
on Learning Representations

Thanks for your attention!

Scan to know More

Technische
Universität
München

TUM

WAYNE STATE
UNIVERSITY

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

MIRMI

Alexander von Humboldt
Stiftung/Foundation

NSF