Verification of Sim-to-Real Transfer in RL-based End-to-End Autonomous Racing Using Visual Domain Randomization

Autumn Kwon[1], Jason M. O'Kane[2], Guni Sharon[2]

[1]Computer Science, Grinnell College

[2]Computer Science and Engineering, Texas A&M University

kwongaeu@grinnell.edu   jokane@tamu.edu   guni@tamu.edu

Motivation

Reinforcement learning (RL) has been an effective method of autonomous racing, in that the AI agent can learn by directly communicating with the environment without any prior data. However, traditional RL training methods require significant time and resources. In particular, RL requires continuously experimenting with various environments with real robots since the learning time takes more than five hours. Therefore, researchers came up with a new method called sim-to-real transfer which trains the models in the various simulated environments and then deploys those models in the real environment. **This study investigates the effects of domain randomization (DR) by augmenting the visual domain with various light conditions to enhance the reinforcement learning (RL)-based autonomous racing algorithm.**

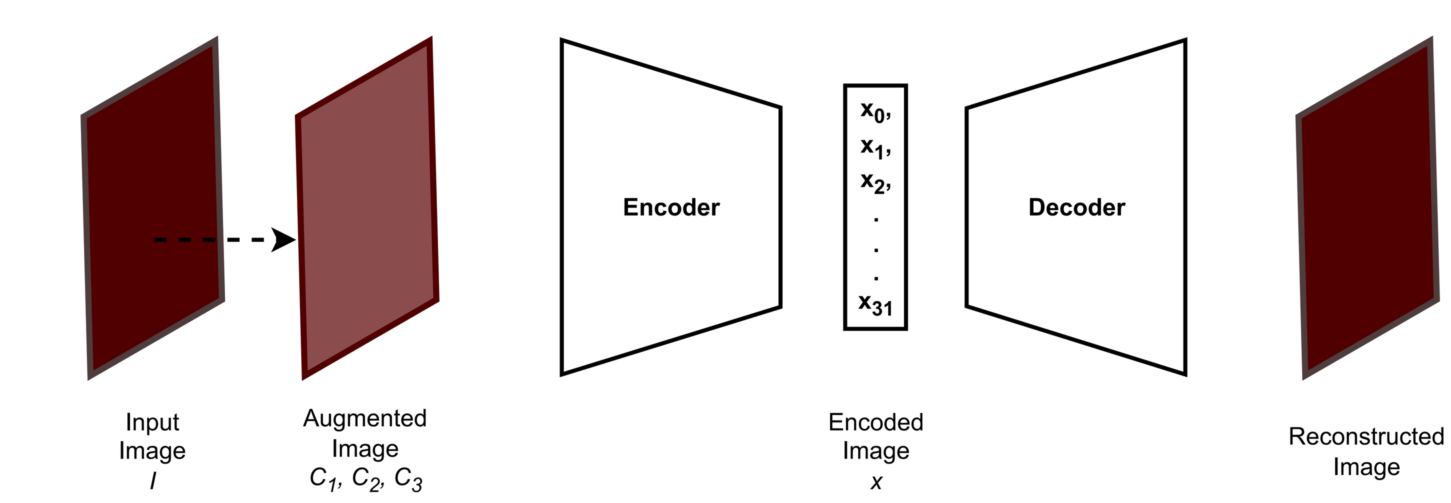
Research Questions

What is the improvement factor in **sample efficiency**, measured by the **number of episodes** required for the AI robot to learn to drive on the track, achieved by implementing **visual domain randomization** with camera sensor images for **sim-to-real transfer** in RL-based end-to-end autonomous racing, compared to standard RL training?

Experimental Settings

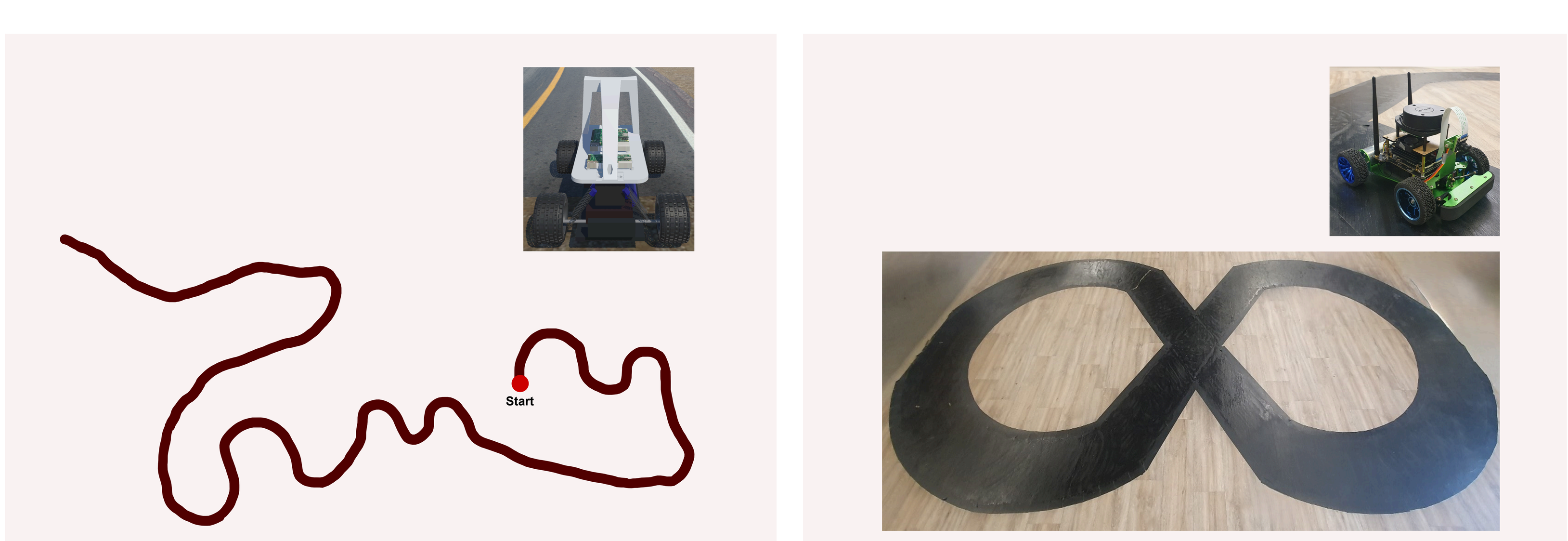
1. Autoencoder (AE) Training with Image Augmentation

This study augments the input images I collected from the AI robot's camera sensor to implement visual domain randomization (visual DR) with three distinct light conditions (C_1, C_2, C_3).[1]



2. Validation and Testing

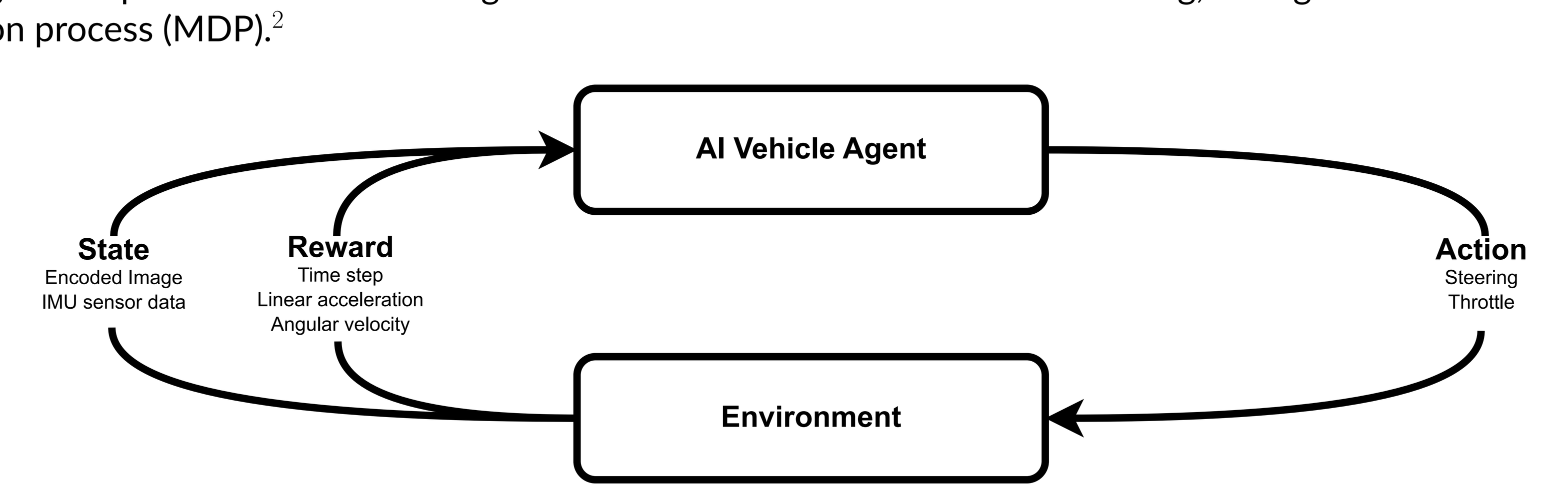
We validated each AE model in the randomly generated road in the DonkeyCar simulator. The model performance was measured by the number of episodes to drive the full track. Then, we deployed the best-performing model in a real-world Figure 8 track setting with JetRacer to validate its successful sim-to-real transfer.



Simulated Environment   Real Environment

Markov Decision Process

During the sequential decision-making interactions in RL-based autonomous racing, the agent follows the Markov decision process (MDP).[2]



In this study, the MDP, denoted by $< S, A, T, \gamma, R >$[3], is defined as:

- S : a set of states where $s_i \in S$ is the encoded image x and Inertial Measurement Unit (IMU) sensor data.
- A : a set of actions where $a_i \in A$ is steering and throttle.
- T : a transition probability function
$$T(s, a, s') = P(s'|s, a) \in (0, 1)$$
where $s, s' \in S$ are the current and new states and $a \in A$ is the taken action, respectively.
- γ : a discount factor
$$\gamma \in (0, 1),$$
which controls how an agent regards future rewards. Low values of γ encourage the agent to maximize short-term rewards, whereas high values of γ cause it to maximize rewards over a longer time frame.
- R : a reward function
$$R(s, a) \rightarrow r(s, a),$$
which is computed by the number of time steps the car stays on track, the change in linear acceleration, and the change in angular velocity.

The MDP goal is to find the optimal policy $\pi^*(a|s)$, which gives the highest expected sum of discounted rewards:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi} \left\{ \sum_{t=0}^{H-1} \gamma^t r_{t+1} | s_0 = s, a_0 = a \right\}$$

for all states $s \in S$ and actions $a \in A$, where $r_t = R(s_t, a_t)$ is the reward at time step t, and H is the maximum number of steps in an episode.

Visual Domain Randomization

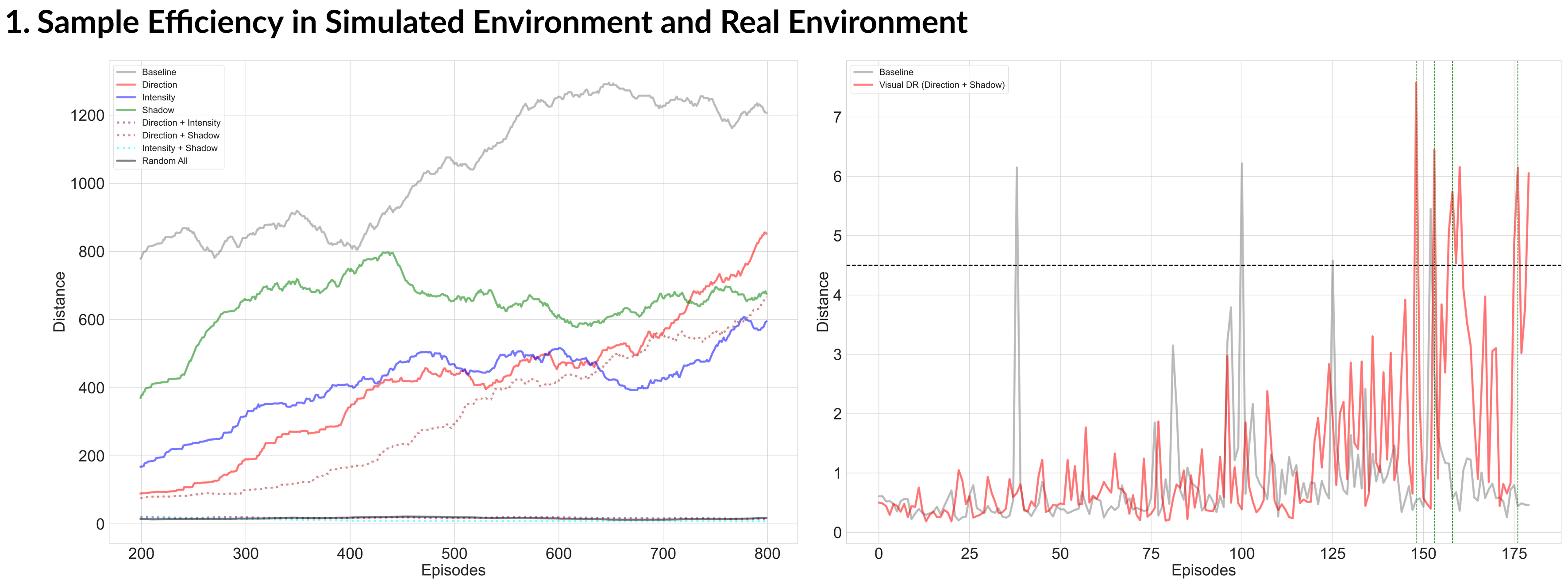
Each AE model gets an RGB image I as an input and outputs an encoded image $x =< x_0, x_1, \cdots, x_{31} >$. Eight AE models are constructed as follows:

Individual Condition Models	Combined Condition Models
$M_1(I)$: Baseline (No Condition)	$M_5(I)$: Light direction and intensity $(C_1 + C_2)$
$M_2(I)$: Light direction condition (C_1)	$M_6(I)$: Light direction and shadow $(C_1 + C_3)$
$M_3(I)$: Light intensity condition (C_2)	$M_7(I)$: Light intensity and shadow $(C_2 + C_3)$
$M_4(I)$: Shadow condition (C_3)	$M_8(I)$: All conditions combined $(C_1 + C_2 + C_3)$

Below are the augmented image exemplars for each light condition:

Results

1. Sample Efficiency in Simulated Environment and Real Environment



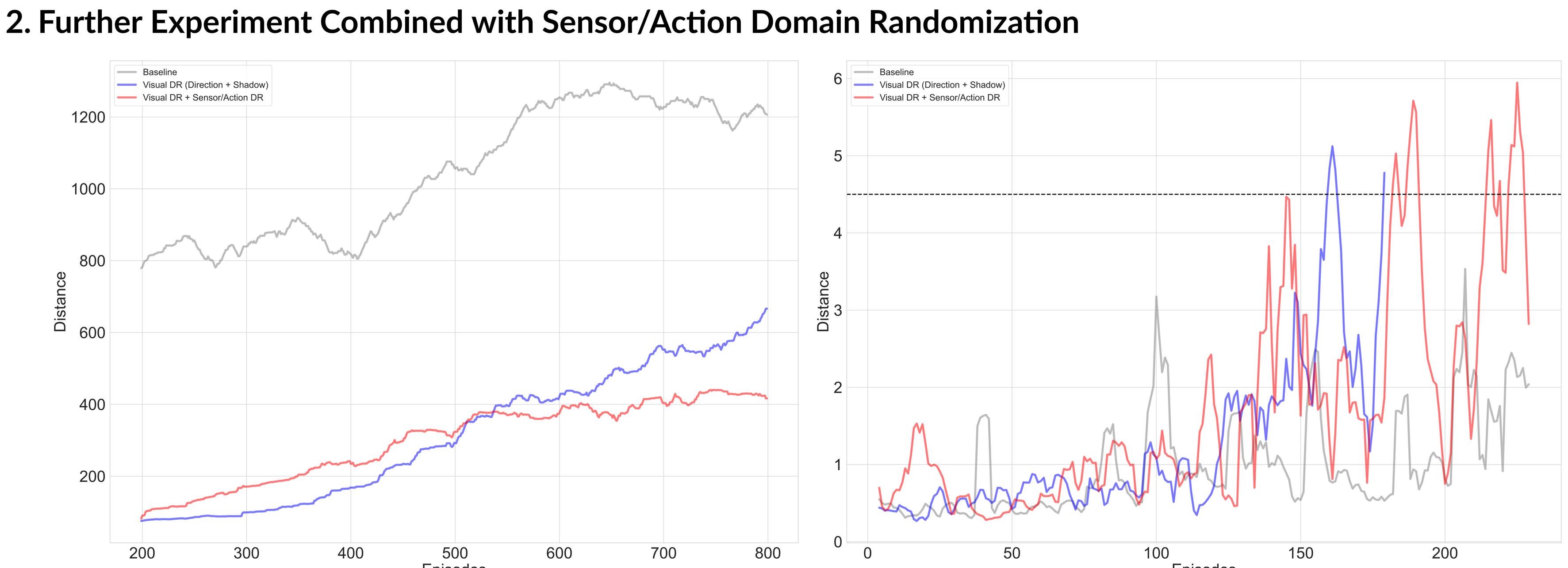
Simulated Environment

- All individual condition models (M_1, M_2, M_3, M_4) successfully learned to navigate the full track.
- If the combined condition models with light intensity condition (C_2) showed no improvement.

Real Environment

- The baseline model M_1 did not show consistent driving after it learned the full track.
- Our model M_6 consistently ran the full track even after we randomly changed the light conditions.

2. Further Experiment Combined with Sensor/Action Domain Randomization



Simulated Environment

- Our model M_6 successfully learned to navigate the full track even when combined with the sensor/action DR model.
- Even though random factors were increased, it showed better performance.

Real Environment

- Our model M_6 remained robust with the sensor/action DR model.
- Not only was the learning speed faster than with a single M_6 model, but it also successfully completed the track after changing lighting conditions.

Conclusion

This study investigates the effects of visual domain randomization (visual DR) by altering various light conditions to enhance the reinforcement learning (RL)-based autonomous racing algorithm. We trained eight autoencoder models, each exposed to three distinct light conditions: light direction, light intensity, and shadow. We found that visual DR can handle random light directions and shadows and perform successful sim-to-real transfer in the model combined with those conditions. We also showed the robustness of our model by adding the sensor and action domain randomization. This proposed approach contributes to more efficient training of RL algorithms for autonomous racing, facilitating their practical application in real-world scenarios.

References

[1] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World," in 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017, pp. 23–30.

[2] R. Bellman, "A Markovian Decision Process," Journal of Mathematics and Mechanics, pp. 679–684, 1957.

[3] B. R. Kiran, I. Sobh, V. Talpaert, P. Mannion, A. A. Al Sallab, S. Yo-gamani, and P. Pérez, "Deep Reinforcement Learning for Autonomous Driving: A Survey," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 4909–4926, 2021.

Acknowledgement

This material is based upon work supported by the National Science Foundation under grant number 2313928. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
I am thankful to Francisco Cavazos and Akhil Kalakota for creating the track for the JetRacer experiment and Meghana Indukuri for sharing the experiment results on sensor and action domain randomization.