Deep Reinforcement Learning for Autonomous Driving by Transferring Visual Features

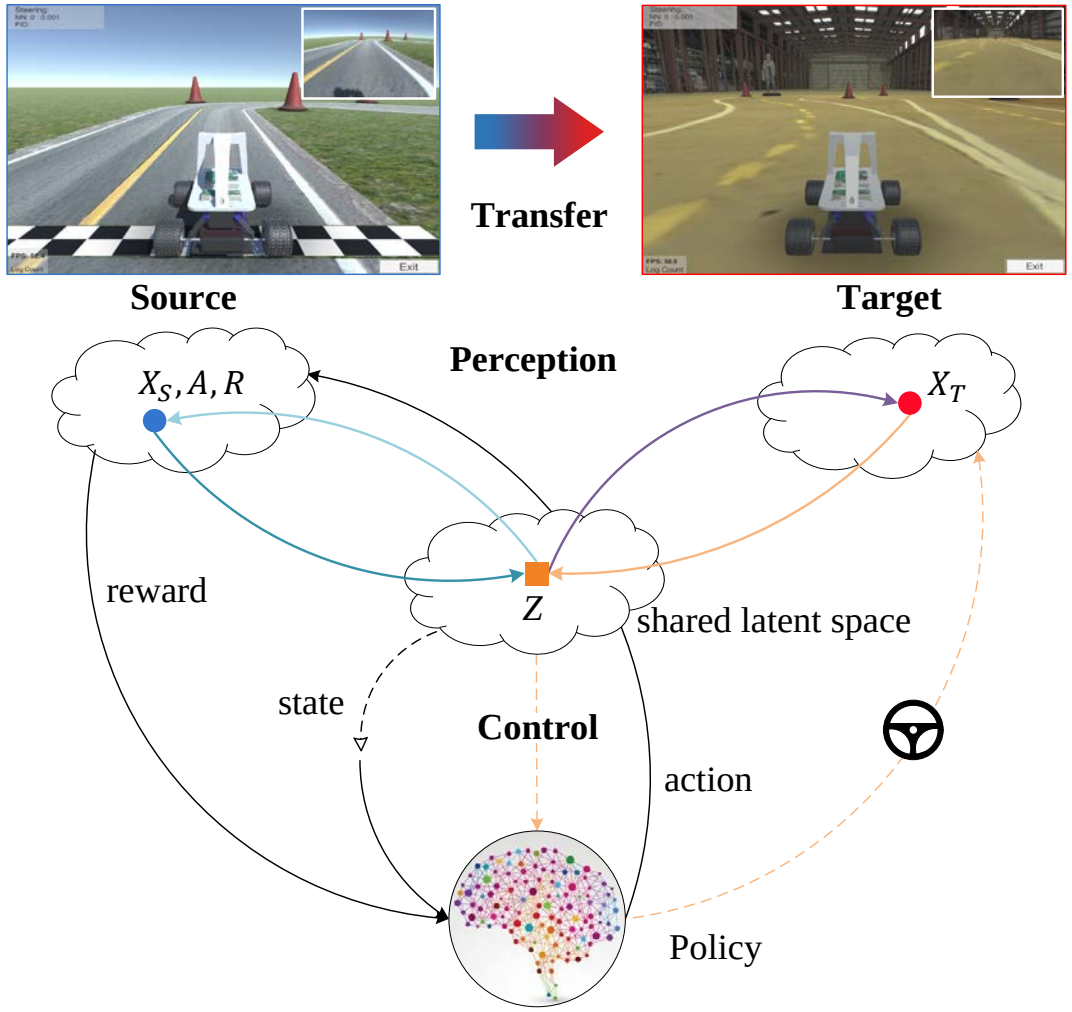
Hongli Zhou[1], Xiaolei Chen[2], Guanwen Zhang[1]*, Wei Zhou[1]
1 School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China
2 CNPC logging Co.,Ltd, China

Reporter: Zhou Hongli

January 10，2021

CONTENTS

Introduction



We propose an effective transfer model that combines adversarial training (Perception) with reinforcement learning (Control).

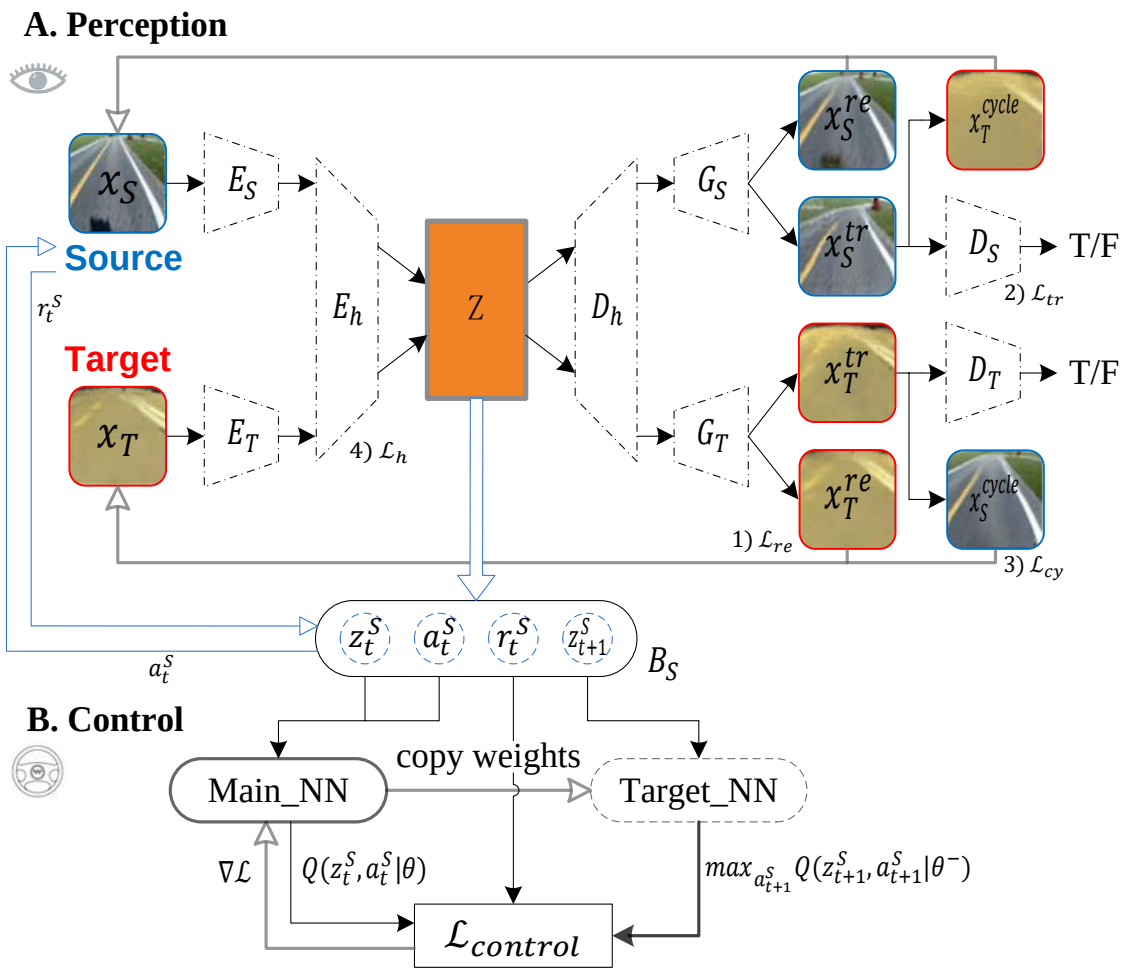
a) Let X_S and X_T be samples from the source and target scenes for perceptual training (colored lines).

b) Define the reward (R) and the action (A) of the source environment for control training (black lines).

c) Perception and control systems are trained synchronously to jointly learns a driving policy from the common latent space Z.

d) This method does not require feedback from the target scene, yet learns a policy which can be applied directly to the target setting (yellow lines).

Method



In this work, we divide the traditional DRL module into two subsystems: perception and control. The perceptual system maps visual images to a shared latent space by various constraints.

a) Reconstruction loss

$$\mathcal{L}_{re} = \mathbb{E}_{x_S\sim P_S}[\| x_S^{re} - x_S \|_1] + \mathbb{E}_{x_T\sim P_T}[\| x_T^{re} - x_T \|_1]$$

b) Translation loss

$$\mathcal{L}_{tr} = \mathbb{E}_{x_S\sim P_S}[(D_T(x_T^{tr}) - c_S)^2] + \mathbb{E}_{x_T\sim P_T}[(D_S(x_S^{tr}) - c_T)^2]$$

c) Cycle consistency loss

$$\mathcal{L}_{cy} = \mathbb{E}_{x_S\sim P_S}[\| x_S^{cy} - x_S \|_1] + \mathbb{E}_{x_T\sim P_T}[\| x_T^{cy} - x_T \|_1]$$

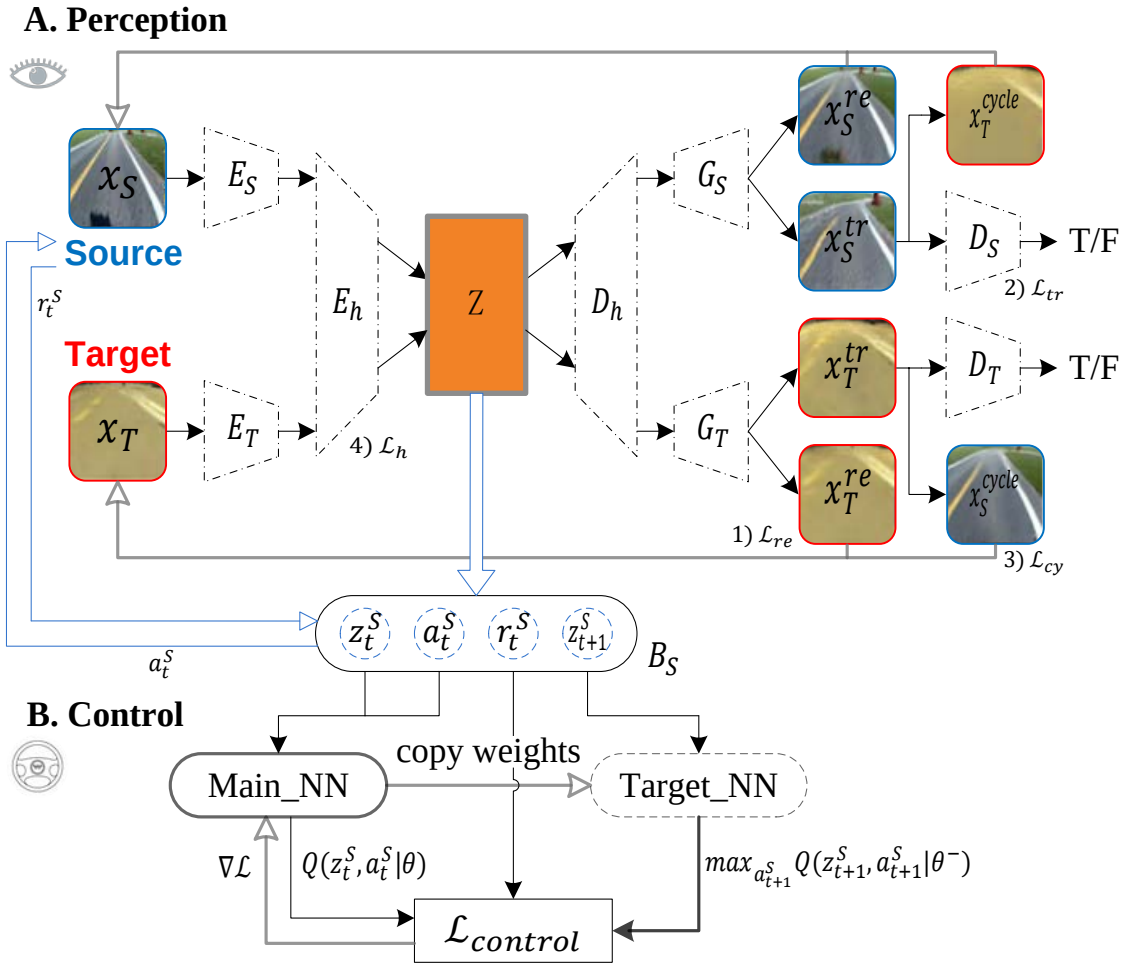
d) Weight sharing

$$\mathcal{L}_h = \mathbb{E}_{x_S\sim P_S}[(E_h(E_S(x_S)) + \eta)^2] + \mathbb{E}_{x_T\sim P_T}[(E_h(E_T(x_T)) + \eta)^2]$$

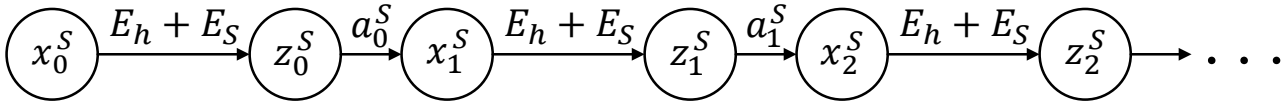
Finally, by combining these individual losses we define the perceptual loss to be:

$$\mathcal{L}_{perceptual} = \lambda_{re}\mathcal{L}_{re} + \lambda_{tr}\mathcal{L}_{tr} + \lambda_{cy}\mathcal{L}_{cy} + \lambda_h\mathcal{L}_h$$

Method



The control system trains driving policy with reinforcement learning in the low-dimensional state. The dynamic process of MDP in the control system:



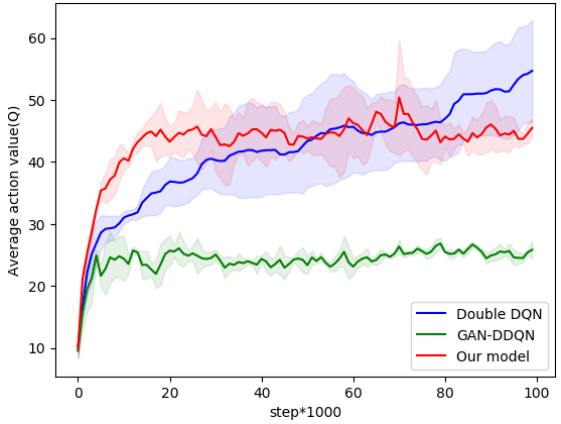
a) Define the future accumulated reward function as:

$$Q^{\pi}(z_t^S, a_t^S) = E\left[r_{t+1}^S + \gamma r_{t+2}^S + \gamma^2 r_{t+3}^S + \cdots \middle| z_t^S, a_t^S\right]$$

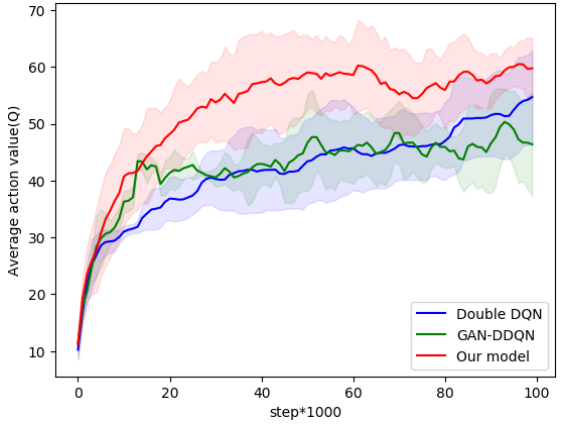
b) Extend Deep Q-Network to design the control module and stabilize the training of driving policy using experience replay and target network:

$$\mathcal{L}_{control} = E[(r_t^S + \gamma max_{a_{t+1}^S} Q(z_{t+1}^S, a_{t+1}^S|\theta^-) - Q(z_t^S, a_t^S|\theta))^2]$$

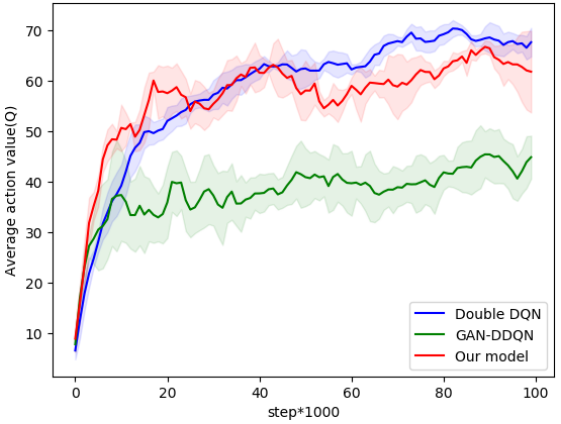
Transfer Results



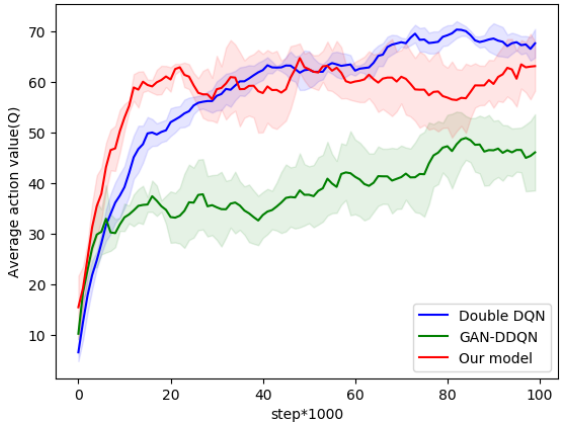
（1）Scene1→Scene2



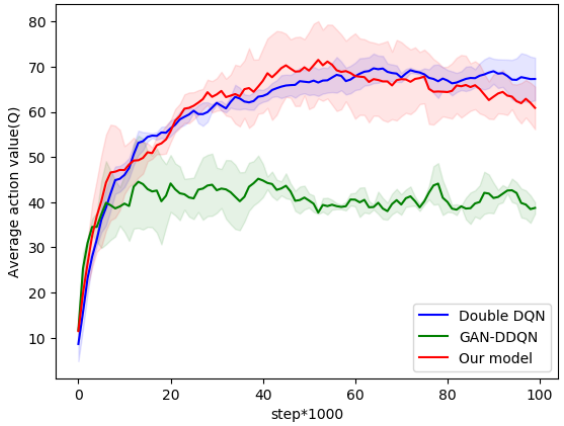
（2）Scene1→Scene3



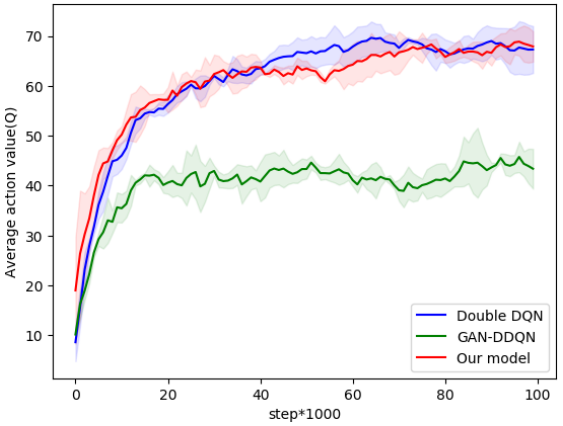
（3）Scene2→Scene1



（4）Scene2→Scene3



（5）Scene3→Scene1



（6）Scene3→Scene2

The Double DQN, GAN-DDQN, and Our model training curves tracking the agents average maximum predicted action-value of six transfer tasks.

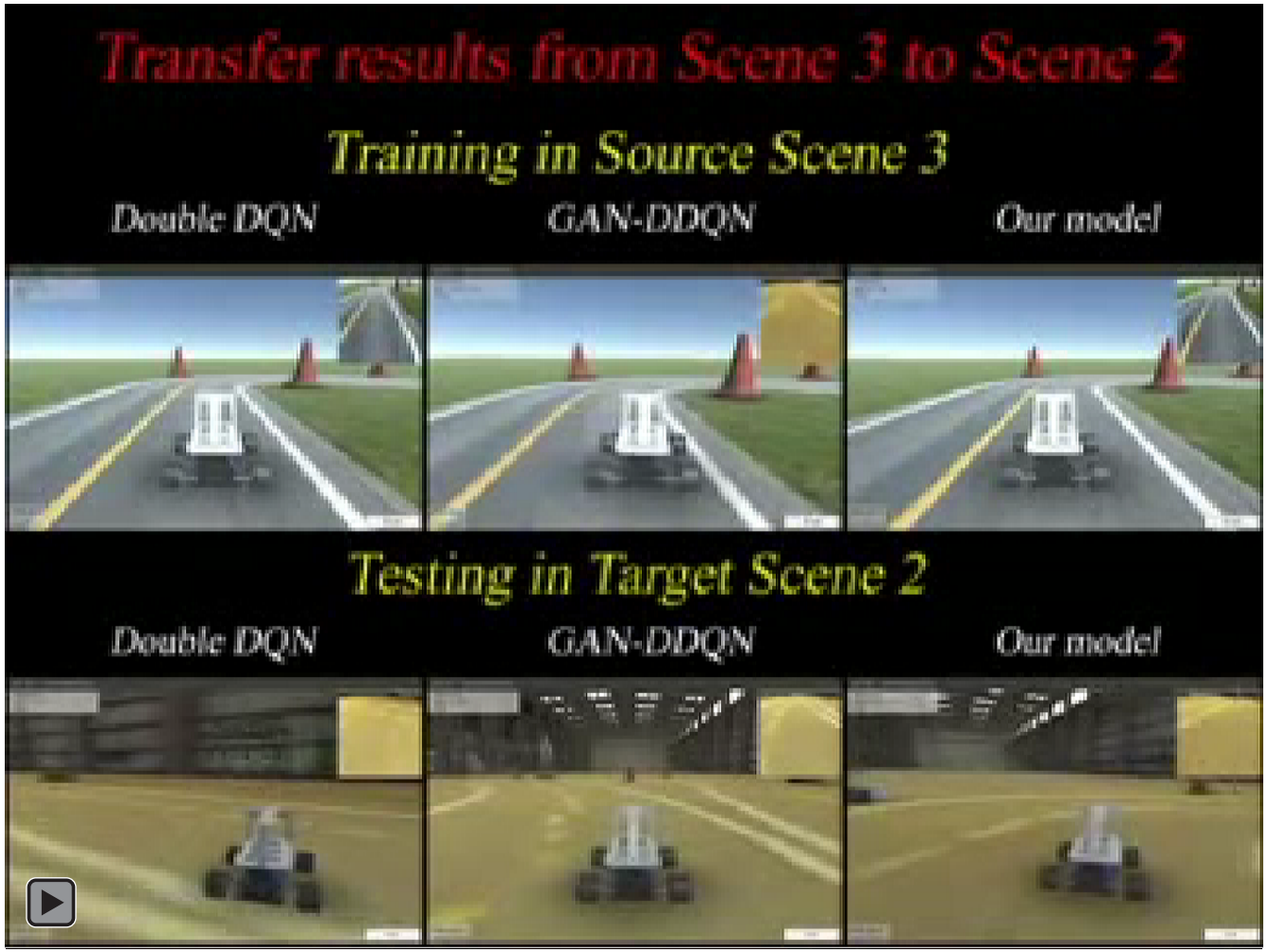
Our model has a good initial performance and converges fastest. It does not over-extract the features of the source environment, and learns a general policy that can adapt to the target scene.

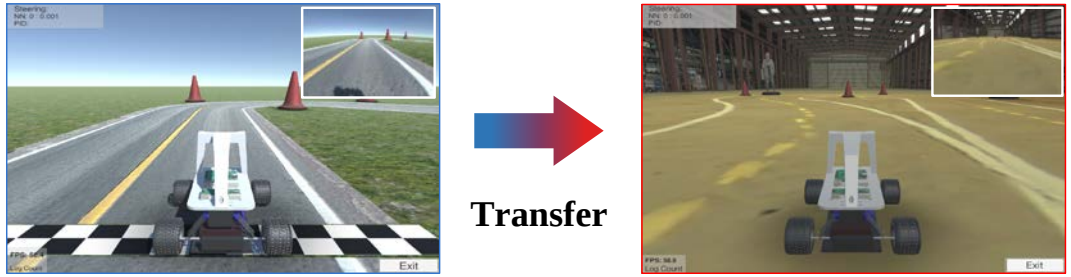
Transfer Results

TRANSFER RESULTS BETWEEN DIFFERENT SCENES

Transfer Results		Double dqn		Gan-ddqn		Our model	
		T R	Laps%	T R	Laps%	T R	Laps%
Scene1→Scene2	Source	6962.2	100%	1742.2	23%	6081.7	100%
	Target	128.8	2.4%	353.1	6.2%	**5844.4**	**100%**
Scene1→Scene3	Source	7022.5	100%	2916.4	39%	6114.8	100%
	Target	261.4	5.4%	**4803.8**	**100%**	3769	100%
Scene2→Scene1	Source	6543.4	100%	2960.2	44%	7231.6	100%
	Target	951.8	20%	3391.4	91%	**5746.5**	**100%**
Scene2→Scene3	Source	7166.4	100%	94.8	2%	1870.3	38.5%
	Target	212.7	3%	5082.5	100%	**5195.8**	**100%**
Scene3→Scene1	Source	2363.4	50%	1100.7	20%	48591	100%
	Target	4603.1	73%	5447.7	79%	**7381.4**	**100%**
Scene3→Scene2	Source	2773.7	57%	125.6	3%	5625.5	100%
	Target	92.6	1.5%	736.8	12%	**6419.6**	**100%**

Transfer Results





Source Target

This video shows the transfer results from scene 3 to scene 2 of the simulator clearly. The driving policy trained in the source scene can be directly applied to the target scene without fine-tuning. Our model performs stable in both source and target environments.

References

① V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,”arXiv preprint arXiv:1312.5602, 2013

② V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, 2015.

③ L. T. P. Yun, Y. Chen, C. Liu, and H. Y. M. Liu, “Visual-based autonomous driving deployment from a stochastic and uncertainty-aware perspective.”

④ X. Zhang, M. Zhou, H. Liu, and A. Hussain, “A cognitively inspired system architecture for the mengshi cognitive vehicle,” Cognitive Computation, vol. 12, no. 1, pp. 140–149, 2020.

⑤ G. Kahn, A. Villaflor, B. Ding, P. Abbeel, and S. Levine, “Selfsupervised deep reinforcement learning with generalized computation graphs for robot navigation,” in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp. 1–8.

⑥ R. S. Sutton, A. G. Barto et al., Introduction to reinforcement learning. MIT press Cambridge, 1998, vol. 135.

⑦ A. Ganesh, J. Charalel, M. D. Sarma, and N. Xu, “Deep reinforcement learning for simulated autonomous driving,” 2016.

⑧ N. Xu, B. Tan, and B. Kong, “Autonomous driving in reality with reinforcement learning and image translation,” arXiv preprint arXiv: 1801.05299, 2018.

⑨ K. Cobbe, O. Klimov, C. Hesse, T. Kim, and J. Schulman, “Quantifying generalization in reinforcement learning,” arXiv preprint arXiv: 1812.02341, 2018.

⑩ C. Asawa, C. Elamri, and D. Pan, “Using transfer learning between games to improve deep reinforcement learning performance and stability.”

⑪ H. Sui, W. Shang, and X. Li, “Transfer of robot perception module with adversarial learning,” IEEE Access, vol. 7, pp. 79 726–79 736, 2019.

⑫ M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in Advances in neural information processing systems, 2017, pp. 700–708.

⑬ Z. Murez, S. Kolouri, D. Kriegman, R. Ramamoorthi, and K. Kim, “Image to image translation for domain adaptation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4500–4509.

⑭ E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, “Adversarial discriminative domain adaptation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7167–7176.

THANKS FOR YOUR ATTENTION