Virtual to Real Reinforcement Learning for Autonomous Driving

Xinlei Pan1, Yurong You2, Ziyan Wang1, Dewu Lu2

1University of California, Berkeley, 2Shanghai Jiao Tong University, *indicates equal contribution

Introduction

Reinforcement learning is considered as a promising direction for driving policy learning. However, •Training autonomous driving vehicles with reinforcement learning in real environment involves non-affordable trial-and-error. •One can first train the model in a virtual environment and then apply it in the real world. But gaps exist between the real and virtual environment, rendering the transferring intractable.

In this paper, we propose a novel realistic translation network to make model trained in virtual environment more adaptable to real world scenario. The proposed framework exploits a key relation between non-realistic virtual images and real images: they all have similar scene structure. Though lacking proper paired data labeling virtual images with corresponding real images, their segmentation can serve as a proxy: we can first convert the virtual images into their scene segmentations, then convert the segmentations into real images, where paired training data are available in both part.

Experiments show that our proposed virtual to real (VR) reinforcement learning (RL) works pretty well. To our knowledge, this is the first successful case of driving policy trained by reinforcement learning that can adapt to real world driving data.

Methods

The overall pipeline is shown in Figure 2.

Inspired by [1], our realistic translation network is composed of two image translation networks: virtual-to-segmentation network: translating virtual images to their segmentations segmentation-to-real network: translating segmented images to their real world counterparts

These two networks are basically conditional GANs. The objective of a conditional GAN can be expressed as,

\[ L_{cGAN}(G, D) = \mathbb{E}_{s \sim \mathcal{D}_{data}}[\log D(s)] + \mathbb{E}_{z \sim \mathcal{N}(0, 1)}[\log (1-D(G(z)))] \]

where \( G \) is the generator that tries to minimize this objective and \( D \) is the adversarial discriminator that acts against \( G \) to maximize this objective. In order to suppress blurring, a \( \ell_1 \) loss regularization term is added. Therefore, the overall objective for the image-to-image translation network is,

\[ L_{cGAN}(G, D) = \mathbb{E}_{s \sim \mathcal{D}_{data}}[\log D(s)] + \lambda_1 \mathbb{E}_{z \sim \mathcal{N}(0, 1)}[\log (1-D(G(z)))] + \lambda_2 \mathbb{E}_{z \sim \mathcal{N}(0, 1)}[\|G(z) - s\|_1] \]

where \( \lambda_1 \) and \( \lambda_2 \) are the weight of regularization.

These two networks are both trained towards the above defined generator, where first network translates virtual images \( s \) to their segmentations \( \hat{s} : \{s\} \rightarrow \hat{s} \) and the second network translates segmented images \( \hat{s} \) into their real counterpart \( y : \{\hat{s}\} \rightarrow y \), where \( y \) is a network that learn to be deterministic outputs. We use the same generator and discriminator architectures as used in [1].

After training of the realistic translation network is finished, we train our reinforcement learning model by using it to filter virtual images to synthetic realistic images and feed these realistic images as state inputs.

For more details about the proposed architecture, including network settings of translation networks and reinforcement learning models, please refer to our paper.

Results

We performed two sets of experiments to compare the performance of our method and other reinforcement learning methods as well as supervised learning methods:

1. virtual to reinforcement learning on real world driving data.
2. transfer learning in different virtual driving environments.

The virtual simulator used in our experiments is TORCS [4]. A total of 1673 images were collected which covers the entire driving cycle of Aalborg environment.

Data

For Segmentation-to-real network:
The real world driving video data are from [2], which is collected in a sunny day with detailed steering angle annotations per frame. We used the image segmentation network design of [3] and their trained segmentation network on the CityScape image segmentation dataset to segment 45k real world driving images from [2]. The network was trained on the CityScape dataset with 11 classes and was trained with 30000 iterations.

For Segmentation-to-real network:
We collected virtual images and their segmentations from the Aalborg environment in TORCS [4].

Accuracy of real world driving

<table>
<thead>
<tr>
<th>Method</th>
<th>OURS</th>
<th>B-RL</th>
<th>SV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>43.40%</td>
<td>28.33%</td>
<td>53.60%</td>
</tr>
</tbody>
</table>

Results show that our proposed method outperforms the baseline method (B-RL), where the reinforcement training agent is trained in a virtual environment without seeing any real data. The supervised method (SV) has the best overall performance, however, was trained with large amounts of supervised labeled data.

Transfer learning in different virtual environments

In this experiment, both the baseline (Randomization Method) and our model are trained in the City--track2 track and evaluate in E-track1 track, which has different visual appearance.

Obviously, standard A3C (Oracle) trained and tested in the same environment gets the best performance. However, our model performs better than the domain randomization method, which requires training in multiple environments to generalize.

Reference