First attempt to build realistic driving scenes using video-to-video synthesis in OpenDS framework

Zili Song, Shuolei Wang, Xiangjun Peng, Weikai Kong, Xu Sun

University of Nottingham
UK | CHINA | MALAYSIA

University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2019

University of Nottingham
UK | CHINA | MALAYSIA

First Attempt to Build Realistic Driving Scenes using Video-to-video Synthesis in OpenDS Framework

Zili Song*
Shuolei Wang*
University of Nottingham
Ningbo, Zhejiang
zy22063@nottingham.edu.cn
zy22067@nottingham.edu.cn

Xiangjun Peng
University of Nottingham
Ningbo, Zhejiang
zy22056@nottingham.edu.cn

Weikai Kong
University of Nottingham
Ningbo, Zhejiang
scywk1@nottingham.edu.cn

Xu Sun
University of Nottingham
Ningbo, Zhejiang
xu.sun@nottingham.edu.cn

Abstract
Existing programmable simulators enable researchers to customize different driving scenarios to conduct in-lab automotive driver simulations. However, software-based simulators for cognitive research generate and maintain their scenes with the support of 3D engines, which may affect users' experiences to a certain degree since they are not sufficiently realistic. Now, a critical issue is the question of how to build scenes into real-world ones. In this paper, we introduce the first step in utilizing video-to-video synthesis, which is a deep learning approach, in OpenDS framework, which is an open-source driving simulator software, to present simulated scenes as realistically as possible. Off-line evaluations demonstrated promising results from our study, and our future work will focus on how to merge them appropriately to build a close-to-reality, real-time driving simulator.

Author Keywords
Video Synthesis; Driving Simulator; Machine Learning;

* stands for equal contributions

AutomotiveUI '19 Adjunct, September 21–25, 2019, Utrecht, Netherlands

ACM ISBN 978-1-4503-6920-6/19/09.
https://doi.org/10.1145/3349263.3351497

Figure 1: A 2D building "sticker".

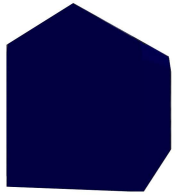

Figure 2: A 3D building example.

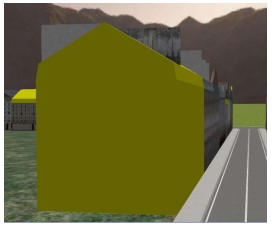

Figure 3: An example of the scene without the "sticker".

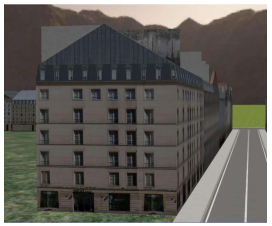

Figure 4: An example of the scene with the "sticker".

Introduction

Existing programmable simulators enable researchers to customize different driving scenarios to conduct in-lab automotive driver simulations. However, software-based simulators for cognitive research generate and maintain their scenes with the support of 3D engines, which may affect users' experiences to a certain degree since they are sufficiently realistic. Now, a critical issue is the question of how to present scenes which are more realistic.

In this paper, we introduce our work-in-progress, which is to build a close-to-reality, real-time cognitive driving simulator to enhance the user experience while undertaking in-lab studies. We first generated several video pieces from OpenDS framework, which is a free, portable and open-source driving simulator [5]. We then transformed those pieces into colored frames based on labeling policy. Finally, We took the video-to-video synthesis, which is a deep learning approach in Computer Vision, as a subsystem to build realistic scenes [6]. Off-line evaluations demonstrated both promising results and outstanding challenges from video-to-video synthesis. Our future work will focus on how to merge them appropriately to be realistic.

Motivation

In this section, we have taken an example to illustrate what was our motivation by analyzing the functionality of OpenDS and its drawbacks.

Before launching OpenDS, a set of images were stored in advance for further scene generation, an example of which is shownin Figure 1. Figure 2 shows an example of a 3D building model, while Figure 3 demonstrates this model in a scenario. Then, OpenDS built the scene by pasting these "stickers"[1] onto 3D building model. Finally, it appears in the simulated scenario and rotated with the changes of view, as shown in Figure 4.

OpenDS has provided freedom for researchers to build and customize their scenarios. However, the rotations of simulated scenes made a original clear image become vague with the changes of the visual field. In particular, buildings by roads suffered from this mostly.

Our Approach: Video-to-Video Synthesis

We chose Video-to-Video Synthesis (vid2vid) to generate realistic scenes. We aimed to achieve a close-to-reality simulation for users, with the minimal adjustments in OpenDS framework to keep its original features. There are three reasons that we chose vid2vid.

First, vid2vid is the most **suitable** framework for video generations. Previous work on building realistic scenes limited its practicality since they applied Image-to-Image synthesis [3], which led to drifts in video flow while emerging images into one video.

Second, vid2vid is an **extensible** framework for different demands of scenes. The current versions of Vid2vid relied on Cityscape, a open-sourced high-resolution data set on Germany street views when driving [1]. It's applicable to be generalized into different places as needed, when there are data support, like Apolloscape (i.e. similar dataset for street views in China) [2].

Third, vid2vid is a **portable** framework. Most implementations of vid2vid were done on PyTorch, a

[1] refers to those 2D building images, like Figure 1.

Figure 5: An standard example of labeled version from vid2vid. Please note that the actual ones used in training are in Grayscale.

Figure 6: An standard example of a realistic driving scene built from vid2vid.

System Configurations

All the implementations and experiments were conducted on a server remotely, with 3 cores and 20G memory. We used Python 3.5.2 and PyTorch 0.4.0. Also, we imported a TITAN X (xl) GPU to support vid2vid. Our host OS is MacOS 10.14.1 and guest OS is Ubuntu 16.04.

cross-platform open-sourced machine learning system [4]. This feature allowed vid2vid to be embedded with OpenDS without resetting Operating Systems.

Our approach works as follow: first, we transformed the simulated forms into labeled versions, as highlighted in Figure 5. Then, it applied vid2vid to create the realistic driving scenes, as shown in Figure 6, from the labeled versions.

Experimental Design

We conducted our experimental study to show the effects of our approach in four steps. First, we selected the driving scenario "Paris", which is a standard scenar-io from OpenDS, and programmed a specific route for the driving simulator to drive automatically. Then, we performed the same routes under three different environmental settings (i.e. Sunny, Rainy and Night time) and recorded them. Next, we trained those videos in Grayscale versions. Finally, we produced the results via vid2vid framework.

The experimental procedure were shown step by step in Figure 7, Figure 8 and Figure 9, which shows the simulated, labeled and synthesized versions respectively. In each figure, the left one was driving in the sunny day, the middle one was driving in the rainy day and the right one was driving at night.

Figure 7: Three examples of driving scenes in OpenDS framework, which are with different environmental settings of a sunny day, a rainy day and at night (from left to right).

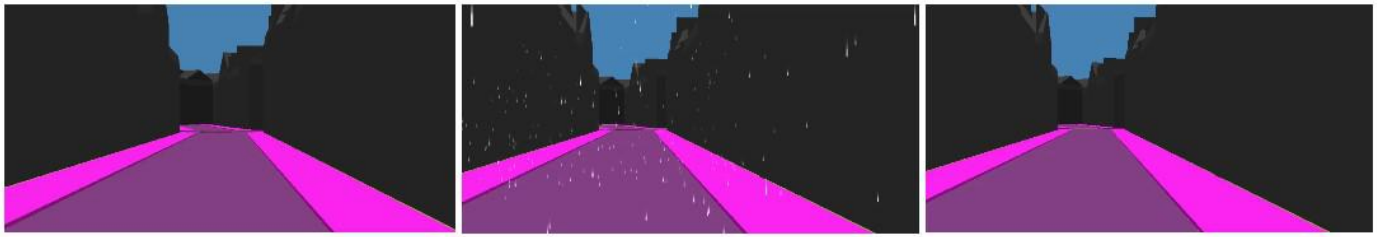

Figure 8: Three images in labeled versions after transformations from Figure 7, which are with different environmental settings of a sunny day, a rainy day and at night (from left to right).

Figure 9: Final results after processing the driving scenes in Figure 8 via vid2vid, which are with different environmental settings of a sunny day, a rainy day and at night (from left to right).

Preliminary Results

Our preliminary results show the overall quality to uild realistic driving scenes using vid2vid is acceptable For example, first, the left one in Figure 9 shows a pretty realistic road scene. However, there are still two issues to be further explored. First, the effects of presenting distant parts of scenes is not well. Second, the edges of the visual field are not very clear too. These two aspects observed may due to the different resolutions ratio between two series of images[2].

The rest of Figure 9 demonstrated relatively poor effects while changing environmental settings. The middle one shows that, raindrops blur the driving scene, which resulted in poor clarity. The right one showed that the night could not be synthesized. This is because the original data set for training the model, which is used to perform Video-to-video Synthesis, didn't contain the situations while driving at night.

[2] one refers to the series of recorded images from OpenDS, and the other refers to the series of images from supporting data set

Discussions

Based on our preliminary results, we summarize two major directions for further optimization, including:

Optimization under different environmental settings. Existing results showed that our approach couldn't be extended to driving scenes under different environmental settings. We planned to optimize it increasing several sample images for model training.

Optimization on Edges. Our results showed that our approach could sketch the street views while driving but couldn't perform well in the edges of the visual field. We plan to optimize it by increasing the differences between classes' labels, which may avoid too much confusions while training the model.

Conclusion and Future Work

In this paper, we explored the usages of vid2vid within OpenDS framework. The preliminary results showed its promising futures and outstanding challenges. Our future work would focus on the optimization of applying vid2vid in OpenDS to build realistic driving scenes

Acknowledgements

We thanked for anonymous reviewers for their valuable feedback. This work is generously supported by the funding body of Ningbo Creative Industry Park and Summer Research Program in University of Nottingham Ningbo China.

References

1. Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR '16), 3213-3223.
2. Xinyu Huang, Xinjing Cheng, Qichuan Geng, Binbin Cao, Dingfu Zhou, Peng Wang, Yuanqing Lin, and Ruigang Yang. 2018. The apolloscape dataset for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (ECCV '18), 954-960.
3. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (CVPR '17), 1125-1134.
4. Nikhil Ketkar. 2017. Introduction to pytorch. In *Deep learning with python*. Apress, Berkeley, CA, 195-208.
5. Rafael Math, Angela Mahr, Mohammad M. Moniri, and Christian Müller. 2012. OpenDS: A new open-source driving simulator for research. In *Adjunct roceedings of the International Conference on Automotive User Interfaces and Interactive Vehicular Application* (Adjunct AutoUI '12), 7-8.
6. Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Nikolai Yakovenko, Anderw Tao, Jan Kautz and Bryan Catanzaro. 2018. Video-to-Video Synthesis. In *Proceedings of the Annual Conference on Neural Information Processing Systems* (NIPS '18), 1152-1164.