Realistic Post-processing of Rendered 3D Scenes

Anastasia Feygina
National Research University
Higher School of Economics
Moscow, Russia
a.g.feygina@gmail.com

Dmitry I. Ignatov
National Research University
Higher School of Economics
Moscow, Russia
dignatov@hse.ru

Ilya Makarov
National Research University
Higher School of Economics
Moscow, Russia
iamakarov@hse.ru

Figure 1: Upgrading GTA SA (2004) graphics (left) to GTA V (2013) level (right) with proposed CycleWGAN deep neural network.

ABSTRACT

In this talk, we show a realistic post-processing rendering based on generative adversarial network CycleWGAN. We propose to use CycleGAN architecture and Wasserstein loss function with additional identity component in order to transfer graphics from Grand Theft Auto V to the older version of GTA video-game, Grand Theft Auto: San Andreas. We aim to present the application of modern art style transfer and unpaired image-to-image translations methods for graphics improvement using deep neural networks with adversarial loss.

CCS CONCEPTS

• **Computing methodologies → Image processing**; **Image-based rendering**; *Scene understanding*; Appearance and texture representations;

KEYWORDS

Computer Graphics, Post-processing Rendering, GAN

ACM Reference Format:
Anastasia Feygina, Dmitry I. Ignatov, and Ilya Makarov. 2018. Realistic Post-processing of Rendered 3D Scenes. In *Proceedings of SIGGRAPH '18 Posters.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3230744.3230764

SIGGRAPH '18 Posters, August 12-16, 2018, Vancouver, BC, Canada

ACM ISBN 978-1-4503-5817-0/18/08.
https://doi.org/10.1145/3230744.3230764

1 INTRODUCTION

Rendering high-quality graphics usually depends on using standard rendering techniques including processing geometry, materials, and light information from the scene. Although, the existing graphics algorithms preform well on such tasks, creating new virtual environments or upgrading graphics built in old-school game engine is both, expensive and time-consuming.

There are several approaches such as neural style transfer [Gatys et al. 2016] or generative adversarial networks [Goodfellow et al. 2014]. The first one has the problem with choosing a single "style" for all the areas of a desired realistic image, while the second one has demonstrated noisiness and blurs when trained for image generation. Then, the idea of image-to-image translation (started in Hertzmann's work [Hertzmann et al. 2001]), which infers a non-parametric texture model from trained mapping two image datasets. We also suggest using a post-processing algorithm able to render realistic high-quality graphics by means of a learning model inferring new graphics over the existing rendering. We could adjust the graphics to a new virtual scene created in low-cost manner with basic graphics and texture quality via training a generative model on new datasets of images or their semantic representation rather than modeling a new virtual scene. Using semantic segmentation, the proposed approach could present the solution for diminishing reality, while editing the corresponding segmented image and then generating a new rendering based on editing within segmentation domain.

2 CYCLEWGAN FOR GRAPHICS POST-PROCESSING

In this section, we discuss a new approach that produces high-quality graphics from two datasets of images taken from two versions of Grand Theft Auto (GTA) video-games. We have collected a dataset from GTA: San Andreas [Rockstar North 2004] released on

October 26th 2004 and containing old generation graphics. Our goal was to train the model on image-to-image translation to the images from the newer version of video game GTA: V [Rockstar Games 2013] released on September 17th 2013 with advanced AAA graphics and the same original place of Los Angeles city.

We use unpaired image-to-image translation approach where the goal is to learn the mapping G between a source domain X and a target domain Y, such that the distribution of images from $range(G)$ is indistinguishable from the distribution Y in terms of adversarial loss, while also making indistinguishable the distributions for X and $range(F)$ for inverse mapping $F = G^{-1}$ and adding cycle consistency loss [Isola et al. 2016]. We use Wasserstein loss for both mappings instead of usual adversarial loss in order to stabilize training process and achieved better realistic post-processing, while adding identity loss in order to preserve the color style palette.

$$L_{D_Y} = | D_Y(G(x)) - D_Y(y) |$$
$$L_{D_X} = | D_X(F(y)) - D_X(x) |$$
$$L_{CYC}(G,F) = || F(G(x)) - x ||_1 + || G(F(y)) - y ||_1$$
$$L_{id}(G,F) = E_{x \sim X} \, || F(x) - x ||_1 + E_{y \sim Y)} \, || G(y) - y ||_1$$
$$L_G(G,F,D_X,D_Y) = L_{D_X} + L_{D_Y} + \lambda L_{CYC}(G,F) + \mu L_{id}(F,G)$$

We compare our method with several GAN architectures, such as CoGAN [Liu and Tuzel 2016], DiscoGAN [Kim et al. 2017] and CycleGAN [Zhu et al. 2017] obtaining more robust and better looking post-processing rendering (see Figure 2).

Figure 2: Two original GTA San Andreas images (bottom) are translated using CoGAN (top left), DiscoGAN (top right), CycleGAN (middle left) and our CycleWGAN (middle right).

3 DISCUSSION

We have suggested totally new approach for realistic image post-processing problem using CycleGAN framework with the new loss function inspired by Wasserstein GAN. We have also collected new large unpaired data set for conducting relevant experiments in video game area and presented experiments showing neural networks competitiveness in image post-processing task using real-time image-to-image translation. The results for average speed of post-processing is 41.75 frames/sec[1] making GANs applicable to real-time image translation, having comparable results with

- CoGAN 44 FPS,
- DiscoGAN 35 FPS,
- CycleGAN 47 FPS, and
- CycleWGAN 41 FPS.

We will compare our approach with the other pix2pix architectures taking into account semantic information of rendered scenes [Chen and Koltun 2017; Wang et al. 2017]. We aim to further improve the results decomposing rendering graphics on several independent layers such as textures [Sun and Hays 2017], illumination [Gardner et al. 2017], scene understanding [Wang et al. 2017], and applying image-to-image translation to video processing improving the robustness of the algorithm taking the optical flow into account [Gupta et al. 2017]. We also need to reduce the training time (currently several days) and to learn how to seamlessly remove in-game labels, writings, and subtitles. We are also planning to evaluate our approach on Open source collected dataset.

ACKNOWLEDGEMENTS

The work of D.I. Ignatov and I. Makarov was supported by the Russian Science Foundation under grant 17-11-01294 and performed at National Research University Higher School of Economics, Russia.

REFERENCES

Qifeng Chen and Vladlen Koltun. 2017. Photographic image synthesis with cascaded refinement networks. In *International Conference on Computer Vision (ICCV)*, Vol. 1. IEEE.

Marc-André Gardner et al. 2017. Learning to Predict Indoor Illumination from a Single Image. *arXiv preprint arXiv:1704.00090* (2017).

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2414–2423.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. NIPS, 2672–2680.

Agrim Gupta, Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2017. Characterizing and Improving Stability in Neural Style Transfer. *arXiv preprint arXiv:1705.02092* (2017).

Aaron Hertzmann, Charles E Jacobs, Nuria Oliver, Brian Curless, and David H Salesin. 2001. Image analogies. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 327–340.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2016. Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004* (2016).

Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jungkwon Lee, and Jiwon Kim. 2017. Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192* (2017).

Ming-Yu Liu and Oncel Tuzel. 2016. Coupled generative adversarial networks. In *Advances in neural information processing systems*. NIPS, 469–477.

Rockstar North Rockstar Games. 2013. Grand Theft Auto: V. http://www.rockstargames.com/V/. [Online; accessed 17-09-2013].

War Drum Studios Rockstar North. 2004. Grand Theft Auto: San Andreas. http://www.rockstargames.com/sanandreas/. [Online; accessed 26-10-2004].

Libin Sun and James Hays. 2017. Super-resolution Using Constrained Deep Texture Synthesis. *arXiv preprint arXiv:1701.07604* (2017).

Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2017. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *arXiv preprint arXiv:1711.11585* (2017).

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593* (2017).

[1] Real-time video processing is available via the link: https://youtu.be/Rn9TxBDQzAw