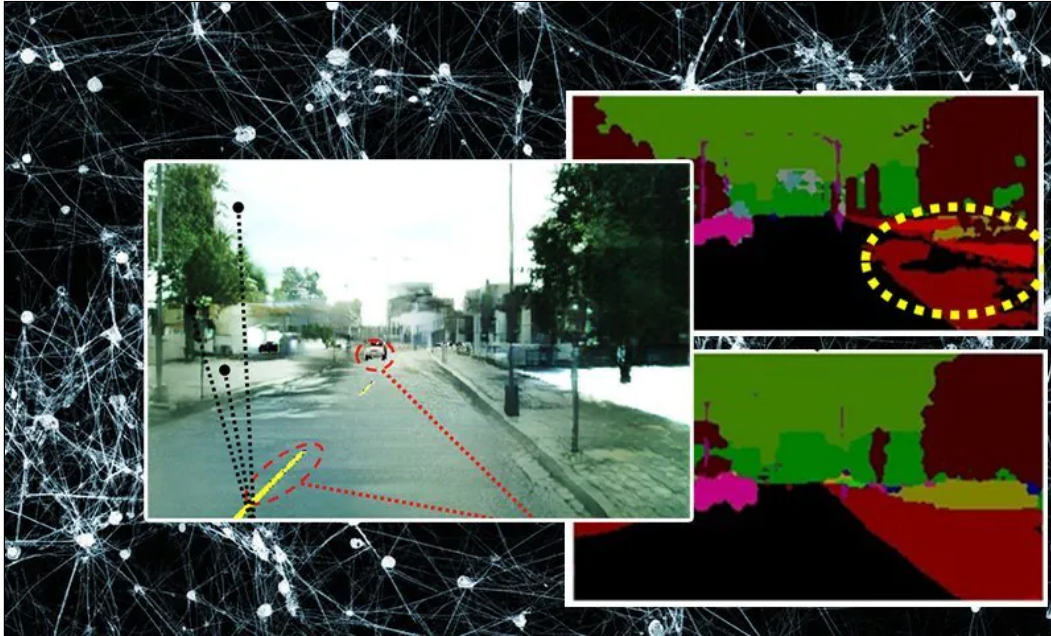


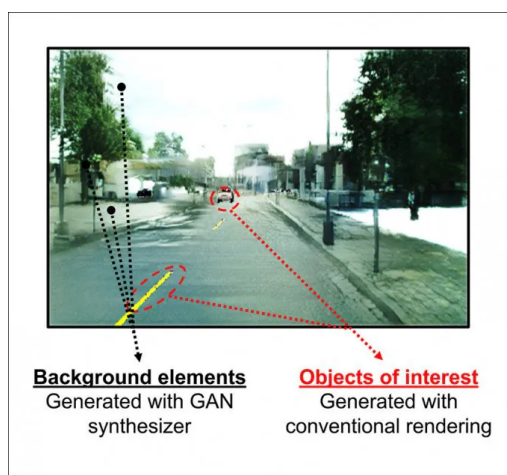
Improving the Photorealism of Driving Simulations with Generative Adversarial Networks

Originally published at <https://www.unite.ai/improving-the-photorealism-of-driving-simulations-with-generative-adversarial-networks/>



A new research initiative between the US and China has proposed the use of Generative Adversarial Networks (GANs) to increase the realism of driving simulators.

In a novel take on the challenge of producing photorealistic POV driving scenarios, the researchers have developed a hybrid method that plays to the strengths of varying approaches, by mixing the more photorealistic output of CycleGAN-based systems with more conventionally-generated elements, which require a greater level of detail and consistency, such as road markings and the actual vehicles observed from the driver's point of view.



Hybrid Generative Neural Graphics (HGNG) offer a new direction for driving simulations that retains the accuracy of 3D models for essential elements (such as road markings and vehicles), while playing to the strengths of GANs in generating interesting and non-repetitive background and ambient detail. [Source](#)

The system, called Hybrid Generative Neural Graphics (HGNG), injects highly-limited output from a conventional, CGI-based driving simulator into a GAN pipeline, where the NVIDIA [SPADE](#) framework takes over the work of environment generation.

The advantage, according to the authors, is that driving environments will become potentially more diverse, creating a more immersive experience. As it stands, even [converting](#) CGI output to photoreal neural rendering output cannot solve the problem of repetition, as the original footage entering the neural pipeline is constrained by the limits of the model environments, and their tendency to repeat textures and meshes.



CLICK TO VIEW ANIMATED GIF

Converted footage from the 2021 [paper](#) 'Enhancing photorealism enhancement', which remain dependent on CGI-rendered footage, including the background and general ambient detail, constraining the variety of environment in the simulated experience. Source: <https://www.youtube.com/watch?v=P1lcaBn3ej0>

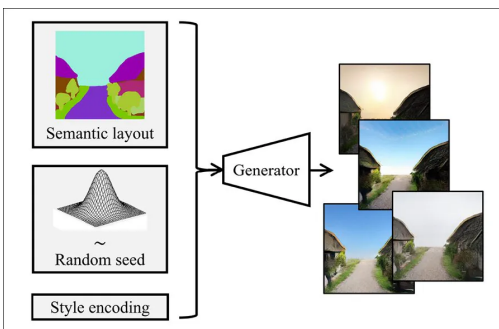
The paper states*:

'The fidelity of a conventional driving simulator depends on the quality of its computer graphics pipeline, which consists of 3D models, textures, and a rendering engine. High-quality 3D models and textures require artisanship, whereas the rendering engine must run complicated physics calculations for the realistic representation of [lighting and shading](#).'

The [new paper](#) is titled *Photorealism in Driving Simulations: Blending Generative Adversarial Image Synthesis with Rendering*, and comes from researchers at the Department of Electrical and Computer Engineering at Ohio State University, and Chongqing Changan Automobile Co Ltd in Chongqing, China.

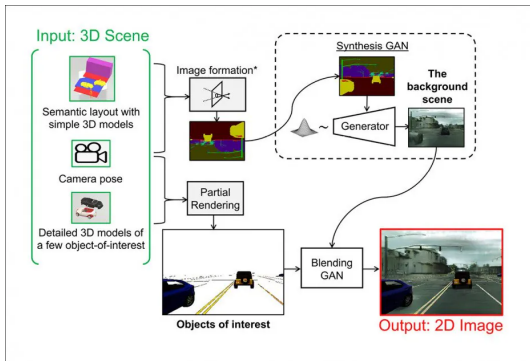
Background Material

HGNG transforms the semantic layout of an input CGI-generated scene by mixing partially rendered foreground material with GAN-generated environments. Though the researchers experimented with various datasets on which to train the models, the most effective proved to be the [KITTI](#) Vision Benchmark Suite, which predominantly features captures of driver-POV material from the German town of Karlsruhe.



HGNG generates a semantic segmentation layout from CGI-rendered output, and then interposes SPADE, with varying style encodings, to create random and diverse photorealistic background imagery, including nearby objects in urban scenes. The new paper states that repetitive patterns, which are common to resource-constrained CGI pipelines, 'break immersion' for human drivers using a simulator, and that the more variegated backgrounds that a GAN can provide can alleviate this problem.

The researchers experimented with both [Conditional GAN](#) (cGAN) and [CycleGAN](#) (CyGAN) as generative networks, finding ultimately that each has strengths and weaknesses: cGAN requires [paired datasets](#), and CyGAN does not. However, CyGAN cannot currently outperform the state-of-the-art in conventional simulators, pending further improvements in [domain adaptation](#) and cycle consistency. Therefore cGAN, with its additional paired data requirements, obtains the best results at the moment.



The conceptual architecture of HGNG.

In the HGNG neural graphics pipeline, 2D representations are formed from CGI-synthesized scenes. The objects that are passed through to the GAN flow from the CGI rendering are limited to 'essential' elements, including road markings and vehicles, which a GAN itself cannot currently render at adequate temporal consistency and integrity for a driving simulator. The cGAN-synthesized image is then blended with the partial physics-based render.

Tests

To test the system, the researchers used SPADE, trained on [Cityscapes](#), to convert the semantic layout of the scene into photorealistic output. The CGI source came from open source driving simulator [CARLA](#), which leverages the Unreal Engine 4 (UE4).



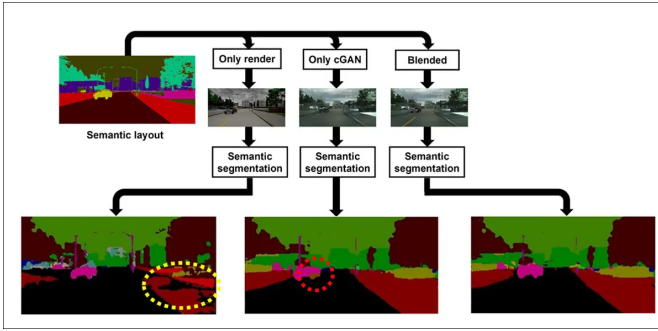
Output from the open source driving simulator CARLA. Source: <https://arxiv.org/pdf/1711.03938.pdf>

The shading and lighting engine of UE4 provided the semantic layout and the partially rendered images, with only vehicles and lane markings output. Blending was achieved with a [GP-GAN](#) instance trained on the [Transient Attributes Database](#), and all experiments runs on a NVIDIA [RTX 2080](#) with 8 GB of GDDR6 VRAM.

SIGGRAPH 2014 - Transient Attributes for High-Level Understanding and Editing of Outdoor Scenes



The researchers tested for *semantic retention* – the ability of the output image to correspond to the initial semantic segmentation mask intended as the template for the scene.



In the test images above, we see that in the 'render only' image (bottom left), the full render does not obtain plausible shadows. The researchers note that here (yellow circle) shadows of trees that fall onto the sidewalk were mistakenly classified by [DeepLabV3](#) (the semantic segmentation framework used for these experiments) as 'road' content.

In the middle column-flow, we see that cGAN-created vehicles do not have enough consistent definition to be usable in a driving simulator (red circle). In the right-most column flow, the blended image conforms to the original semantic definition, while retaining essential CGI-based elements.

To evaluate realism, the researchers used [Frechet Inception Distance](#) (FID) as a performance metric, since it can operate on paired data or unpaired data.

Three datasets were used as ground truth: Cityscapes, KITTI, and [ADE20K](#).

The output images were compared against each other using FID scores, and against the physics-based (i.e., CGI) pipeline, while semantic retention was also evaluated.

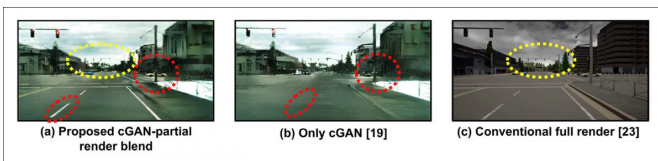
Method	Semantic retention $\uparrow$
Baseline: only render [23]	0.819
only Cycle GAN [17]	0.343
Proposed	
cy-GAN alpha blend	0.362
cy-GAN pyramid blend	0.353
cy-GAN GAN blend	0.318
cGAN alpha blend	0.879
cGAN pyramid blend	0.868
cGAN GAN blend	0.846

In the results above, which relate to semantic retention, higher scores are better, with the CGAN pyramid-based approach (one of several pipelines tested by the researchers) scoring highest.

Method	FID $\downarrow$		
	Cityscapes[45]	KITTI[50]	ADE20K[51]
Only render [23]	231.768	285.222	361.496
Proposed			
Cy-R alpha blend	175.832	220.223	272.069
Cy-R pyramid blend	196.911	228.277	279.704
Cy-R GAN blend	194.191	234.087	266.615
c-R alpha blend	188.809	220.161	272.877
c-R pyramid blend	202.120	214.488	265.603
c-R GAN blend	194.898	217.663	260.404

The results pictured directly above pertain to FID scores, with HGNG scoring highest through use of the KITTI dataset.

The 'Only render' method (denoted as [23]) pertains to the output from CARLA, a CGI flow which is not expected to be photorealistic.



Qualitative results on the conventional rendering engine ('c' in image directly above) exhibit unrealistic distant background information, such as trees and vegetation, while requiring detailed models and just-in-time mesh loading, as well as other processor-intensive procedures. In the middle (b), we see that cGAN fails to obtain adequate definition for the essential elements, cars and road markings. In the proposed blended output (a), vehicle and road definition is good, whilst the ambient environment is diverse and photorealistic.

The paper concludes by suggesting that the temporal consistency of the GAN-generated section of the rendering pipeline could be increased through the use of larger urban datasets, and that future work in this direction could offer a real alternative to costly neural transformations of CGI-based streams, while providing greater realism and diversity.

** My conversion of the authors' inline citations to hyperlinks.*

First published 23rd July 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More