35th European Modeling & Simulation Symposium
20th International Multidisciplinary Modeling & Simulation Multiconference

2724-0029 © 2023 The Authors.
doi: 10.46354/i3m.2023.emss.038

The Synergistic Relationship between Artificial Intelligence and Simulation for Real-World Applications

Jorge Luis Jiménez Aparicio[1,*] and Jürgen Roßmann[1]

[1]Institute for Man-Machine-Interaction, RWTH Aachen University, Ahornstrasse 55, Aachen 52074, Germany

[*]Jimenez@mmi.rwth-aachen.de

Abstract

Artificial Intelligence (AI) has gained importance in several fields, but challenges still remain when developing effective AI systems. Generating rich datasets and testing AI's performance in special conditions are primary obstacles in AI development. Simulation offers a possible solution by generating synthetic data for training and flexible testing environments. However, the generation of highly detailed simulation implies a lot of effort and sometimes cannot even be integrated into AI development due to lack in performance and quality. AI could help close these gaps in simulation. We propose a framework with which the mutually beneficial relationship between simulation and AI can be studied. It focuses on how AI can help improve simulation that in turn will be used for the benefit of AI deployed in real applications, which also serves as feedback to improve the former. In this survey-like contribution, we also highlight examples of current AI developments that can be useful to enhance some key factors of realistic simulation and analyze how simulation can help with training and testing of AI for real applications. Finally, we show our preliminary results on human vs. AI perception in detection networks and on video to video synthesis, to further validate the necessity of research on the AI's influence on simulation.

Keywords: Simulation; Artificial Intelligence; sensor simulation

1. Introduction

Different fields of application nowadays demand the use of Artificial Intelligence (AI) to tackle problems that are too hard to code manually by engineers and developers. Although AI seems like a compelling solution to this kind of problems, it is usually not easy to develop functioning and reliable AI in a cost-effective way. Most AI developing technologies face the same problem: scarce availability of varied, meaningful training data. For instance, in the automotive sector only a few companies (e.g. Tesla) with fleets of semi-autonomous cars that continuously collect data to feed their networks, have access to a big and widely varied dataset to achieve acceptable AI behavior. But even with such data acquisition methods, they miss crucial information from special cases that fully autonomous AI needs to handle. Moreover, testing of the AI performance becomes a problem due to the high costs in fields of application like space missions. Virtual Testbeds (VTBs) Rossmann et al. (2016), thorough 3D simulation environments that portray all important aspects of the applications, provide a possibly effective solution to these problems. They offer a cost-effective, fast, safe and very flexible alternative to generate virtual environments Jochmann et al., which could be used for AI testing or to create datasets for AI training. However, this often requires highly detailed and realistic scenarios that need a lot of human labor, expert knowledge and computational effort.

AI cloud also be use to help overcome limitations of virtual worlds, and close the remaining gap between simulation and reality. Following the Pareto principle, while the first steps in simulation yielded major improvements with little effort, the last necessary improvements require the most effort. Several current AI developments can be implemented in simulated environments to increase their

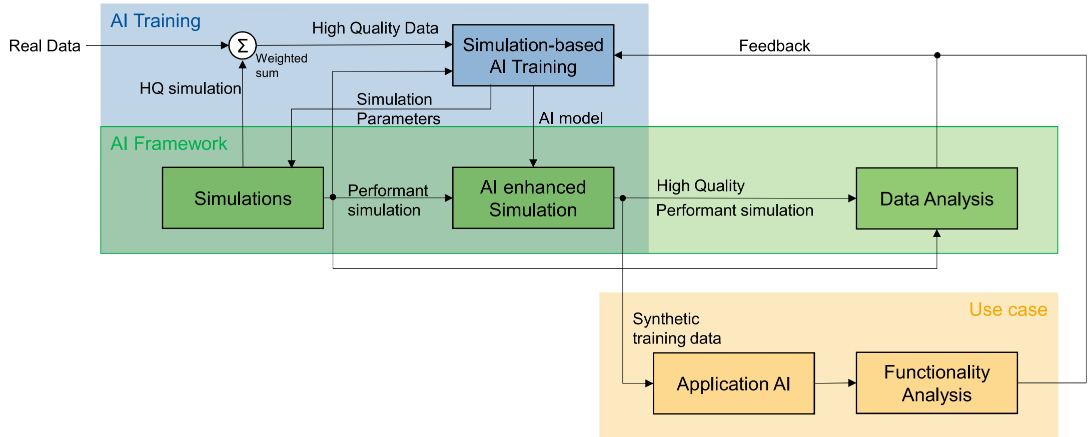

Figure 1. Diagram of the framework proposed to study the mutually beneficial relationship between Simulation and Artificial Intelligence.

performance and realness, which in turn leads to better synthetic data for AI training and testing.

We want to investigate if AI can be used to enhance simulation and ultimately help produce better AI for real applications ("real-AI"). The purpose of this "simulation-AI" is to either improve upon current working classical methods of simulation or benefit from them, rather than replace them. It is important to keep in mind that the enhancements in simulation must be beneficial in the eyes of the real-AI, i.e. either the performance or the ease of use of the real-AI are increased due to the improvements in the virtual worlds. Our motivation is then to scrutinize the benefits and limitations of using AI to improve simulation and vice versa.

For this purpose, we propose a framework to study the potentially mutually beneficial relationship between simulation and AI (Figure 1). This includes the use of synthetic data to train and test real-AI by easily configuring and validating virtual scenarios according to specific use cases, focusing on the real-AI to be used. In the following sections, we give an insight to the proposed framework and highlight related work relevant to each part of it, we show preliminary results on human vs. AI perception and video-to-video translation, and we provide our conclusions and the focus of our future work.

1.1. The relationship between simulation and artificial intelligence (concept)

Virtual Testbeds are able to produce high quality data with good performance. Quality results often require a great human effort to obtain, while good performance demands expensive computations which leads to the need of expensive hardware (e.g. GPUs with high parallelism). AI can help overcome this drawbacks for the VTBs to be effectively used in real-AI development. Our proposed framework allow us to investigate the use of AI to close the remaining gap to realism of the VTBs, while focusing on use case applications from the perspective of the real-AI to be used.

The diagram of our proposed framework to study the relationship between simulation and AI is seen in Figure 1. It is centered on AI enhanced simulation, for which the training of AI models is needed. As with most AI, training data plays a major role in the success of the algorithm; in this case, high quality data (e.g. high resolution or ray tracing images) is used to train such models. The proportion of synthetic and real training data will depend on the goal of the simulation-AI, which can be the improvement of simulation performance, quality or environment realness. The AI enhanced simulation needs to be perceived as performant and of high quality in the real-AI's perspective. With this, not only can we use the synthetic data to train AI that will work in real life, but we can also use the enhanced virtual worlds to test this real-AI. Furthermore, both the analysis of the obtained simulated data and the performance of real-AI trained with such data, serve as feedback for further iterations of the simulation-AI.

All in all, this concept can be viewed as having two main paths of thought which, with the help of the related work to this topic, will be discussed further in the following sections. These paths of thought and the goals of our research are:

- The benefits that AI brings to simulation in order to close the gap between real and simulated systems.
- The benefits that simulated environments can bring to AI applications, specially for training and testing.

2. State of the art: Improving simulation with AI

2.1. Simulation

Simulation environments are constantly improving and are showing very impressive graphical fidelity, especially in the video game industry. Leaders in this sector are Unreal Engine 5 (UE5) and Unity. These engines are sometimes used for research and testing in sectors like the automotive one. A recently released simulation environment platform called Omniverse NVIDIA focuses on automotive and industry sectors. They state that their platform can be used to generate synthetic data for training AI perception networks. This platform is the closest to the intended capabilities of VTBs. However, their environments are not boosted by AI and they are assuming that their synthetic data works in real applications without any feedback of the AI performance to the 3D world generation tools. Both of which we intend to investigate and analyze in this and future publications.

In the following, we go through the different ways in which AI can help improve simulation by highlighting some state of the art developments that could prove useful.

2.2. Performance

Performance of the simulation can be increased by leaving taxing computations to the AI. Some examples follow:

2.2.1. *Quality*

We refer to quality as how the simulation is perceived by the human eye, and how it compares to real life, including real sensors (e.g. cameras, LiDAR).

The use of image-to-image translators from both supervised (Wang et al. (2018b), Park et al. (2019)) and unsupervised (Liu et al. (2017), Huang et al. (2018)) methods produce a high quality frame from a low quality or even semantic frame. This could be used at every step of the simulation to generate a higher quality output at the lower computational cost of simulating semantic frames. However, this rises temporal consistency problems with the material properties (e.g. color).

Video-to-video translators based on Generative Adversarial Networks (GANs) Wang et al. (2018a) aim to solve consistency problems and achieve a high quality, short-term temporally coherent output. World consistent video-to-video translation Mallya et al. (2020) builds on top of this. Using a guiding image, it achieves a long-term temporal consistency over the entire 3D world. However, this translator requires inputs of the same quality (e.g. real images) as the desired output to achieve temporal consistency.

As this defeats the purpose of using synthetic data, alternative ways to calculate the parameters for temporal consistency are needed. Thankfully, simulated environments can provide additional data of the world (so called Ground Truth data) which, together with modifications of the translator's structure, should be sufficient to calculate said parameters and achieve comparable results. This means that this AI framework could be used to transform semantic data of the simulation into photorealistic or ray tracing looking ones. Figure 2 shows our objective: research the use of AI to turn a semantic input into a ray tracing output during simulation to reduce the computational effort, thus reducing the hardware requirements.

2.2.2. *Resolution*

Deep Learning Super Sampling (DLSS) from NVIDIA Burnes (2020) uses a Convolutional Autoencoder to produce high resolution images from lower resolution images and their motion vectors. This produces comparable results to natively generated super resolution images and a substantial increase in performance. FidelityFX AMD from AMD and DirectML Microsoft from Microsoft are alternatives to DLSS.

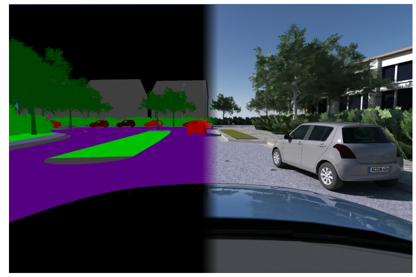

Figure 2. Example of one of our goals in performance improvement. From a semantic input the AI will be able to generate a ray tracing output. This will be integrated to the simulation pipeline to reduce computational effort. Virtual scenario and sensor data generated using VEROSIM VEROSIM and the framework presented in Thieling and Rosmann.

2.2.3. *Denoising*

The AI-accelerated denoiser from NVIDIA, introduced with OptiX 5.0 Chaitanya et al. is a post-processing feature based on Autoencoders. This AI can "dramatically reduce the time to render a highly fidelity visually noiseless image", which could lead to performance improvements. Similarly, the denoiser in Mildenhall et al. (2021) is capable of taking extremely noisy raw images captured in dark environments and generate a synthetic high dynamic range view.

2.2.4. *Physics and Dynamics of the system*

Computation of the physics and dynamics of a system can be very costly. Some AI developments in this area have proven that deep networks are capable of boosting performance for simulation problems with a known solution. In Holden et al., they use a feed-forward neural network to simulate subspace physics that interacts with external objects and generates close to ground truth results significantly faster than classic methods. Moreover, in Sanchez-Gonzalez et al. they use Graph Networks (GNs) together with an encoder and a decoder to describe complex physical systems composed of particles, showing that they can accurately simulate thousands of particles to solve complex physics problems in simulation.

2.3. Realism

Implementing models of the simulation agents (e.g. sensors) that are not ideal but rather describe their behavior as close to reality as possible with AI help, is the goal of this subsection.

Figure 3. Example of the development in Thieling et al. (2019) applied to our simulation pipeline. Color transfusion from vegetation is present in some objects due to the distribution of trees and houses in the limited real dataset used in that research. Virtual scenario and sensor data generated using VEROSIM VEROSIM and the framework presented in Thieling and Rosmann.

2.3.1. *Sensor simulation results*

For simulated camera sensors, the goal is to achieve photorealistic images that integrate error models and noise typical of real sensors. The work in Thieling et al. (2019) is a first approach to use unsupervised GANs based image-to-image translation to achieve photorealistic outputs out of rasterized simulation. It takes advantage of *transfer learning*, an approach in Machine Learning where one uses a previously trained network as the starting point of training; as the network already knows abstract concepts of the initial domain, that are also found in the desired domain, the amount of new data required to train is drastically reduced. Figure 3 shows an example of this image refiner in our simulation pipeline, it shows a more realistic image, although some artifacts and color transfusion from the vegetation can be seen.

Furthermore, the research in Richter et al. (2021) takes rendered images from the video game GTA V and produces enhanced images that look realistic. They use G-buffer encoders to get typical rendering information of the input and pass them to their image enhancement network together with the input to produce the enhanced output. This research shows very promising results for using AI to simulate the behavior of real camera sensors.

Another important type of sensor is the LiDAR. The research in Elmadawi et al. (2019) presents a structure using a fully convolutional Deep Neural Network (DNN) together with Polar Grid maps and a Histogram classifier to simulate the realistic behavior of a LiDAR sensor with promising results.

2.3.2. *Modeling*

When modeling objects, it is often necessary to simplify the model so that the engine is able to simulate the scene at a reasonable computational cost. Using AI to learn models that have the necessary complexity to fully imitate the real physics and dynamics of the system is a step further in improving realism. This way, better digital twins of the objects and the 3D world can be achieved.

2.3.3. *Textures*

Textures and materials are important for the realness of the scene, as they contain information on how light particles should behave when in contact with objects. The work in Ramesh et al., although primarily a text-to-image translator with focus in art, has been shown to be effectively used in the creation of realistic looking materials in modeling software like Blender Brown. Ultimately, the simulation-AI can help set up simulation parameters to achieve more realistic materials.

2.4. Data Analysis

We need a way to objectively measure if our enhancements are improving realism. For instance, a super resolution high quality simulation may look realistic for some but not for others. In this part of the framework, we try to tackle that subjectiveness by proposing the use of some metrics that may give more insight in how realistic a simulation is.

Non reference image quality metrics like CEIQ Yan et al. for contrast, BRISQUE Mittal et al. for space quality or NoR-VDPNet Banterle et al. for visibility and luminance (or their combination); could offer a first approach into measuring realness of synthetic data. Ultimately, virtual worlds need to work with real-AI, so it is important that in this AI's perspective, the simulation is realistic. Therefore, valid AI performance metrics have to be compared with the results of the realism metrics.

3. State of the art: How simulation can help AI

Two major steps in real-AI development can be improved with the help of simulation: training and testing. In the following we take a closer look at them and at how functionality analysis can help simulation-AI.

3.1. Training

One of the problems of training in machine learning is the scarce availability of meaningful, varied and labeled datasets. Even when there is available data, e.g. the Cityscapes dataset Cordts et al. (2016), samples of extreme conditions are missing to achieve a reliable AI. These are gaps that can be filled with the help of simulation. Virtual worlds allow the automatic generation of bigger, varied, labeled datasets, that are produced at a faster rate and lower cost than real datasets. Furthermore, these worlds can be used to generate indispensable data of special situations that are hard, if not impossible, to recreate in real life, like: abnormal, dangerous or extreme situations.

Agriculture is a sector where training data is not easily available. The German project "Feldschwarm" Feldschwarm is developing a fleet of smaller tractors that work in a swarm-like interactive manner, where the surroundings recognition system is based on a real time object detector. They have done research on using synthetic data to train AI for this purpose Jiménez Aparicio et al. (2020). They tried using only synthetic data, a combination with real data and *transfer learning*, with the later showing the best results. They state that using simulated sensor data for AI training is viable but the simulation technology is not good enough yet to have the performance of comparable real datasets, showing that improving simulation can be beneficial to real-AI.

All in all, the better approach is to use synthetic data to complement already existing real data, when available, to help improve the performance of real-AI.

3.2. Testing

Testing an AI application to be sure it is reliable and safe for usage in real life can take a lot of time. In the automotive sector, testing an AI's reliability for its safe deployment could take up to hundreds of years of driving time Kalra and Paddock. Furthermore, some cases are too dangerous or expensive to repeatedly test in real life. Simulated worlds not only offer more flexibility in test environments but also allow testings to be done in parallel (e.g. by running different tests in several computers at the same time), significantly reducing the time needed and risks taken. This also hints at the benefits of increasing simulation performance for real-AI.

Although the bulk of the tests and the special cases can be handled in virtual worlds, the better approach is to complement real life tests, as they are unavoidable.

3.3. Functionality Analysis

Real-AI's performance is not only important for its given task but also for simulation-AI. As real-AI needs to be successfully trained and tested with synthetic data, its performance can serve as feedback for the simulation-AI, to determine if the improvements on simulation are relevant or not. Naturally, different real-AI applications use different metrics to measure performance, e.g. in object detection and classification the commonly used metric is the mean Average Precision (mAP) Everingham et al. (2010). These metrics need be cross checked against the proposed metrics in section 2.4, to corroborate if the simulated data is also more realistic to the real-AI. In this part of the work it is also important to test how vulnerable is such an AI to being fooled by tampered images.

4. Preliminary Results and Discussion

To corroborate the importance of the real-AI's view in our proposed framework, our first approach was to evaluate the difference in a human's and an AI detector's perception of an image. We started by testing two different resolutions for the same virtual scenario (Figure 7). If you take a look at Figure 4, both images appear to be the same. However, when calculating the pixel by pixel difference, we can see that there is indeed a slight difference between them, this is portrayed in Figure 5.

But is this difference relevant for the AI? To test this, we implemented one of the state of the art real time capable AI detectors, i.e. You Only Look Once (YOLOv3) Redmon and Farhadi (2018). We used the standard weights published by the author and run the inference on the two resolutions. As our focus was not to measure the performance of the AI, but to test if there is any difference from its perspective, we did not base the detection on the ground truth of the scene but on each other. I.e. we took one of the resolutions (2560x1440) as our ground truth for the detection. Additionally, YOLOv3 takes 416x416 images as input, this means that both datasets are scaled down before running inference. If the initial resolution had no impact in the detection, the Average Precision of each category, as well as the mAP, would be 100%. Figure 6 shows the AP for each detected category and states a mAP of 71.55% for this test. With this, it can be concluded that the AI behaves differently for different initial resolutions and that the two datasets are in fact different for the AI despite being practically the same for the human eye. This also demonstrates the need for appropriate metrics to measure these images, as the human eye is incapable of perceiving differences that the AI would.

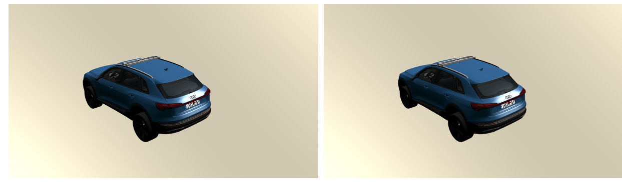

Figure 4. Two images of the same asset with different original resolution. One image has an original resolution of 2560x1440 (left) and the other of 1440x720 (right).

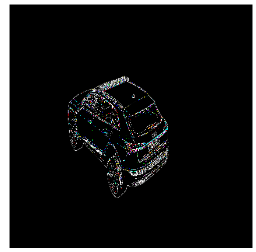

Figure 5. Pixel by pixel difference between images in Figure 4 after they have been passed through YOLO's pre processing, where they get rescaled to 416x416.

Moreover, Figure 7 shows two example frames, where cars are not being detected in the lower resolution, despite having a detection in the higher resolution. These are examples of why the mAP ended up not being 100%.

We tested the translator "vid2vid" Wang et al. (2018a), mentioned in section 2.2, with the semantic images of the virtual version of the KITTI dataset (VKITTI2) Cabon et al. (2020); Gaidon et al. (2016) and run the results through YOLO. Figure 8 shows the Precision vs Recall curve with the average precision (AP) of 20.02% for the car class, together with two examples were both the results of the translation and the inference are seen. The AP shows that although the detections look relatively promising, the performance is rather poor with respect to the ground truth

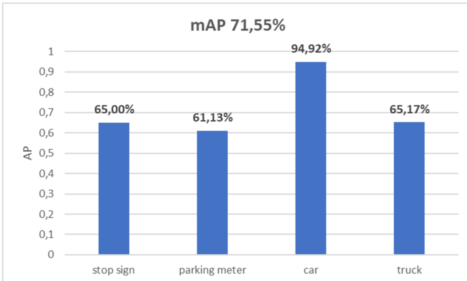

Figure 6. Average Precision (AP) for four detected categories, and the calculated mAP for the dataset. Here the high resolution dataset is taken as the ground truth to test if there is any difference in the detections of both datasets. This does not measure the performance of YOLOv3.

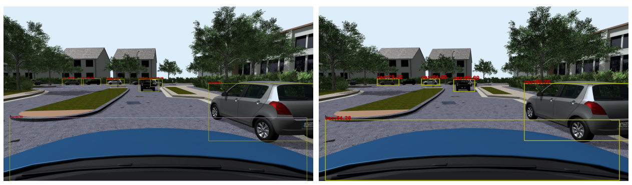

Figure 7. Two examples of the inference run with YOLOv3 Redmon and Farhadi (2018), both show instances where the cars are being miss detected in the lower resolution dataset. Images on the left have the original resolution of 2560x1440 and images on the right of 1440x720. Virtual scenario and sensor data generated using VEROSIM VEROSIM and the framework presented in Thieling and Rosmann.

of the simulation (i.e. perfect pixel-resolution annotation). This implies that further changes in the implementation are needed in order for this AI to improve simulation in the perspective of real-AI. Furthermore, Both images have many artifacts and blends of objects but follow semantic input information relatively well. However, some of the blends of cars are being detected with a very high confidence. This happens because the AI fills that specific area with features and shapes common to cars and the detector gets "tricked" into thinking that it should be classified as a car, which hints to the need of testing real-AI's robustness against tampered data.

Summarizing, our first approach to study the integration of the simulation-AI developments in simulated worlds, shows that even seemingly equal datasets have a different performance on real-AI. Metrics to calculate the realness and, in turn, the performance of synthetic datasets in the real-AI's perspective are necessary before any improvements in simulation for real-AI training and testing can be accurately measured. Furthermore, although the translation from simulated semantic to photorealistic videos shows promising results, it still lacks temporal consistency and realness to be effectively used in real-AI training and testing.

5. Conclusion

We proposed a framework to study the mutual benefits of the integration between AI and simulation, as well as its limitations. The framework shows a cyclical relationship between both, were AI is used to improve simulation, which in turn would improve real-AI, and give feedback for further enhancements of the simulation-AI. We also highlighted examples of the areas of simulation that could be improved with recent AI developments. Furthermore, we analyzed how the use of simulated environments could positively impact the training and testing of real-AI.

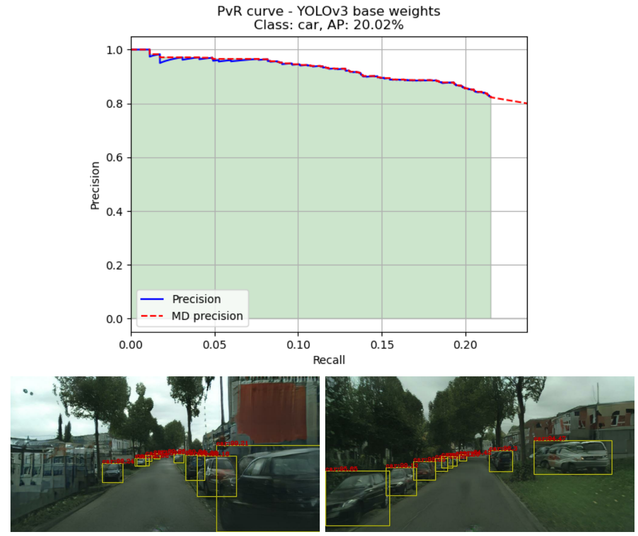

Figure 8. (top) Precision vs Recall curve for car detection from YOLOv3 Redmon and Farhadi (2018) on the vid2vid Wang et al. (2018a) output using semantic images from VKITTI2 Cabon et al. (2020); Gaidon et al. (2016), giving an average precision of 20%. (bottom) Two examples of the results.

Our preliminary results show that perception AI can distinguish datasets that the human eye could not (e.g. different original resolution rescaled to the same one). This gives two conclusions: first, appropriate metrics to measure the realness of synthetic data is needed, human perception is too subjective and the real-AI's perspective is more important as it will be developed with said data. Second, resolution plays a role in perception networks and its up-scaling by AI would impact the real-AI's performance. Furthermore, video to video translation showed promising results in the conservation of semantic information but lacks temporal consistency; world consistency is left for future research. It also hinted at the importance of training real-AI for robustness against falsified data.

The research in simulation and AI that can be done with our proposed framework will help determine in which cases AI is useful as an improvement or replacement of classical methods, and when it can be used to generate parameters of the simulation, all from the perspective of real-AI. Our future work will be based on the developments of simulation- and real-AI, with a special focus on ease of use enhancement, and integration and validation of them.

6. Funding

This work is part of the project "ViTOS-3", supported by the German Aerospace Center (DLR) with funds of the

German Federal Ministry of Economics and Technology (BMWi), support code 50 RA 2120.

References

AMD. AMD Radeon™ Software | FidelityFX | AMD.

Banterle, F., Artusi, A., Moreo, A., and Carrara, F. Norvdpnet: A no-reference high dynamic range quality metric trained on hdr-vdp 2. In *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE.

Brown, S. Dall-e2 to substance tutorial by stan brown.

Burnes, A. (2020). NVIDIA DLSS 2.0: A Big Leap In AI Rendering.

Cabon, Y., Murray, N., and Humenberger, M. (2020). Virtual kitti 2.

Chaitanya, C. R. A., Kaplanyan, A. S., Schied, C., Salvi, M., Lefohn, A., Nowrouzezahrai, D., and Aila, T. Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Elmadawi, K., Abdelrazek, M., Elsobky, M., Eraqi, H. M., and Zahran, M. (2019). End-to-end sensor modeling for lidar point cloud. IEEE.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*.

Feldschwarm. The Feldschwarm project website.

Gaidon, A., Wang, Q., Cabon, Y., and Vig, E. (2016). Virtual worlds as proxy for multi-object tracking analysis. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4340–4349.

Holden, D., Duong, B. C., Datta, S., and Nowrouzezahrai, D. Subspace neural physics. In *Proceedings of the 18th annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. ACM.

Huang, X., Liu, M.-Y., Belongie, S., and Kautz, J. (2018). Multimodal unsupervised image-to-image translation. In *ECCV*.

Jiménez Aparicio, J. L., Thieling, J., Roßmann, J., Robert, M., and Bosdorf, R. (2020). Evaluation of virtual methods for training neural networks in agricultural applications. In *2020 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor)*.

Jochmann, G., Thieling, J., Dahmen, U., and Rossmann, J. Virtual testbeds for the development and validation of sensor-based systems using digital twins. In *2022 International Symposium on Electrical, Electronics and Information Engineering (ISEEIE)*. IEEE.

Kalra, N. and Paddock, S. M. *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* RAND Corporation.

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Mallya, A., Wang, T.-C., Sapra, K., and Liu, M.-Y. (2020). World-consistent video-to-video synthesis. In *European Conference on Computer Vision (ECCV)*.

Microsoft. Direct Machine Learning (DirectML).

Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P. P., and Barron, J. T. (2021). NeRF in the dark: High dynamic range view synthesis from noisy raw images. *arXiv*.

Mittal, A., Moorthy, A. K., and Bovik, A. C. No-reference image quality assessment in the spatial domain. 21(12):4695–4708.

NVIDIA. Omniverse Platform for 3D Design Collaboration and Simulation | NVIDIA.

Park, T., Liu, M.-Y., Wang, T.-C., and Zhu, J.-Y. (2019). Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. Hierarchical text-conditional image generation with clip latents.

Redmon, J. and Farhadi, A. (2018). YOLOv3: An Incremental Improvement.

Richter, S. R., AlHaija, H. A., and Koltun, V. (2021). Enhancing photorealism enhancement. *arXiv:2105.04619*.

Rossmann, J., Schluse, M., Rast, M., and Atorf, L. (2016). E-Robotics Combining Electronic Media and Simulation Technology to Develop (Not Only) Robotics Applications. In *E-Systems for the 21st Century*. Apple Academic Press.

Sanchez-Gonzalez, A., Godwin, J., Pfaff, T., Ying, R., Leskovec, J., and Battaglia, P. W. Learning to simulate complex physics with graph networks.

Thieling, J., Elspas, P., and Roßmann, J. (2019). Neural networks for end-to-end refinement of simulated sensor data for automotive applications. In *2019 IEEE International Systems Conference (SysCon)*.

Thieling, J. and Rosmann, J. Scalable sensor models and simulation methods for seamless transitions within system development: From first digital prototype to final real system. 15(3):3273–3282.

VEROSIM. 3D-Simulationssoftware.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Liu, G., Tao, A., Kautz, J., and Catanzaro, B. (2018a). Video-to-video synthesis. In *Conference on Neural Information Processing Systems (NeurIPS)*.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., and Catanzaro, B. (2018b). High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Yan, J., Li, J., and Fu, X. No-reference quality assessment of contrast-distorted images using contrast enhancement.