Applications and Limitations of Machine Learning in Computer Graphics

Chengyu Fu

Department of Computer Science, The University of Sheffield, Sheffield, United Kingdom
cfu10@sheffield.ac.uk

Abstract. Just as in many other fields, the application of machine learning methods has become a prominent topic in the realm of computer graphics. This paper reviews and summarizes some ongoing applications of machine learning in both the academic and industrial aspects of computer graphics, along with certain challenges and limitations being faced. The applications encompass the three primary facets of computer graphics: rendering, modelling, and animation. Machine learning's roles in areas such as anti-aliasing, ambient occlusion, model generation, and motion capture techniques are discussed and also compared with the traditional methods. On the one hand, the advantages of machine learning are shown by the applications part. On the other hand, limitations are also introduced. The limitations section primarily highlights the need for labelled data and associated cost concerns, as well as the inflexibility and lack of creativity inherent to machine learning. In the end, the article provides a summary of the entire content and offers a glimpse into future trends.

Keywords: Computer graphics, Machine learning, Rendering, Modelling, Animation

1 Introduction

Computer graphics is a subject in Computer Science that studies the generation of images and art using computers. It was first coined by computer graphics researchers Verne Hudson and William Fetter of Boeing in 1960. In general, computer graphics could be either 2D or 3D, while nowadays 3D computer graphics is more common. Computer graphics is an essential technology in many fields especially in video games, films, virtual reality (VR), augmented reality (AR) and digital art, and mainly consists of three parts which are modelling, rendering and animation.

Machine learning (ML) is one of the most important parts of the study of artificial intelligence (AI) and it was coined by Artur Samuel in 1959 [1, 2]. The study of machine learning is an interdisciplinary subject that involves probability theory, statistics, algorithm complexity etc. Instead of using a human-developed algorithm to tell the machines what to do explicitly to solve a problem [3], machine learning methods aim to help the machines discover and improve their 'own algorithms' [4]. With the rapid development of artificial intelligence, machine learning has already been widely used in almost every aspect.

© The Author(s) 2024
B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_55

It is not surprising that in the last few years machine learning has also been merged with computer graphics. Companies like Nvidia and AMD have already developed several computer graphics techniques that use machine learning and have shown magnificent results with both improvements in performance and quality, which prove the use and the meaning of machine learning in computer graphics without doubts. This would encourage more and more people to explore the chemical reactions between these two charming subjects in the future.

Currently, the research on the applications and limitations of machine learning in computer graphics is still not comprehensive and needs to be further discussed. This paper's main content will be laid out by summarizing and pointing out some symbolic examples for both applications and limitations, aiming to give a clearer view and examination of the present situation of this topic. The application part will include all three core parts of computer graphics – modelling, rendering and animation. On the other hand, the limitations will mainly be introduced from two angles.

2 Applications

The fusion of machine learning and computer graphics has significantly impacted various domains, encompassing rendering, modelling, and animation. This section delineates the machine learning applications within anti-aliasing, ambient occlusion, 3D model generation, and motion capture.

2.1 Anti-Aliasing

The traditional offline anti-aliasing methods, such as supersampling, require a high level of computing power that could result in limited performance in interactive applications [5]. The Temporal Anti-Aliasing (TAA) method is widely used in real-time renderers and game engines. It shows an impressive performance in efficiency as it serves for real-time rendering, however, as a blur filter is applied during the process, some fine details would be removed and causing the final image to seem hazy and leading to a relatively low quality [6].

Deep learning has been applied to try to solve these drawbacks. By training a convolutional neural network, the network could generate the sequence of samples with about 16 times the number of samples rendered in the original sequence [5]. The convolutional neural network could share the computing stress in supersampling, to generate images with higher quality in real-time rendering. In particular, the Deep learning anti-aliasing (DLAA), which is created by Nvidia is a great example of the application of machine learning in the anti-aliasing aspect. Compared to TAA, although both are spatial anti-aliasing methods which means they require the data of past frames and both fit the real-time rendering, DLAA represents a higher level of ability in resolving details and so the final image quality is also much better than TAA [7, 8].

Like DLAA but may be more famous, Deep Learning Super Sampling (DLSS) is one of the most popular anti-aliasing techniques in the game industry. It is also created by Nvidia and the first version DLSS 1.0 was released in February 2019. DLSS and

DLAA share the same machine learning-aided anti-aliasing method while DLSS consists of an upscaling function that could improve the performance, DLAA doesn't have any upscaling or downscaling functionality but just focus on the visual quality produced by anti-aliasing.

2.2 Ambient Occlusion

In the 3D Computer Graphics aspect, ambient occlusion is referred to as a technique that calculates the extent of how each point in a scene is exposed to ambient lighting. It plays an essential role in real-time rendering as it makes the image more realistic. Screen Space Ambient Occlusion (SSAO) is one of the most popular methods for calculating occlusions.

Inevitably, in a highly sophisticated scene, the ambient occlusion often makes mistakes and results in ignoring some details [9]. As a result, in the past several years, much research on using machine learning to help improve ambient occlusion algorithms has been raised.

Many methods in the machine learning aspect have been tried to be applied to the ambient occlusion algorithm. Neural Network Ambient Occlusion (NNAO) is one example. By using a neural network to learn the SSAO algorithm, it could minimise the errors that a normal SSAO algorithm would produce, with the use of several cost functions [10]. The researchers compared the NNAO with other ambient occlusion methods that are SSAO, SSAO+ and Horizon Based Ambient Occlusion (HBAO) and the result showed that NNAO produced a much lower level of error numbers and a generally good runtime performance with a slightly lower performance in low sample count scenario but better performance in high sample count scenario [10]. Another application in research is AOGAN, which is an SSAO algorithm uses the Generative Adversarial Network (GAN), the result also presented an improvement in quality with no significant increase in cost for running time to achieve real-time performance [9]. Nonetheless, this study faces a common constraint seen in all screen-space solutions, whereby the G-buffers lack complete scene information, potentially leading to artefacts at screen boundaries or in culled scenes [9]. In general, both NNAO and AOGAN showed a great performance in improving the accuracy and runtime.

2.3 3D Model Generation

The traditional way of 3D modelling manually could be time-consuming, and impossible in a large-scale scenario that involves too many objects. For the latter, the Perlin Noise function developed by Ken Perlin became one of the most important functions used in procedural generation. The procedural generation is commonly used for creating enormous scenarios such as a city, terrain and even a whole planet that require randomness in appearance and contain massive objects. Recently, machine learning has also been combined with it. After capturing the images of landscapes by using unmanned aerial vehicles (UAVs) or satellites, the GAN model could be applied to learn a large amount of real-world terrain and environment information [11]. This could help the final model contain the similarity of some real-world scenarios to make the environment created more realistic.

Except for procedural generation, traditional 3D modelling is also facing a revolution. In the computer graphics industry, the modelling process usually requires experienced staff and highly detailed design graphs. Especially when the high-precision model is required. In the past few years, the machine learning method has made the modelling much simpler - by only using a single 2D image, a 3D model could be generated. Furthermore, some text-based model generation applications are also invented, high-quality models can be created by typing in sentences that include the key features that the user desired, one example is the Imagine 3D produced by Luma AI. The invention of text-based model generation means the users not only have no need to know the modelling knowledge but also do not have to prepare a 2D image in advance. By only using a keyboard to type in what they have in mind, the model with precise requirements can be generated. Some screenshots of models generated on the Imagine 3D website are represented as examples. (Fig.1 and Fig.2) Compared to the traditional way of modelling, these methods could be easier and more friendly for users without modelling knowledge and useful for personal or small projects. However, for large projects, the precision of these models may not fit complicated requirements.

Fig. 1. Model generated by the text "a single colorful expensive koi fish", on Luma AI, Imagine 3D [12]

Fig. 2. Model generated by the text "a huge Egyptian statue with a lion head", on Luma AI, Imagine 3D [13]

2.4 Motion Capture

Motion Capture is a process of tracking and recording the movement of people or objects. It is a technique that has been widely used in the movie, game, and other multimedia industries, even in the military, sports, and medical aspects. Compared to traditional means of creating animations for models, motion capture could provide a real-time sequence of movements with low latency. This can reduce the costs of keyframebased animation in entertainment applications [14]. Motion Capture is an important part of the animation part of computer graphics and a cross-field of computer vision and computer graphics. Machine learning methods are commonly used for pose estimation, as an important part of motion capture and computer vision. Motion capture requires support from expensive motion capture devices, machine learning could make the motion capture markerless.

Except for lowering the cost, machine learning could also help the motion capture to be more accurate and efficient. Through learning the dataset of existing motion capture data, machine learning models could successfully predict the movement of the subject according to the input data. This would not only speed up the process but also improve the quality of captured movements, preventing the affection of unexpected motion to some extent. Upon this, motion capture animation could be generated totally without hardware or motion data. The researchers from Nvidia proposed a new framework for training motion synthesis models by pose estimation from raw video and the result demonstrated impressive performance improvement in both pose estimation and motion synthesis results [15].

3 Limitations

Although the application of machine learning in computer graphics showed significant potential and advantages, limitations still exist for now. Two symbolic examples will be discussed as follows.

3.1 Requirements and costs for label marking

Supervised learning in machine learning necessitates the annotation of datasets to provide accurate labels for classification purposes. However, the field of computer graphics demands copious amounts of images and intricate datasets, which can pose challenges for proper annotation. The acquisition of high-quality labelled data becomes imperative for the generation of superior outcomes. At this juncture, the process of data annotation can be time-consuming, presenting a concern as machine learning endeavors to enhance the efficiency of computer graphics algorithms.

Furthermore, the task of accurately annotating labels typically requires the expertise of professionals well-versed in the nuances of such actions. Presently, certain companies offering label annotation services enlist non-specialized personnel. Although outsourcing the label annotation task to these enterprises may reduce time-related costs to some extent, the resultant quality of the annotated dataset remains a formidable issue. This, in turn, can yield disappointing final results in the generated images.

3.2 Inflexibility and lack of creativity

Machine learning is a data-driven discipline, wherein any generated outcome is rooted in the dataset from which it is learned. This became a well-recognized issue in the context of machine learning, highlighting its lack of genuine creativity. While the study of

AI strives to imbue machines with human-like thinking, a current challenge remains unaddressed: data-driven models possess the capability to generalize proficiently, yet they falter when attempting to reliably extrapolate beyond the confines of the data and prior information [16].

When applied to the realm of computer graphics, this predicament assumes critical importance, particularly in certain scenarios. Computer graphics necessitates creativity, especially in artistic contexts. While certain issues, such as the previously mentioned anti-aliasing problem, may not demand extensive creative input, matters like rendering style generation are substantially impacted. Rendering styles demand specificity and adaptability, often entailing the conveyance of emotions sought by the creator. Machine learning can solely generate outcomes rooted in the dataset, although it may assist by training models on datasets with specific styles. However, it's important to note that AI-generated works currently lack the ability to convey emotions akin to humans, as emotions remain beyond the scope of AI. Furthermore, the inherent inflexibility of machine learning methods becomes evident when stylistic requisites evolve. Altering the style prompts these models to undergo relearning from a new dataset, a process consuming considerable time. In contrast, human artists can swiftly adapt styles through parameter adjustments. Moreover, in extreme cases where the required style is exceedingly rare and insufficient data is available for model training, machine learning methodologies might even prove entirely futile.

4 Conclusions

This paper focuses on the application and certain limitations of machine learning in the field of computer graphics through illustrative examples. Including DLSS and DLAA in anti-aliasing, NNAO and AOGAN in ambient occlusion, machine learning in procedural generation, image-based and text-based model generation and motion capture. The limitation part presented the defects of machine learning methods in label marking, flexibility, and creativity. Undoubtedly, the fusion of machine learning with computer graphics has garnered widespread attention and adoption in both academia and industry. Overall, the integration of machine learning has ushered in a new era in computer graphics, exerting a profound influence on the future trends of this field. It not only enhances the efficiency of practitioners and reduces certain development costs but also optimizes performance while improving image quality and accuracy, providing a comprehensive problem-solving approach.

However, whether in the context of applying machine learning to computer graphics or in the broader domain of machine learning and AI, limitations and bottlenecks persist. An overreliance on machine learning without careful consideration can lead to negative consequences. Traditional computer graphics and machine learning-assisted computer graphics should complement each other, drawing mutual inspiration. Depending on the situation, the appropriate approach should be chosen to achieve the most satisfactory outcomes. The synergy between computer graphics and machine learning will foster increased communication between these two domains. As both fields continue to evolve, the future holds boundless possibilities.

References

1. Samuel, Arthur (1959). "Some Studies in Machine Learning Using the Game of Checkers". IBM Journal of Research and Development. 3 (3): 210–229.
2. R. Kohavi and F. Provost, (1998) "Glossary of terms," Machine Learning, vol. 30, no. 2–3, pp. 271–274.
3. Koza, J.R., Bennett, F.H., Andre, D. and Keane, M.A., (1996). Automated design of both the topology and sizing of analog electrical circuits using genetic programming. Artificial intelligence in design'96, pp.151-170.
4. Ethem Alpaydin (2020). Introduction to Machine Learning (Fourth ed.). MIT. pp. xix, 1–3, 13–18.
5. Keller, A., Křivánek, J., Novák, J., Kaplanyan, A. and Salvi, M., (2018). Machine learning and rendering. In ACM SIGGRAPH Courses . 1-2.
6. "GTC 2020: DLSS 2.0 - Image Reconstruction for Real-time Rendering with Deep Learning". NVIDIA Developer. 2020-06-09. Retrieved 2022-06-26.
7. Liu, E., 2020, March. DLSS 2.0-Image reconstruction for real-time rendering with deep learning. In GPU Technology Conference (GTC) (Vol. 8).
8. Yang, L., Liu, S. and Salvi, M., (2020), May. A survey of temporal antialiasing techniques. In Computer graphics forum, Vol. 39, No. 2, pp. 607-621.
9. Ren, L. and Song, Y., (2022). AOGAN: A generative adversarial network for screen space ambient occlusion. Computational Visual Media, 8(3), pp.483-494.
10. Holden, D., Saito, J. and Komura, T., (2016). Neural network ambient occlusion. In Special Interest Group on Computer Graphics, 2016 Technical Briefs (pp. 1-4).
11. Panagiotou, E. and Charou, E., (2020). Procedural 3D terrain generation using Generative Adversarial Networks. arXiv preprint arXiv:2010.06411.
12. NovySan, on Luma AI. Available at: https://lumalabs.ai/dashboard/imagine?slug=poise-playfully-xv-113031 (Accessed: 2 September 2023).
13. Gabrieldesign, on Luma AI. Available at: https://lumalabs.ai/dashboard/imagine?slug=mesmerizes-renown-5p-184905 (Accessed: 2 September 2023).
14. "Xsens MVN Animate – Products". Xsens 3D motion tracking. Retrieved 2019-01-22.
15. Xie, K., Wang, T., Iqbal, U., Guo, Y., Fidler, S. and Shkurti, F., 2021. Physics-based human motion estimation and synthesis from videos. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 11532-11541.
16. Carbone, M.R. (2022) When not to use machine learning: A perspective on potential and limitations. MRS Bulletin 47, 968–974. https://doi.org/10.1557/s43577-022-00417-z

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.