Artificial Intelligence for Computer Graphics: A Survey

Ryan Tang, Nguyen Thahn Lam, Fitria Zusni Farida, Bryan Anthony, Marietta Claudine

I. Abstract

***Abstract*—Artificial Intelligence, or AI, has as one of the most prominent keywords in recent years, impacting every facet of our daily lives. It is not an exaggeration to say that AI has become integral to every aspect of people's lives. In this rapidly changing technological landscape, we are on the verge of witnessing another breakthrough in AI, this time within the realm of Computer Graphics. The fusion of Artificial Intelligence (AI) and Computer Graphics has ushered in a breakthrough era in visual content creation and rendering. Artificial Intelligence (AI) has become a cornerstone in the contemporary technological landscape, permeating every facet of our daily lives. Within this technological upheaval, the amalgamation of AI with Computer Graphics stands out as a transformative force, ushering in a new era of visual content creation and rendering. This survey delves into the convergence of these two domains, offering a deeper exploration of the intricate applications of AI that revolutionize how we perceive and interact with digital imagery. Specifically, the survey aims to explore the current and potential application of AI-assisted computer graphics rendering in the industry, discovering new potentials in utilizing these state-of-the-art rendering technologies.**

II. Introduction

Rendering, a pivotal component in computer graphics, has experienced remarkable progress with the infusion of deep learning methodologies. Through the lens of Neural Networks, we explore innovative techniques, challenges, and breakthroughs that have reshaped how we perceive and generate visual content. In recent years, the integration of Artificial Intelligence (AI) techniques has revolutionized the field of Computer Graphics, unlocking unprecedented possibilities for realistic rendering, image enhancement, and super-resolution. As computational power continues to surge and neural network architectures evolve, the synergy between AI and Computer Graphics has paved the way for substantial advancements in visual content creation.

One facet of this survey focuses on the utilization of deep learning algorithms for image quality enhancement. Traditional image processing techniques often struggle with limitations in capturing intricate details and variety features. The appearance of the Deep Learning models, with their ability to discern complex patterns, offer a paradigm shift in image enhancement. We delve into the intricacies of these models, their training processes, and their impact on elevating visual fidelity.

Another significant sub-topic explored in this survey is AI chips. the fusing of AI chips in rendering processes is a central theme in our exploration. Rendering, a computational-intensive task, benefits significantly from specialized hardware designed to accelerate AI workloads. We dissect the architecture of AI chips that is suitable for rendering tasks, analyzing their impact on real-time graphics, and the overall user experience.

A noteworthy subtopic is the usage of AI real-time ray tracing in computer graphics. We discuss the integration of AI to optimize ray tracing algorithms, focusing on denoising, lighting, and material representation. Various techniques are introduced to enhance rendering efficiency and reduce noise, and the complexities of achieving accurate global illumination through sophisticated optimization strategies are also discussed, as well as alternative AI techniques for noise elimination in rendered images.

The primary objectives of this survey include providing an insightful overview of the evolving landscape where AI intersects with Computer Graphics, identifying key trends, and understanding the impact of Neural Networks and state of the art algorithms on rendering techniques. By examining specific sub-topics like **Deep Learning for Image Enhancement**, **AI chips**, **Adaptive Sampling and Denoising 135**, we aim to contribute to the broader understanding of how AI technologies are reshaping the visual computing landscape, as well as elucidate the mechanisms by which dedicated hardware optimizes the performance of AI algorithms, facilitating the seamless integration of intelligent applications.

III. Current Technologies used in Computer Graphics

A. Deep Learning for Image Enhancement

Image super-resolution is a pivotal task in computer vision, aimed at enhancing the resolution of low-quality images. This process serves not only to enhance the visual appeal of images, as exemplified by the implementation of techniques such as Deep Learning Super Sampling (DLSS) in video games, but also plays a crucial role in computer vision applications, including object detection and classification. Various methodologies have been employed to achieve image super-resolution. One approach involves identifying similarities within an image through geometric estimation within an expanded search space, as demonstrated by Dong in their work on image super-resolution [1]. Another technique employs sparse-coding methods, as discussed by Yang. [2]. In this context, a pipeline is utilized to encode image patches into a dictionary, subsequently reconstructing them within a higher-resolution dictionary, thereby achieving higher quality encoded patches. However, these existing techniques exhibit limitations and lack optimization. For instance, sparse-coding methods struggle to capture fine structures, necessitating the incorporation of additional mappings in the dictionary. This, in turn, results in computationally expensive processes. To address these shortcomings, utilizing neural networks or deep learning present a promising solution. By leveraging the learning capabilities of neural networks, it becomes possible to mitigate these issues without the need to explicitly encode all details of an image into memory, thereby reducing the overall computational cost in super sampling.

In the realm of deep learning for super resolution (DLSR), the classification can be dichotomized into two primary categories: supervised and unsupervised resolution. In the case of supervised DLSR, High Resolution (HR) images serve as reference benchmarks to evaluate the efficacy of image enhancement. Each model is characterized by distinct components encompassing upsampling techniques, network designs, and learning strategies, yielding unique outcomes in resolution enhancement [3].A conventional approach to supervised DLSR involves upsampling images through bicubic interpolation, followed by further upsampling utilizing traditional Convolutional Neural Network (CNN) methods. Nevertheless, this interpolation process incurs time costs, prompting the exploration of alternative methods. One such method involves the incorporation of a deconvolution layer to achieve upsampling effects, as exemplified by the Fast Super-Resolution Convolutional Neural Network (FSRCNN) [4]. Another strategy involves cascading, wherein images are progressively reconstructed to higher resolutions, as observed in the Laplacian Pyramid Super-Resolution Network (LapSRN). Additionally, iterative approaches have been proposed, wherein the model iteratively computes the reconstruction error of an HR image and employs the error to refine the HR image resolution.

Conversely, unsupervised DLSR primarily focuses on the task of downgrading high-resolution images to low-resolution counterparts and subsequently employing these downgraded images for the reconstruction of the original high-resolution images, in most cases in unsupervised learning, these models mostly operate as a generative adversarial network (GAN). GANs consist of two neural networks— a generator and a discriminator— engaged in a competitive learning process. The generator fabricates realistic data instances, while the discriminator evaluates and distinguishes between genuine and generated data. This dynamic leads to the refinement of the generator's ability to produce increasingly authentic outputs [5]. Various techniques have been introduced in the realm of unsupervised DLSR. A notable exemplar is the Progressively Cascaded Residual Network (PCARN). PCARN's network design is derived from the Cascading Residual Network (CARN), a type of recursive learning network that progressively learns higher-level features without introducing additional parameters. The model employs a recursive learning approach, featuring a cascading structure to self-learn and enhance efficiency. This design facilitates the incorporation of features not only from the preceding layer but also directly from the input, establishing multi-level shortcut connections. Such a design expedites the propagation of information from lower to higher layers and vice versa, particularly advantageous when utilizing backpropagation for parameter updates [6]. The inclusion of VGG loss in the model's learning strategy is noted for its capability to produce more photorealistic images compared to pixel-based error functions. Furthermore, the introduction of "Residual-E blocks" aims to reduce network parameters and depth [7].

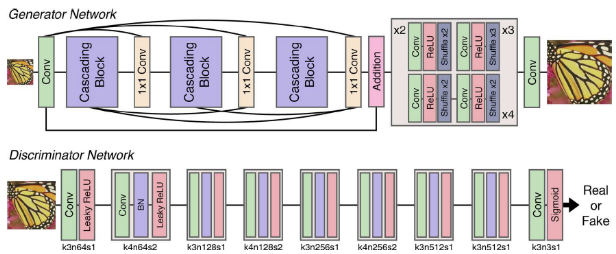

Fig. 1. The architecture of PCARN

Various techniques are employed to enhance the performance of both Unsupervised and Supervised DLSR. Notably, data augmentation is utilized to expose models to diverse scenarios, incorporating edge cases such as color modification, rotation, and scaling. This practice enables the model to adeptly handle a broader range of situations during training [8].Furthermore, improvements in network structures have proven instrumental in achieving enhanced accuracy in super resolution. These modifications involve the introduction of novel components, among which is pyramid pooling which is a layered pooling function [9], and also attention mechanism. The attention mechanism allows for the utilization of weighted features, enabling the model to focus on and learn crucial features pertinent to super resolution.

B. AI chips

This explores the pivotal role of AI chips, particularly GPUs, in the realm of computer graphics. Originally tailored for processing graphics-intensive tasks in games, GPUs are inherently designed with parallelism. Their exceptional performance, specifically suited for demanding deep learning AI algorithms reliant on parallel processing, positions GPUs as an ideal choice for AI hardware.

The widespread adoption of GPUs extends beyond gaming; they are extensively utilized in cloud and data centers for AI training, and find applications in automotive and security sectors. Presently, GPUs stand as the most widely used and versatile AI chips, encompassing over 30% of the market share. NVidia's series of GPUs, in particular, are extensively employed in the cloud for classification and deep neural network training. The GPU's multitude of computational cores enables an application throughput 10-100x greater than standalone CPUs, making them the primary choice for machine learning in numerous web and social media companies.

Additionally, specialized AI chips significantly contribute to performance enhancements. For instance, the Google Tensor Processing Unit (TPUv1) is renowned for various AI inference tasks in the cloud, including search queries and translation [10].

AI's Utilization of Big Data and the Phases of Model Execution: AI leverages big data as a cornerstone for training neural network models. These newly trained models, derived from extensive training datasets, acquire the ability to infer conclusions from fresh datasets. The training phase demands substantial computational power, necessitating high-end servers with advanced parallel performance. This phase, typically executed in the cloud using hardware, handles vast, diverse, and highly parallel datasets.

Conversely, the inference phase can occur either in the cloud or on edge devices. In comparison to training chips, inference chips require meticulous consideration of power usage, latency, and cost. Achieving a balance in these factors becomes crucial when executing the inference phase, particularly on edge devices.

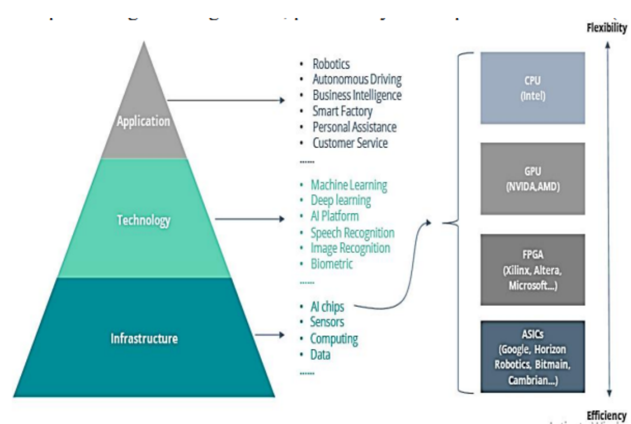

Fig. 2. Role of AI chip in the layers of AI.

Role of AI Chips in Computer Graphics: The Dominance of GPUs: This paper delves into the pivotal role of AI chips, specifically GPUs, in computer graphics. GPUs, originally designed to handle graphic-intensive tasks like gaming, possess inherent parallelism. Their high performance and parallel processing capabilities make them well-suited for demanding deep learning AI algorithms, establishing GPUs as a prime choice for AI hardware.

The widespread utilization of GPUs extends beyond gaming. They are extensively employed in cloud and data centers for AI training, as well as in automotive and security sectors. GPUs, currently the most flexible and widely used AI chips, dominate the market share, accounting for over 30%. NVidia's GPU series, in particular, is heavily employed in the cloud for classification and deep neural network training. Their thousands of computational cores enable a 10-100x application throughput compared to standalone CPUs, making them the conventional choice for machine learning in major web and social media companies.

While GPUs excel in performance, specialized AI chips also contribute significantly. The Google Tensor Processing Unit (TPUv1) stands as a prime example, extensively used for various AI inference tasks in the cloud, such as search queries and translation.

AI's Utilization of Big Data and the Phases of Model Execution: AI relies on big data as a foundation for training neural network models. These models, trained using large datasets, acquire the capability to infer conclusions from new datasets. The training phase demands substantial computational power, necessitating high-end servers with advanced parallel performance. This phase, typically executed in the cloud using hardware, processes vast, diverse, and highly parallel datasets.

Conversely, the inference phase can occur either in the cloud or on edge devices. In comparison to training chips, inference chips require more meticulous consideration of power usage, latency, and cost. Balancing these factors is crucial while handling the inference phase, especially when executing it on devices at the edge.

Memory Hierarchy and AI Chip Performance: One of the critical elements enhancing the performance and energy efficiency of AI chips lies in efficient data access through the memory hierarchy. Various types of memory are employed to empower AI chips, each serving specific purposes:

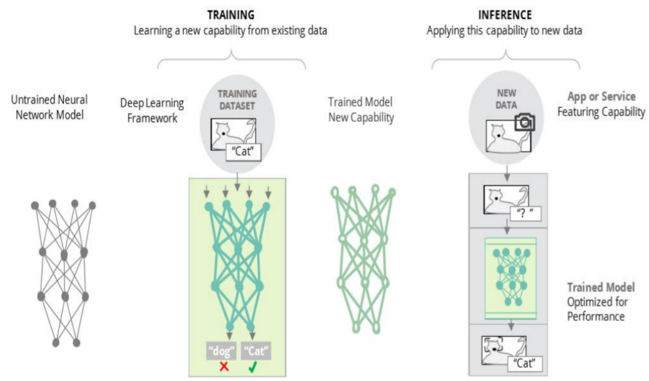

Fig. 3. Phases of deep learning.

- **AI-friendly memory:** AI and big data processing demand memory with high bandwidth and extensive storage capacity to facilitate parallel data access. Conventional nonvolatile memory (NVM) faces challenges in continual scaling. Emerging NVMs, with their growing bandwidth and capacity, offer promising solutions for AI chip memory technologies.
- **Commodity memory:** Off-chip memory, such as DRAM and NAND Flash memory, boasts dense cell structures and substantial capacity. Innovations like 3D integration, achieved through stacking multiple dies using through silicon via (TSV) technology or monolithic fabrication, effectively increase bandwidth and capacity. High bandwidth memory (HBM) and hybrid memory cube (HMC) are representative advancements in this realm.
- **On-chip (Embedded) Memory:** SRAM, an indispensable on-chip memory, facilitates seamless interaction between logic and memory circuits and aligns perfectly with logic devices. Its performance and density benefit from ongoing CMOS scaling. However, its volatility necessitates the use of on- or off-chip NVMs. Despite the widespread use of NOR Flash as on-chip NVM, its limitations, such as relatively slow access time and high write energy, impact system performance.

C. Real-Time Ray Tracing

Global illumination has been extensively studied as the need for high-quality 3D rendering for mixed reality, video games, and simulation has grown. A potent method for creating realistic images is Monte-Carlo path tracing, which can incorporate important effects like motion blur, depth of field, multiple illumination, caustics, color bleeding, and shadows. It is sluggish since it must trace a large number of rays in order to get a respectable quality. However, if the sample rate is too low, there will be a lot of noise in the produced data, which will prevent it from being used in real-time applications.

AI can be used to optimize ray tracing algorithms, improving the efficiency of rendering realistic lighting and reflections in scenes by enhancing different aspects, such as denoising, lighting, or material representation. Numerous researchers have proposed an end-to-end training framework that combines temporal reprojection, adaptive sampling, and deep learning to produce high-quality, temporally stable path traced animations at speeds that are almost real-time[11] to able to produce trustworthy estimation. Through the combination of adaptive sampling and temporal reuse, the network gains the ability to reconstruct challenging temporal effects like view dependent shading and disocclusion. Supported through a learned error-predicting module[12] Adaptive Sampling has some areas with bigger variance like edges or highly reflective surfaces might require a higher sample count.

Deep Adaptive Sampling and Reconstruction[13] (DASR): enhances low sample count rendering through the utilization of two Sampling Importance Convolutional Neural Networks (CNNs) in a UNET architecture. One CNN generates the sampling heatmap, output frame to contain sparse recommendations, while the other is responsible for denoising the averaged sampled pixel values[14], exploit the recommendation values information at different scales. DASR uniquely enables the back-propagation of gradients to the sampling importance network.

The traversal of multiple rays follows a random distribution, and for optimal outcomes, sparse sampling techniques[15] can be adapted to leverage additional information inherent in a ray tracer, such as depth, normal, and albedo data. These samples may then be confined to a specific density or organized into a regular grid with slight perturbations. This ensures a well-understood framework for image interpolation (polynomial within each triangle) and reconstruction, applicable in scenarios like super-resolution, image compression, and stereo rectification.

Reinforcement Learning-based Adaptive Sampling (REAP): reinforcement learning is employed to optimize the sampling importance network, avoiding the need for explicit numerical gradient approximations. Notably, this method refrains from aggregating sampled values per pixel through averaging. Instead, it retains all sampled values, feeding them into a latent space encoder. This encoder replaces manually crafted spatiotemporal heuristics with learned representations within a latent space. Finally, a neural denoiser is trained to refine the output image.

Numerous research efforts have focused on denoising algorithms because generating fully converged, noise-free images through rendering is typically prohibitively expensive and labor-intensive. The most effective strategy for mitigating or minimizing Monte Carlo noise involves deploying extensive convolutional neural networks (CNNs) to forecast distinctive feature-dependent filter kernels for each pixel[16]. The primary difficulty in this context lies in determining the suitable weights for each feature while simultaneously maintaining the intricate details of the scene. Spatial denoising as postprocessing is used to reduce the computational budget.

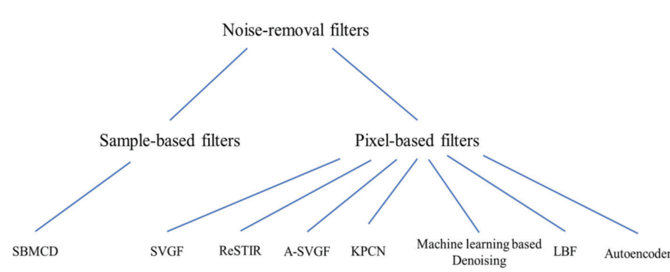

Fig. 4. Taxonomy of denoising methods.

Adaptive Spatiotemporal Variance-Guided Filtering (A-SVGF[17]): a cutting-edge approach addresses issues like flicker by employing real-time adaptive temporal filtering, estimating gradients. This technique incorporates spatiotemporal information from previously denoised and averaged rendered frames, utilizing various metrics such as averages, higher-order statistics, or the sampled pixel values directly to gain a comprehensive understanding of the sampled distribution. The integration of a temporal filter into the deferred renderer yields a notable improvement, ranging from 5% to 47%, in the alignment with reference images when compared to earlier crossover filters. To mitigate spatiotemporal loss of information, a spatiotemporal reservoir is employed, effectively creating a spatiotemporal latent space. While the encoded information in the latent space is primarily beneficial for the denoiser network, its utility for the sampling importance network is limited[14].

The Learning-based Filter (LBF) employs a multilayer perceptron neural network and features generated by the rendering system to produce filtered images with a broad range of distributed effects. The Reservoir-based Spatiotemporal Importance Resampling (ReSTIR) facilitates interactive sampling of direct lighting from thousands or millions of dynamic emissive triangles, supporting real-time rendering of millions of polygonal elements with shadows by reducing near-optimal variance and ensuring equal weighting. This filtering technique transforms denoising into a concurrent process rather than a post-processing step, integrating seamlessly with the rendering completion. Additionally, the Kernel-predicting Convolutional Networks (KPCN) method, known for its low Sample Per Pixel (SPP) requirement, outperforms the A-SVGF method with some adjustments due to its utilization of deep learning and neural networks. Enhanced outcomes with a reduced number of SPPs can be achieved by refining Sample-Based Motion Blur, Many Light Transmission Paths, and Depth-of-Field (SBMCD) methods, where individual samples are strategically placed on adjacent pixels[18].

Recursive denoising AutoEncoders[19](RAE): Neural networks following the encoder-decoder architecture, such as UNETs, are designed to correct imperfect inputs and replicate them in the output. The autoencoder's objective is to acquire the ability to reverse the corruption in its input. When the hidden layer of the encoder operates in a nonlinear fashion, its behavior differs from Principal Component Analysis (PCA), avoiding the loss of detailed image information and enabling the capture of multi-modal aspects within the input distribution. To achieve an optimal solution, training involves employing fuzzy-based techniques, treating the network as an activation function over a finite period. This approach aids in reducing uncertainty within stacks of autoencoders[20].

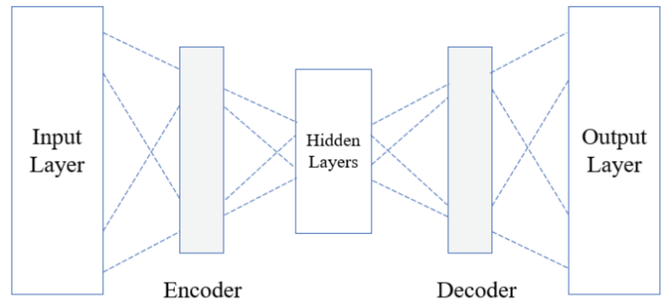

Fig. 5. Overview of autoencoder structure.

Achieving accurate global illumination involves complex algorithms to implement a camera-centric approach with multiple light bounces. Developing, optimizing, and debugging such algorithms can be time-consuming and require advanced knowledge of computer graphics and ray tracing techniques. Hence, achieving real-time performance necessitates sophisticated optimization strategies to balance visual quality and computational efficiency.

Various optimization techniques like bounding volume hierarchies (BVH), parallel processing using GPU (CUDA or OpenCL), and denoising algorithms to reduce noise in rendered images are employed to meet real-time rendering requirement.

K-Dimensional Trees, CUDA (Compute Unified Device Architecture), and Iso Surfaces can be used in conjunction to optimize various aspects of the ray tracing pipeline. A KD-tree can accelerate the intersection tests between rays and scene geometry, improving the efficiency of the ray tracing process by reducing the number of calculations needed as it is a spatial data structure used to organize the scene geometry into a hierarchical structure, allowing for efficient traversal. Developed by NVIDIA, CUDA enables developers to harness the parallel processing power of GPUs for speeding up various computationally intensive workloads. It can parallelize each independent graphics rendering operations, which performed independently for each pixel or ray, like ray-object intersections, shading, etc. Leveraging the GPU's processing power and significantly speeding up computations, allowing for real-time or interactive rendering. In computer graphics, iso surfaces are often used to enhance the visual representation of volumetric data within a scene, such as medical imaging or scientific simulations, providing realistic rendering effects. They are particularly useful for identifying regions with specific properties or values within a three-dimensional space with its Marching Cubes algorithm used to extract iso surfaces from volumetric data.

Some game developers adopt hybrid approaches, combining rasterization techniques with ray tracing for specific effects, achieving a balance between performance and realism. Moreover, alternative artificial intelligence techniques such as neural networks and fuzzy logic can be applied to eliminate noise in the resultant images[18].

IV. Evaluation of Technologies Used

A. DLSR: Current Application and Potential Usage in the Future

Application of Deep Learning Super Resolution (DLSR) varies significantly across industries. For instance, NVIDIA has developed its proprietary adaptation of DLSR, known as "Deep Learning Super Sampling" (DLSS), where it introduced antialiasing technology for better rendering, specifically to enhance the gaming experience for users,hence this technology is targeted mainly towards the video game industry. While the detailed architecture of DLSS is not publicly disclosed, NVIDIA utilizes its Tensor cores to train models for DLSS. Users then receive updated models, enabling them to utilize the latest advancements in graphic rendering.[21]

In the realm of remote sensing, where information about objects or phenomena is acquired without physical contact, deep learning super resolution finds applications, especially in satellite imagery. The primary focus is often on small object detection, such as vehicles and people. Super resolution techniques contribute by enhancing images, introducing additional distinguishable features that aid in improved classification and detection. Various neural architectures have been proposed to address this challenge. [22]

Deep Learning Super Resolution (DLSR) finds application in the reconstruction of turbulence, as highlighted by Kim et al. [23]. Specifically, the utilization of the cycle-consistent generative adversarial network (CycleGAN) has been instrumental in this context. CycleGAN employs cycle-consistency losses as its primary loss function, a critical aspect contributing to the assurance of data dependency by constraining the image and domain space. This mechanism enhances the reliability of generated data.

In the context of turbulence reconstruction, a Convolutional Neural Network (CNN) is integrated into the methodology to predict the target flow field. The objective function governing this CNN is subjected to regularization through L2 regularization. Additionally, the overall loss function is based on the summation of Mean Squared Error (MSE), signifying a comprehensive approach to optimizing the predictive accuracy of the model in reconstructing turbulent phenomena. This integrated framework demonstrates the synergistic utilization of generative adversarial networks and convolutional neural networks for turbulence reconstruction, underscoring the versatility of DLSR methodologies in diverse scientific domains.

For example, the Deep Memory Connected Network (DMCC), a novel neural network architecture, has been introduced as an advancement over previous models like SRCNN and LGCNet. The motivation behind DMCC arises from the limitations of earlier architectures in achieving satisfactory image reconstruction [24]. DMCC incorporates deeper neural layers to capture more intricate features and employs an hourglass model as a sequence model for efficient upsampling and downsampling of images, thereby reducing the time complexity associated with image reconstruction.

This application exemplifies the ongoing evolution of neural network architectures tailored to address specific challenges. In many instances, there is a notable emphasis on advancing deep learning super resolution techniques for image enhancement and analysis. Despite these strides, challenges persist, particularly when confronted with the processing of large quantities of images from satellites.

The endeavor to process all images simultaneously using super resolution methods introduces complexities [25].Furthermore, challenges are encountered in the domain of data quality. Given the inherent imperfections in real-world images, perfection is unattainable. Issues such as image degradation may persist, and data augmentation may prove insufficient in addressing certain anomalies. Additionally, some images may inherently possess super low resolution, compounding the overall difficulty in training models. As a result, the development of a robust training model becomes imperative to effectively navigate and overcome these challenges.

B. Advantages and Potential Drawbacks of AI Chips in GPU Computer Graphics

1) Advantages: AI chips, particularly in the context of GPU-driven computer graphics, offer numerous advantages in addressing critical technological challenges[26]:

1) **Data Security and Privacy:** Edge AI chips allow local data processing, reducing the risk of sensitive information exposure. They enable devices like security cameras to analyze and selectively transmit relevant video segments, mitigating privacy concerns and minimizing data sent to the cloud.
2) **Low Connectivity:** Embedded machine learning in devices like drones facilitates real-time decision-making without requiring constant internet connectivity, enhancing safety measures, such as identifying swimmers in perilous conditions without relying on internet connection.
3) **Handling Vast Data:** Integration of machine learning processors, like vision processing units (VPUs) in cameras, enables real-time data review, reducing the burden of transmitting all data to the cloud. This significantly cuts down storage costs and bandwidth usage.
4) **Power Efficiency:** AI chips designed for low-power consumption, such as ARM chips, enable efficient data analysis without draining device batteries excessively. This allows for innovative healthcare applications, like analyzing inhalation lung capacity in respiratory inhalers, aiding personalized care for asthma patients.
5) **Low Latency Requirements:** Edge AI chips offer nearly instantaneous on-device data processing, critical for tasks demanding real-time responses. For instance, in autonomous vehicles, these chips enable immediate data analysis and decision-making for safe navigation, significantly reducing latency compared to remote data center computations.

2) Potential Drawbacks: While AI chips offer significant advantages, there are potential limitations to consider:

- **Complexity of Implementation:** Implementing and optimizing AI chips, especially specialized ones, might require substantial technical expertise and resources.
- **Initial Costs:** Initial investment costs associated with integrating AI chips into existing systems could be a deterrent for some entities.
- **Hardware Limitations:** Constraints in processing power, memory, or compatibility with evolving AI methodologies might persist in certain AI chip implementations.

Overall, while AI chips provide substantial benefits in GPU-related computer graphics by addressing various technological challenges, careful consideration of potential drawbacks is crucial for their successful integration and utilization.

C. An Academic Exploration: Leveraging AI-Assisted Computer Graphics

The exploration of computer graphics has been proposed for a long time, especially the surge in AI art generators for two-dimensional output. However 3D modeling pipelines are still limited because of the complexity and require specialized knowledge.The 3D models created are also not optimized and are not created for rendering and are often incomplete. The proposed solution suggests leveraging advances in AI, specifically natural language processing and visual content generation, to create 3D models through text prompts in interoperable formats [27]. The aim is to address existing constraints, simplify the development of content for extended reality (XR), and pave the way for Web 3.0 and immersive social environments in the metaverse.

With the rise of the metaverse and immersive environments, there is an increasing demand for 3D asset creation. Popular software tools like Sketchfab, Blender, Maya, and Autodesk 3ds Max are used for this purpose. Assets can be viewed in these applications or imported into game engines like Unreal Engine 5 and Unity. The review notes the existence of marketplaces within game engines where developers can download and sell 3D assets. However, it emphasizes a limitation in the finite resources of these marketplaces, leading to a reliance on existing assets that may need modification. The conclusion is that there is a growing need for high-quality, editable, and reconfigurable 3D assets, particularly in industries focused on immersive content creation for the metaverse [28]. The application of Artificial Intelligence to accelerate computer graphic for rendering metaverse will be delved as follow [29]:

1) **Game Engines and Cinematics:** 3D models and game engines are increasingly used in industries such as film. The mainstream film industry has witnessed a surge in the adoption of game development software like Unreal Engine 5. Notably, the television series ”The Mandalorian” is cited as a pioneering example, utilizing game engine technology to create realistic virtual sets that can dynamically change based on needs and camera positions, while also accurately reflecting lighting information. This innovative approach has streamlined the filmmaking process, saving time in post-processing, and has received praise from actors who can now interact with a virtual world on set instead of working in front of a green screen.
2) **3D Modelling Process:** Despite the advancements in game-engine technology, 3D modeling remains a specialized task due to the intricacies of the development pipeline. Traditional 3D modeling starts with basic geometry, like polygons, often involving complex techniques such as retopologizing and the use of rendering tools. Game models, optimized for performance, can consist of thousands of polygons. This complexity demands specialized skills and techniques, impacting the ability to efficiently process and render polygons for real-time applications like games. The next section will delve into two key aspects of the asset creation pipeline: model creation and optimization for use in game engines or other real-time applications.
3) **3D Scanning and Photogrammetry:** The technology of 3D scanning, although not new, has seen continuous improvement in quality, accessibility, and adaptability. Modern scanners, such as handheld devices like Artec 3D or smartphone applications like Scaniverse and Polycam utilizing LiDAR, allow users to create 3D models of objects, large and small, as well as entire areas. Photogrammetry involves the generation of 3D models from photographs or other data. While not a new concept, photogrammetry has evolved from creating 2D information using photographs, sonar, and radar. The technique has found applications in various fields, including entertainment, where it has been used in films like ”The Matrix” and in the development of video games. [30]
4) **AI Generated Content:** The evolving role of AI in art creation is transitioning from aiding artists to becoming a primary method for generating 2D and 3D art. Notably, AI art generators like DALL-E 2 create imaginative works, evolving toward photorealistic art with each iteration. AI's ability to learn from previous versions is emphasized, with artists using generated images for inspiration. The shift from 2D to 3D content generation introduces AI systems transforming text prompts into 3D models. While 3D model creation from text prompts is in early stages, AI systems like Nvidia's Instant NeRF and Kaedim use 2D images as references for 3D model generation. Kaedim, designed for 3D artists, considers technical requirements but requires human reviewers. Nvidia's NeRF utilizes inverse rendering, creating 3D models from a collection of 2D images. These AI-driven 3D model creation methods are becoming more user-friendly, accessible via smartphone applications, removing barriers to 3D model creation.
5) **3D Model Optimization:** In the proposed immersive development pipeline, another critical aspect is optimizing 3D models for real-

time applications. While the methods described earlier can create detailed models, they often result in dense meshes unsuitable for performance and animation [31]. This issue is prevalent in current processes, particularly in 3D modeling for entertainment that often involves digital sculpture, as seen in popular applications like ZBrush [32]. Digital sculpture can lead to dense, unusable meshes, necessitating a time-consuming retopology process. This involves creating a lower-resolution, optimized version of the model suitable for game engines, with high-resolution details added during rendering. The process includes baking higher-resolution details into color images called normal maps, a practice dating back to the early 2000s [33]. Recognizing the challenges of retopology, software developers are introducing auto-retopology tools, with Nvidia being a pioneering force in graphics technology. Nvidia's recent innovations, NGX and NeRF, leverage AI and deep learning, enhancing graphic output. While NGX can modernize older, lower-resolution graphics, it primarily focuses on textural graphics, making it ideal for lower resolution models. Higher resolution models still require optimized geometry, often achieved through retopology [34]. Epic Games, through Unreal Engine 5, introduces a novel approach to address retopology challenges. The Nanite system, built on earlier research, allows users to import non-optimized high-resolution 3D models directly, eliminating the need for retopology. Nanite dynamically adjusts the level of detail in real-time and automatically occludes invisible polygons, optimizing performance for real-time applications. The system also improves rendering with virtualized textures, all handled automatically. These optimizations remove technical obstacles, allowing developers to focus on enhancing the viewer experience. Nanite and similar systems mark a significant shift by eliminating time-consuming and technically challenging aspects of 3D asset creation.

V. Conclusion

In conclusion, the survey on ”Artificial Intelligence for Computer Graphics” has navigated the expansive landscape where the realms of AI and graphics seamlessly converge. The synergy of these domains has not only reshaped our understanding of visual content creation but has also set the stage for unprecedented advancements.

Deep Learning Super Resolution has emerged as a beacon of innovation within computer graphics. Through the lens of neural networks, the project explored intricate techniques, unveiling a spectrum of possibilities in the enhancement of visual content. The journey through this sub-section highlighted the transformative impact of deep learning methodologies on image quality, realism, and overall user experience. As we stand at the intersection of AI and image enhancement, it is evident that the fusion of these technologies heralds a new era in the field of computer graphics.

Simultaneously, the exploration of AI Chips underscored the pivotal role hardware plays in catalyzing the potential of artificial intelligence. The survey delved into the architecture, capabilities, and implications of AI chips, unraveling the nuanced relationship between hardware advancements and the efficiency of AI algorithms. The significance of optimized hardware for accelerated computations in graphics-intensive tasks became apparent, signifying a symbiotic relationship that propels the field forward.

Global illumination techniques are crucial for high-quality 3D rendering in mixed reality, video games, and simulations. Monte-Carlo path tracing produces realistic images but is slow for real-time use due to noise. AI optimizes ray tracing, enhancing efficiency in denoising and lighting. Researchers propose frameworks combining temporal reprojection, adaptive sampling, and deep learning for high-quality, nearly real-time path-traced animations. Techniques like Deep Adaptive Sampling and Reinforcement Learning-based Adaptive Sampling improve low sample count rendering. Denoising algorithms, including Learning-based Filters and Recursive denoising AutoEncoders, refine images. Optimizations like bounding volume hierarchies and GPU parallel processing contribute to real-time rendering, especially with hybrid approaches blending rasterization and ray tracing in game development.

This paper also discusses the evaluation of the technologies discussed through their pros and cons or their current application in the industry, and brings out one big potential in utilizing AI for computer graphics in general, which is rendering the metaverse. Techniques proposed such as natural language processing and visual content generation can simplify the development of 3D assets which are in high demand in the metaverse. Other aspects of how AI can improve rendering are also explored, including the application of AI in game engines, cinematography,3D modeling processes, 3D scanning and photogrammetry, and art generation.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," 2015.

[2] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," no. 1, pp. 1–8, 2008.

[3] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 2021.

[4] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," *CoRR*, vol. abs/1608.00367, 2016. [Online]. Available: http://arxiv.org/abs/1608.00367

[5] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014.

[6] N. Ahn, B. Kang, and K.-A. Sohn, "Efficient deep neural network for photo-realistic image super-resolution," *Pattern Recognition*, vol. 127, p. 108649, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320322001303

[7] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1637–1645.

[8] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *CoRR*, vol. abs/1712.04621, 2017. [Online]. Available: http://arxiv.org/abs/1712.04621

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *CoRR*, vol. abs/1406.4729, 2014. [Online]. Available: http://arxiv.org/abs/1406.4729

[10] S. M. Viswanathan, "Ai chips: New semiconductor era," vol. 7, pp. 14 687–14 692, 08 2020.

[11] J. Hasselgren, J. Munkberg, M. Salvi, A. Patney, and A. Lefohn, "Neural temporal adaptive sampling and denoising," *Computer Graphics Forum*, vol. 39, no. 2, p. 147–155, 2020.

[12] B. Xu, J. Zhang, R. Wang, K. Xu, Y.-L. Yang, C. Li, and R. Tang, "Adversarial monte carlo denoising with conditioned auxiliary feature modulation," *ACM Transactions on Graphics*, vol. 38, no. 6, p. 1–12, 2019.

[13] A. Kuznetsov, N. K. Kalantari, and R. Ramamoorthi, "Deep adaptive sampling for low sample count rendering," *Computer Graphics Forum*, vol. 37, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:49341310

[14] A. Scardigli, L. Cavigelli, and L. K. Müller, "Rl-based stateful neural adaptive sampling and denoising for real-time path tracing," *ArXiv*, vol. abs/2310.03507, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:263671619

[15] T. Viitanen, M. Koskela, K. Immonen, M. J. Mäkitalo, P. O. Jääskeläinen, and J. H. Takala, "Sparse sampling for real-time ray tracing," in *VISIGRAPP*, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:45171909

[16] S. Bako, T. Vogels, B. McWilliams, M. Meyer, J. Novák, A. Harvill, P. Sen, T. DeRose, and F. Rousselle, "Kernel-predicting convolutional networks for denoising monte carlo renderings," *ACM Transactions on Graphics (TOG)*, vol. 36, pp. 1 – 14, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:31004998

[17] C. Schied, A. Kaplanyan, C. Wyman, A. Patney, C. R. A. Chaitanya, J. Burgess, S. Liu, C. Dachsbacher, A. E. Lefohn, and M. Salvi, "Spatiotemporal variance-guided filtering: Real-time reconstruction for path-traced global illumination," *Proceedings of High Performance Graphics*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:202716856

[18] S. Malekmohammadi Faradounbeh and S. Kim, "Evaluation of artificial intelligence-based denoising methods for global illumination," *Journal of Information Processing Systems*, vol. 17, pp. 737–753, 08 2021.

[19] C. R. A. Chaitanya, A. Kaplanyan, C. Schied, M. Salvi, A. E. Lefohn, D. Nowrouzezahrai, and T. Aila, "Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder," *ACM Transactions on Graphics (TOG)*, vol. 36, pp. 1 – 12, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:3350221

[20] B. Costa and J. Jain, "Fuzzy deep stack of autoencoders for dealing with data uncertainty," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2019, pp. 1–6.

[21] A. Watson, "Deep learning techniques for super-resolution in video games," *CoRR*, vol. abs/2012.09810, 2020. [Online]. Available: https://arxiv.org/abs/2012.09810

[22] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, 04 2019.

[23] H. Kim, J. Kim, S. Won, and C. Lee, "Unsupervised deep learning for super-resolution reconstruction of turbulence," *Journal of Fluid Mechanics*, vol. 910, p. A29, 2021.

[24] W. Xu, G. XU, Y. Wang, X. Sun, D. Lin, and Y. WU, "High quality remote sensing image super-resolution using deep memory connected network," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 8889–8892.

[25] Y. Liu, Y. Qiao, Y. Hao, F. Wang, and S. F. Rashid, "Single image super resolution techniques based on deep learning: Status, applications and future directions," *Journal of Image and Graphics*, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239043521

[26] S. M. Viswanathan, "Ai chips: New semiconductor era," vol. 7, pp. 14 692–14 693, 08 2020.

[27] M. Oussalah, "Ai explainability. a bridge between machine vision and natural language processing," in *Pattern Recognition. ICPR International Workshops and Challenges*, A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini,

H. J. Escalante, and R. Vezzani, Eds. Cham: Springer International Publishing, 2021, pp. 257–273.

[28] J. J. Korbel, U. H. Siddiq, and R. Zarnekow, "Towards virtual 3d asset price prediction based on machine learning," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 17, no. 3, pp. 924–948, 2022.

[29] J. Ratican, J. Hutson, and A. Wright, "A proposed meta-reality immersive development pipeline: Generative ai models and extended reality (xr) content for the metaverse," *Journal of Intelligent Learning Systems and Applications*, vol. 15, 2023.

[30] A. R. C. Mamani, S. R. R. Polanco, and C. A. C. Ordonez, "Systematic review on photogrammetry, streaming, virtual and augmented reality for virtual tourism," in *HCI International 2022–Late Breaking Papers: Interacting with eXtended Reality and Artificial Intelligence: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, vol. 13518. Springer Nature, 2022, p. 46.

[31] E. Touloupaki and T. Theodosiou, "Performance simulation integrated in parametric 3d modeling as a method for early stage design optimization—a review," *Energies*, vol. 10, no. 5, p. 637, 2017.

[32] B. Raitt and G. Minter, "Digital sculpture techniques," *Interactivity Magazine*, vol. 4, no. 5, 2000.

[33] T. Tasdizen, R. Whitaker, P. Burchard, and S. Osher, "Geometric surface processing via normal maps," *ACM Transactions on Graphics (TOG)*, vol. 22, no. 4, pp. 1012–1033, 2003.

[34] J. Wu, C. Dick, and R. Westermann, "A system for high-resolution topology optimization," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 3, pp. 1195–1208, 2015.