Virtual Staging Technologies for the Metaverse

Muhammad Tukur+
College of Science and Engineering
HBKU, Doha, Qatar

Yehia Boraey+
Department of Computer Science
Qatar University, Doha, Qatar

Sara Jashari+
College of Science and Engineering
HBKU, Doha, Qatar

Alberto Jaspe Villanueva
VCCVIS, High Perf. Visualization
KAUST, Jeddah, Saudi Arabia

Uzair Shah
College of Science and Engineering
HBKU, Doha, Qatar

Mahmood Al-Zubaidi
College of Science and Engineering
HBKU, Doha, Qatar

Giovanni Pintore
Visual and Data-intensive Computing
CRS4, Cagliari, Italy

Enrico Gobbetti
Visual and Data-intensive Computing
CRS4, Cagliari, Italy

Jens Schneider
College of Science and Engineering
HBKU, Doha, Qatar

Marco Agus*
College of Science and Engineering
HBKU, Doha, Qatar

Noora Fetais*
Department of Computer Science
Qatar University, Doha, Qatar

***Abstract*—We discuss virtual staging technologies, focusing on two primary pipelines for creating and exploring immersive indoor environments in the metavers: an AI-based image processing pipeline and a LIDAR-based pipeline. The AI-based image processing pipeline leverages advanced AI algorithms for tasks such as clutter removal, semantic style transfer, and super-resolution, enabling rapid generation of high-quality, photorealistic virtual environments from single panoramic images. The LIDAR-based pipeline captures measurable 3D models of indoor spaces, facilitating immersive editing and collaborative design through real-time interaction with high-fidelity virtual environments. A qualitative comparative analysis of these technologies highlights their strengths and limitations in various applications. The practical implications of these pipelines are discussed, particularly their potential to transform industries such as real estate, furniture retail, interior design, construction, remote collaboration, and immersive training. The paper concludes with suggestions for future research, including conducting user studies, integrating the two pipelines, and optimizing technologies for mobile and edge devices to enhance accessibility and usability.**

***Index Terms*—Metaverse, Virtual Staging, LIDAR, AI, Immersive Environments, Collaborative Design**

I. INTRODUCTION

The metaverse represents a convergence of virtually enhanced physical and digital realities, creating a collective virtual shared space [43]. This expansive digital universe allows users to interact, socialize, and engage in various activities through avatars and digital representations. Applications of the metaverse span from gaming and social networking, to education and virtual commerce [12], necessitating advanced technologies for creating and interacting with immersive environments [44].

A potential application of these technologies is Virtual Staging, where the goal is to create and interact with believable alternative designs of real indoor environments. Virtual staging is particularly useful in real estate and interior design, where it helps visualize captured furnished spaces without physical furniture, replacing it with virtual alternatives [23]. This requires end-to-end solutions, going from the capture of 3D environment, to their processing to recover the required data to perform editing, the implementation of the design, and the virtual immersive presentation.

This paper contributes to the field by analyzing two distinct pipelines for the production and exploration of immersive environments: an AI-based image processing pipeline for exploratory purposes, and a LIDAR-based pipeline focusing on immersive editing and collaborative design [38].

1) *AI-Based Image Processing Pipeline:* This pipeline leverages advanced AI algorithms that are integrated and synergized to process and enhance panoramic images of indoor environments for virtual staging applications in the metaverse (see Fig. 1). Key integrated technologies include a system for clutter removal [23], multitask dense prediction, semantic photorealistic style transfer [42], super-resolution [46], a rendering system [41], and a system for the automatic generation of stereoscopic environments for metaverse applications [30].
2) *LIDAR-Based Pipeline:* This pipeline focuses on capturing measurable 3D models of indoor spaces using LIDAR technology. The captured data is then used for immersive editing and collaborative design. This approach allows users to interact with and modify virtual environments in real-time, providing a high level of detail and accuracy. For example, the use of LIDAR scanning in applications like Spatial enables detailed

---

This publication was made possible by NPRP-Standard (NPRP-S) 14th Cycle grant 0403-210132 AIN2 from the Qatar National Research Fund (a member of Qatar Foundation).

+: Equal contribution

*: Corresponding authors: magus@hbku.edu.qa,n.almarri@qu.edu.qa.

virtual walkthroughs and object placement, enhancing the realism of virtual staging.

A qualitative comparison of these two technologies is also presented, demonstrating their respective strengths and limitations in enabling immersive editing and collaborative design. For instance, while LIDAR provides high accuracy and detailed spatial data, AI-based methods offer flexibility and efficiency in image processing, making them suitable for rapid staging and exploratory tasks. This analysis aims to highlight the practical implications of adopting these technologies and offers insights into future developments in virtual staging for the metaverse. The practical implications of these technologies in various industries are also discussed, highlighting their potential to transform fields such as real estate, furniture retail, interior design, construction, remote collaboration, and immersive training.

II. RELATED WORK

This work deals with technologies for immersive editing of 3D indoor environments and with methods for automatic processing of panoramic images. For space reasons, we do not aim to provide here an extensive survey of the related literature: we refer interested readers to the surveys related to scene understanding from panoramic images [10], 3D geometry estimation from 360 images [6], on 3D reconstruction of indoor environments [31], and on construction and maintenance of building geometric digital twins [8]. In the following, we discuss the most recent technologies related to our work.

a) Immersive editing for virtual staging: In general, the application of immersive technologies and digital twins in the construction industry is still in the early stages [1]. For what concerns interior design, immersive frameworks focusing on material selection [48], or lighting design [33] have been proposed, but the usage of immersive technologies gained most of popularity during the pandemic for the generation of virtual tours experiences [28], [37], especially for real estate applications [5], and interior design [45], through the usage of 3D scanning technologies or specialized cameras [47]. Very recently, methods exploiting deep learning technologies has been considered for the automatic generation of 3D indoor scenes, through graph convolutional networks [9], or text-based generative models [11], but these methods are still far from real-world application. In this work, we show how current LIDAR technologies can be used to support Metaverse-enabled design of indoor spaces.

b) AI-based technologies for indoor panoramic images: Omnidirectional cameras are very popular for the fast and accurate acquisition of indoor scenes since they can capture most of the 3D content with few shots [31]. In the last decade, various data-driven technologies have been developed to support the automatic creation of digital content from panoramic images. Examples include: extraction of room layouts [14], [25], [26], geometry estimation [27], [35], signal extraction for inverse rendering [34], clutter removal [23], style transfer [42], and novel pose synthesis for virtual exploration [29], [30]. Those methods have been applied to pipelines for inverse rendering [17], and for virtual staging of indoor scenes [13], [41]. This paper illustrates the integration of various AI panoramic image processing components to automatically generate immersive exploration experiences of refurnished indoor environments.

III. AI-IMAGE-BASED PIPELINE

We developed an end-to-end AI-image-based pipeline based on semiautomatic processing and editing of single panoramic images for performing image-based virtual staging tasks and deploying the generated environments in Metaverse-ready web resources for immersive exploration. The framework is depicted in Figure 1 and integrates various components for automatic processing, generation, and stereoscopic exploration of immersive indoor environments. Starting from single panoramic images (left), our framework removes clutter, estimates the necessary signals for both cluttered and decluttered scenes, and applies semantic style transfer to enhance visual aesthetics. The pipeline further increases detail by employing a super-resolution model, enabling high-fidelity editing. The result is the presentation of immersive, high-resolution spherical indoor scenes that can be explored using VR setups on lightweight WebXR viewers (c.f. Fig. 1(C-D)), making them ready for Metaverse applications (right). In the following sections, we will provide an overview of the various components and discuss how they are integrated and synergized to generate an immersive environment for virtual staging applications in the metaverse.

a) Instant clutter removal: For this task, we use the model presented in [23], which is a deep learning architecture to automatically remove clutter from panoramic indoor images, providing both a photorealistic view and depth estimation of the empty scene. This end-to-end solution leverages a lightweight deep learning network to process 360-degree panoramic images, distinguishing between permanent architectural features and removable clutter. The method begins by generating an attention mask to identify cluttered regions based on geometric differences between cluttered and uncluttered scenes. This mask guides the subsequent image and depth generation, using gated convolutions and high-order geometric constraints during training to ensure the output is both visually and geometrically plausible (c.f. Fig. 2 [A-top]).

The approach is unique in its holistic treatment of the entire scene, as opposed to traditional methods that focus on single object removal. It shifts the computational burden to the training phase, allowing for interactive-speed performance during inference.

b) Multitask dense prediction: Signals such as semantic segmentation, depth, color-coded normals, and intrinsic decomposition signal distinguishing reflectance (albedo) and shading are crucial for generating immersive and interactive environments for virtual staging applications in the metaverse. To this end, we developed a deep-learning framework designed to infer multiple pixel-wise signals from a single panoramic

Fig. 1. **AI-Image-Based Pipeline:** Several AI models are integrated and synergized to process and generate the virtual staging environment. The framework is composed of four integral modules: *(A) Input Module:* A single RGB panoramic image of an indoor environment taken using a 360○ camera. *(B) Deep Learning Models:* This component comprises a network designed for the extraction of various signals. Our comprehensive approach autonomously procures essential signals such as depth, semantics, shading, reflectance, color-coded normals, and empty scene representation, thereby synergizing the generation of virtual staging applications for the metaverse. *(C) Example of virtual staging process leveraging the AI-Based Image Processing Pipeline:* [42] is used for semantic style transfer, a super-resolution model [4] is used to increase details, a rendering system [41] is used for interactive exploration, object insertion, and editing etc. *(D)* A framework for the automatic generation and exploration of immersive scenes representing indoor stereoscopic environments, which can be navigated using VR setups on lightweight WebXR viewers, making it ideal for Metaverse applications [30].

image [34]. The framework is able to concurrently extract diverse types of information—such as depth, normals, semantic segmentation, reflectance, and shading—from indoor panoramic images (See Fig. 2 [B] for examples). This is achieved through a transformer-based encoder-decoder architecture that leverages multiple heads for dense estimation. By incorporating a context adjustment layer, the framework ensures effective knowledge distillation between the encoder and various decoder heads, enhancing the quality of predictions for each signal.

c) Indoor style transfer: Style transfer is used for changing the appearance of indoor environments to look like target scenes. It enhances user engagement, making it a valuable tool for virtual staging. In our pipeline, we integrated a semantic photorealistic style transfer approach tailored for indoor panoramic images proposed in [42].

The methodology integrates several components into a generative adversarial network (GAN) framework. Firstly, it employs a shading decomposition scheme to separate reflectance (albedo) from shading (c.f. Sec. III-0b), thus preventing shading-related artifacts during the style transfer process. This ensures that the style transfer affects only the intrinsic colors of surfaces, not their illumination. Secondly, the architecture incorporates strong geometry constraints through the use of layout and depth inference during training, enforcing shape consistency between the generated and ground truth scenes. This is achieved by introducing custom geometry-aware losses that account for the 3D characteristics of indoor scenes, including clutter, layout, and edges. Additionally, the method applies the GAN-based super-resolution technique (See Sec. III-0d) to enhance the detail and resolution of the generated images, making them suitable for immersive applications. The visual results (c.f. Fig. 2 [C-D]) confirm the effectiveness of the method in producing realistic and visually pleasing indoor scenes.

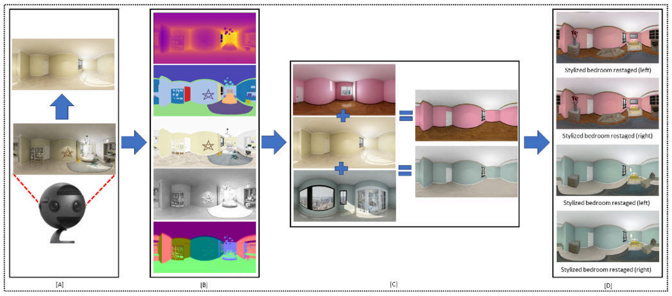

Fig. 2. **AI-Image-Based Pipeline Demo:** *(A)* Input: a spherical shot of a panoramic indoor scene (bottom) and an example of instant automatic emptying of the scene (top). *(B)* Example of multiple inferences obtained with a multitask dense prediction model on a single synthetic RGB image from Structured3D [49]. From top to bottom, the depth prediction, semantic inference, reflectance, shading, and color-coded normals. *(C)* Semantic photorealistic style transfer examples generated using [42] (from 2 different style images). A super-resolution model [4] is also applied to enhance details. *(D)* A rendering system [41] allows users to compose a new scene by placing virtual objects. A framework for the automatic generation of stereoscopic environments [30] enables users to view the scene from different angles, creating an impression of depth and solidity.

d) Extending resolution of indoor spherical representations: In general, current CNN architectures used for generating signals from spherical images are limited to a maximum resolution of 1024×512. This falls significantly short of the native resolutions of RGB panoramas, which can exceed 4096×2048 for modern commodity 360° cameras. Consequently, this limitation poses a considerable obstacle for VR applications. For instance, a 90°-by-90○ view generated from a 1024×512 panorama would have a resolution of only 256×256. Super-resolution (SR) is a common technique to address this limitation, aiming to restore a high-resolution image from a single or sequence of low-resolution (LR) images. Current state-of-the-art solutions utilize data distribution learning with Generative Adversarial Network (GAN) and spectral normalization (SN) regularization (Real-ESRGAN [46]) or hybrid attention mechanisms together with image transformers (HAT [4]) to boost the number of input pixels for reconstruction. Recently, Deng et al. [7] proposed a method targeting spherical images, involving a latitude adaptive upscaling network (LAU-Net) that splits a spherical image into different latitude bands and hierarchically upscales them with different factors. In previous research [41], we benchmarked LAU-Net, Real-ESRNet, and HAT for usage with spherical indoor imagery, finding that HAT provides the best accuracy in terms of W-PSNR, while LAU-Net and Real-ESRNet offer a better compromise between accuracy and processing times. Regarding depth signals, recent techniques exploit features compression [24] and perspective depth estimation combined with sampling over spherical tessellation and registration methods [2], [22]. However, such these approaches appear to consume too much time for VR applications. In [41], we also benchmarked state-of-the-art image super-resolution methods [4], [7], [46] on a high-resolution indoor spherical imagery depth dataset (Replica 360 4K). For our pipeline, we use [46] for generating super-resolution for both RGB and depth signals, as it strikes a balance between accuracy and processing times, making it more suitable for VR-related applications.

e) Editing system: Editing systems such as [41], could significantly advance the capabilities of virtual staging by providing a robust, scalable, and interactive solution for transforming panoramic images into high-resolution 3D environments. Their contributions are vital for the development and deployment of immersive experiences in the metaverse, making it an indispensable tool for various XR applications. In this work, we utilized [41], which is a comprehensive system for transforming single 360° panoramic images into interactive, high-resolution 3D representations suitable for various Extended Reality (XR) applications. The framework integrates a series of advanced deep-learning models to process, edit, and render spherical images of indoor environments. The core components of the system include a novel architecture for geometric and semantic information extraction (c.f. Sec. III-0b), a super-resolution module for enhancing image resolution (c.f. Sec. III-0d), and an interactive editor for scene manipulation and exploration.

The methodology begins with the acquisition of a single panoramic image, which is processed to infer depth and semantic segmentation using gated and dilated convolutions. These inferred signals are then fed into a super-resolution module based on image transformers [4], significantly improving the resolution of both color and depth signals. The enhanced high-resolution data allows for the creation of detailed 3D models that can be explored and edited interactively. Users can perform various operations, such as virtual object insertion, scene refurnishing, and deferred shading, on the reconstructed indoor scene (c.f. Figs. 1 [C] and 2 [D]). The system supports rendering in multiple modalities, including point cloud, polygonal, and wireframe representations, making it versatile for different applications.

f) Automatic stereo generation: In this study, we utilized the method proposed in [29], [30] to automatically generate stereoscopic environments for metaverse applications. This method's capabilities for view synthesis from single panoramic images provide a robust foundation for creating immersive, interactive, and visually compelling environments, essential for advancing virtual staging in the metaverse and enhancing user experiences across various industries. The technique introduces a framework to create and explore immersive stereoscopic scenes using single panoramic images. The core methodology is a data-driven architecture designed for panoramic monocular depth estimation and view synthesis. The framework starts by inferring a fixed set of panoramic stereo pairs from a single panoramic image, which are then seamlessly fused to cover the entire viewing workspace when explored through VR headsets (See Fig. 1 [D]). This architecture utilizes a lightweight gated network for depth estimation and view synthesis, ensuring scalability and low latency. The depth map is estimated from the input panoramic image and used to generate new views by reprojecting and inpainting the scene. This approach maintains high visual detail and stereo consistency, achieved through a combination of photometric loss and GAN-based super-resolution techniques. The result is a set of precomputed omnidirectional stereo pairs that provide a seamless and photorealistic stereoscopic experience. The system is integrated into WebXR viewers, making it accessible on various VR headsets and demonstrating effective performance across different indoor scenes. For a more visual demonstration, access this link.

IV. LIDAR-BASED PIPELINE

For the sake of comparison with the image-based pipeline described in Scc. III, we also implemented an immersive virtual staging pipeline based on modern LIDAR scanning and integrated with the Spatial.io ecosystem. LiDAR technology has emerged as a powerful tool in the construction of building-scale 3D acquisition [47]. While this technology has existed for a long time and is well-established for indoor reconstruction [31], its recent integration in commodity multi-purpose devices is expanding its range of applications. In particular, its inclusion as a standard feature in Apple's flagship smartphones over the past four years highlights its increasing significance for non-professional users [16]. These phone-based LiDAR scanners excel in generating detailed oriented point cloud data of real-world environments with good precision [19], [39]. Coupled with surface reconstruction techniques (e.g., the well established Screened Poisson Surface Reconstruction [15]), they can create complete 3D models that can be further enhanced with additional 3D objects, creating immersive and interactive experiences. Collaborative platforms like Spatial.io [36] could leverage these LiDAR-scanned spaces to facilitate meaningful user interactions, a core aspect of the metaverse. Within Spatial.io, users can engage in activities such as virtual staging, where they populate scanned rooms with furniture and other objects, fostering a dynamic and collaborative virtual environment. Additionally, users can voice chat and interact across multiple platforms, including VR, web, and mobile, without the need to manually configure the user experience for different platforms.

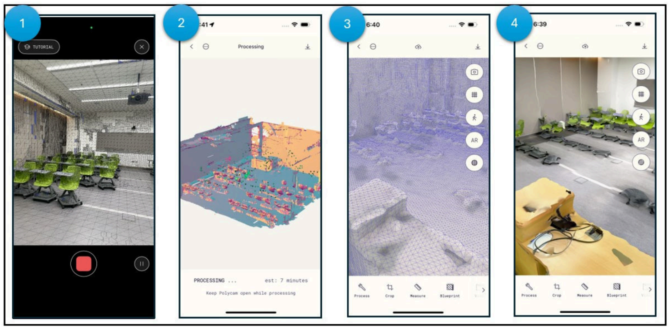

Fig. 3. **Polycam LiDAR Scanning:** This figure demonstrates the process and outcomes of using Polycam's LiDAR mode to scan a typical classroom. The images show the initial scanning phase and the resulting detailed 3D model. The process highlights the efficiency and precision of mobile LiDAR scanning for generating accurate and interactive 3D models of real-world environments, which can then be utilized within the metaverse for various applications, including virtual staging and collaborative design.

a) LIDAR scanning: LiDAR technology embedded in Apple's mobile devices is a successful example of integration of 3D sensing in mobile multi-purpose devices. Numerous mobile applications have leveraged Apple's LiDAR sensor for 3D scanning, each utilizing distinct algorithms and integrating data from other sensors such as the camera and gyroscope. Some applications, e.g. SiteScape, strive to produce dense and fairly accurate point cloud data, while others target fast acquisition or the automatic generation of low-poly processed 3D models, e.g., EveryPoint and 3D Scanner App [39]. While a number of solutions perform capture on the device and reconstruction on connected servers, Polycam leverages the device's built-in LiDAR sensor to directly capture depth information from the environment and process it on device, leading to a faster 3D model generation [32]. Using the LiDAR sensor in an iPhone, Polycam collects data points representing the room's surface. This data is then processed and refined by the app to automatically generate a complete 3D model of the room. The efficiency of capturing a room with Polycam's LiDAR mode depends on the room's size. For instance, a typical classroom of 84 square meters can be scanned and processed in approximately 10 to 15 minutes. (See Fig. 3)

b) Spatial Creator Toolkit: The Spatial Creator Toolkit, a free plugin for the widely used Unity game engine, facilitates the development of interactive, multiplayer experiences. By streamlining the creation process for social spaces, games, and collaborative environments within the metaverse, the toolkit offers a suite of built-in features. These include matchmaking functionalities, synchronized object and variable management, integrated voice and text chat options, and sandbox testing environments. By lowering the barrier to entry for metaverse development, it empowers a broader range of developers to contribute to the evolving metaverse landscape. Notably, creators using the toolkit do not have to manually address networking and user interaction across multiple platforms, including VR, web, and mobile. However, the platform does not fully support many of Unity's components, which limits the creation of advanced experiences. Despite this limitation, the tool effectively provides a simple way to share experiences across various platforms within the metaverse. For more advanced applications game engines could be utilized for building a custom implementation of the features provided in Spatial.io.

c) Metaverse exploration: Unity can be leveraged to reimagine a room scanned with LiDAR by importing the low poly 3D models generated from PolyCam into the Unity game engine. Once imported, it is possible to populate the virtual room with furniture and other objects to create an interactive and realistic environment. Through integration with Spatial.io, users can interact within the populated room, utilizing features such as voice and text chat to communicate and provide real-time feedback. This collaborative approach allows users to experience and modify the virtual environment, enhancing the design process and ensuring that the final result meets their needs and expectations. Moreover, it is possible to enable object placement through edge devices on multiple platforms, such as VR and mobile. This capability requires a set database of 3D objects and custom components that facilitate object placements. By combining Unity's powerful rendering capabilities with Spatial.io's interactive features, and leveraging edge devices for seamless object manipulation, developers can create immersive and engaging virtual spaces within the metaverse. This facilitates meaningful user interactions and feedback, advancing the possibilities for virtual staging and collaboration in the evolving metaverse landscape.(See Fig. 4)

Fig. 4. **Polycam classroom creation:** A demo showcasing a classroom environment created with Polycam, an iOS application that utilizes iPhone LiDAR sensors, allows users to interact with the environment, communicate via text or audio, or watch an educational video through Spatial.io services.(click the link to access the metaverse using Mobile,VR,or the Web).

V. QUALITATIVE COMPARISON

We performed a qualitative analysis of the various features of the presented pipelines for the generation of immersive indoor environments. Table I contains the main insights of this analysis. Based on the comparison, the AI-image-based pipeline is better suited for virtual staging applications. Virtual staging does not require users to navigate the entire scene to view the redecorated room, nor does it necessitate a detailed collision mesh for complex physics interactions. While these

Feature	**AI-Image-Based Pipeline**	**LiDAR-Based Pipeline**
Data Acquisition	Single panoramic image of the environment	LiDAR scan using a mobile device or professional scanner
Processing	Deep learning models for clutter removal, signal generation, style transfer, and super-resolution	Algorithms like screened poisson surface reconstruction to create 3D models from point cloud data
Output	A photo-realistic 360 image with semantic segmentation, depth information, and stylistic customization, ready for immersive stereoscopic virtual staging	Point cloud data reflecting real-world spatial relationships; intensive data processing and manual configuration required for high-quality complex models. Automatic generation yields low-poly 3D models, ready for virtual staging
Advantages	- Faster processing - Unique and stylized environments - Well-suited for indoor spaces	- High fidelity of real-world spatial details - Full physics integration for realistic simulations - Full scene exploration
Limitations	- Quality dependent on input image - Limited field of view	- Requires specialized equipment - Less detailed for large or outdoor spaces - Struggles with complex or cluttered environments - Does not support clutter removal - Struggles with transparent and reflective surfaces.
Realism	High (photo-realistic)	Low (automatic low-poly model generation)
Physics	Limited (using depth maps)	Full (using collision mesh derived from the 3D mesh)

TABLE I
QUALITATIVE COMPARISON OF AI-IMAGE-BASED AND LIDAR-BASED PIPELINES

attributes are valuable advantages of the LiDAR-based pipeline and may justify its use in other contexts, virtual staging prioritizes photo-realism and clutter removal, which the AI-image-based pipeline demonstrably excels at. In contrast, the LiDAR pipeline, although capable of automatic generation on edge devices [32], produces low-poly and unrealistic 3D models. Additionally, the LiDAR system not only lacks automatic clutter removal but also requires significant technical expertise to generate high-quality 3D models from point-cloud data [18]. These limitations further hinder its effectiveness for virtual staging applications. Therefore, the AI-image-based pipeline's emphasis on automatic clutter removal, photo-realistic stereoscopic experience, and accessibility makes it a more convenient and effective solution for virtual staging applications.

VI. PRACTICAL IMPLICATIONS

The integration of advanced AI and LIDAR-based technologies for virtual staging has transformative potential across several industries. Below, we discuss how these technologies can be effectively applied in real estate, furniture retail, interior design, the construction industry, remote collaboration, and immersive training.

a) Real estate: Virtual staging technologies, particularly those leveraging LIDAR and AI-based image processing, offer transformative capabilities for the real estate industry [5], [20], [44]. These technologies enable detailed interactive virtual tours and immersive experiences of properties [41]. Utilizing AI-based image processing, such as semantic style transfer [42] and super-resolution [4], real estate agents can create photorealistic and visually appealing representations of properties. This allows buyers to visualize different interior design options and spatial arrangements, enhancing their decision-making process. The ability to remove clutter [23] and present clean, staged environments also makes properties more attractive and marketable. Furthermore, using LIDAR scanning, accurate 3D models of properties can be created, allowing potential buyers to explore homes remotely as if they were physically present [41]. This not only enhances the buying experience but also broadens the market reach, as international buyers can tour properties without the need for travel. AI-based style transfer can further enhance these tours by enabling dynamic visualization of different interior styles and layouts, helping buyers to envision the potential of each property.

b) Furniture retail: In the furniture retail sector, virtual staging technologies can revolutionize the way products are showcased and sold [21], [41]. These technologies enable customers to visualize how different pieces of furniture will look in their homes before making a purchase, eliminating the need to visit the furniture showroom. Using AI-based systems for rendering [41] and super-resolution [4], retailers can create high-quality, interactive 3D models of furniture within various room settings. This not only improves the shopping experience by allowing customers to explore different styles and arrangements but also reduces return rates by providing a clearer expectation of how products will fit and look in their intended spaces. Moreover, by integrating AI-based image processing pipelines, retailers can create virtual showrooms where customers can visualize furniture in various settings and configurations. This interactive experience can help customers make more informed purchasing decisions by seeing how different pieces fit together and complement existing decor. Super-resolution models [4], [7], [46] and semantic style transfer [42] can enhance the realism of these virtual showrooms, making the virtual furniture appear as realistic as possible.

c) Interior design: Interior designers benefit from virtual staging technologies by being able to present clients with a range of design options without the need for physical samples [45], [48]. AI models for multitask dense prediction (See Sec. III-0b and semantic style transfer [42] allow designers to experiment with different colors, materials, and layouts in a virtual environment. This speeds up the design process and enhances client satisfaction by providing a clear and realistic preview of the final outcome. The ability to generate stereoscopic environments [30] facilitates a more immersive and engaging client presentation. AI-based image processing can generate high-fidelity, photorealistic visualizations of design proposals, allowing clients to see precisely how their spaces will look after redesign. The ability to quickly switch between different styles and layouts using virtual staging tools can facilitate better client communication and faster decision-making. Additionally, these technologies allow designers to experiment with various elements without the need for physical materials, saving time and resources.

d) Construction industry: The construction industry can leverage virtual staging technologies to improve project visualization and collaboration. LIDAR-based pipelines provide accurate 3D models of construction sites, which can be used for planning and monitoring progress [38]. AI-based tools for depth estimation and view synthesis [30] help create detailed virtual environments that reflect the current state of a project. These technologies enable stakeholders to conduct virtual walkthroughs, identify potential issues early, and make informed decisions, thus enhancing efficiency and reducing costs. Moreover, these technologies can be used to visualize building projects before completion, ensuring they are constructed as designed. Building owners can leverage these technologies to showcase how unfinished buildings will appear once completed, enabling them to market and sell properties before construction is finalized. This capability also improves project coordination and stakeholder communication, ultimately leading to more efficient project execution.

e) Remote collaboration: The integration of virtual staging technologies into remote collaboration platforms can significantly enhance the way teams work together. High-resolution [4], interactive virtual environments [30], [41] can facilitate better communication and collaboration among team members who are geographically dispersed [38]. For example, architects and engineers can collaboratively review and edit virtual models of their projects in real-time, making adjustments and discussing changes as if they were in the same room, leading to more cohesive teamwork and faster project timelines.

f) Immersive training: In the context of immersive training, virtual staging technologies offer a safe and controlled environment for training simulations. AI-driven models for semantic segmentation, depth estimation, and interactive rendering III-0e create realistic scenarios that can be used for training purposes in fields such as healthcare [40], emergency response, and military operations [3]. The ability to create detailed and interactive virtual environments ensures that trainees can practice and hone their skills in a lifelike setting, improving the overall effectiveness of the training programs.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a comprehensive exploration of virtual staging technologies within the context of the metaverse, focusing on two primary pipelines: an AI-based image processing pipeline, and a LIDAR-based pipeline. Our qualitative comparison highlighted the strengths and limitations of each approach, demonstrating their potential to revolutionize various industries, including real estate, interior design, and immersive training. The LIDAR-based pipeline excels in generating highly accurate 3D models of indoor environments, making it ideal for applications requiring detailed spatial data and full physics integration. Its ability to provide real-time, interactive editing and collaboration in virtual spaces showcases its suitability for immersive design and construction monitoring. However, the need for specialized equipment and the challenges associated with processing complex environments limit its broader application. On the other hand, the AI-based image processing pipeline offers significant advantages in terms of flexibility, efficiency, and ease of use. By leveraging advanced AI algorithms for tasks such as clutter removal, semantic style transfer, and super-resolution, this pipeline can rapidly generate high-quality, photorealistic virtual environments from single panoramic images. Its applicability in virtual staging, where photo-realism and rapid deployment are critical, underscores its potential for transforming the real estate and furniture retail sectors. Future research should focus on conducting user studies to evaluate immersive experiences, integrating the strengths of both LIDAR-based and AI-based pipelines for enhanced versatility, and developing tailored metaverse solutions for specific industries. Additionally, improving real-time collaboration features and optimizing technologies for mobile and edge devices will enhance accessibility and usability, driving broader adoption and impact of virtual staging technologies across various sectors. Furthermore, as a next step, we also plan to extend our evaluation by incorporating additional virtual scenes, conducting comprehensive user testing, and exploring more diverse datasets and virtual environments to further assess the applicability, robustness, and user experience of the proposed technology.

ACKNOWLEDGMENTS

This publication was supported by NPRP-S 14th Cycle grant 0403-210132 AIN2 from the Qatar National Research Fund (a member of Qatar Foundation). The findings herein reflect the work and are solely the responsibility, of the authors.

REFERENCES

[1] W. AlBalkhy, D. Karmaoui, L. Ducoulombier, Z. Lafhaj, and T. Linner. Digital twins in the built environment: Definition, applications, and challenges. *Automation in Construction*, 162:105368, 2024. 2

[2] M. R. Area, M. Yuan, and C. Richardt. 360monodepth: High-resolution 360° monocular depth estimation. In *Proc. CVPR*, 2022. 4

[3] J. Baughman. Enter the battleverse: China's metaverse war. *Military Cyber Affairs*, 5(1):2, 2022. 7

[4] X. Chen, X. Wang, J. Zhou, and C. Dong. Activating more pixels in image super-resolution transformer. *arXiv preprint arXiv:2205.04437*, 2022. 3, 4, 6, 7

[5] P. Chhikara, H. Kuhar, A. Goyal, and C. Sharma. Digitour: Automatic digital tours for real-estate properties. In *Proc. 6th Joint International Conference on Data Science & Management of Data (10th ACM IKDD CODS and 28th COMAD)*, pages 223–227, 2023. 2, 6

[6] T. L. da Silveira, P. G. Pinto, J. Murrugarra-Llerena, and C. R. Jung. 3d scene geometry estimation from 360 imagery: A survey. *ACM Computing Surveys*, 55(4):1–39, 2022. 2

[7] X. Deng, H. Wang, M. Xu, Y. Guo, Y. Song, and L. Yang. Lau-net: Latitude adaptive upscaling network for omnidirectional image super-resolution. In *Proc. CVPR*, pages 9189–9198, 6 2021. 4, 6

[8] V. Drobnyi, Z. Hu, Y. Fathy, and I. Brilakis. Construction and maintenance of building geometric digital twins: state of the art review. *Sensors*, 23(9):4382, 2023. 2

[9] L. Gao, J.-M. Sun, K. Mo, Y.-K. Lai, L. J. Guibas, and J. Yang. Scenehgn: Hierarchical graph networks for 3d indoor scene generation with fine-grained geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 2

[10] S. Gao, K. Yang, H. Shi, K. Wang, and J. Bai. Review on panoramic imaging and its applications in scene understanding. *IEEE Transactions on Instrumentation and Measurement*, 71:1–34, 2022. 2

[11] A. Gradinaru, M. Vrejoiu, F. Moldoveanu, and M.-A. Balutoiu. Evaluating the usage of text to3d scene generation methods in game-based learning. In *Proc. CSCS*, pages 633–640. IEEE, 2023. 2

[12] E. Hopkins. Virtual commerce in a decentralized blockchain-based metaverse: Immersive technologies, computer vision algorithms, and retail business analytics. *Linguistic and Philosophical Investigations*, 21:203–218, 2022. 1

[13] G. Ji, A. O. Sawyer, and S. G. Narasimhan. Virtual home staging: Inverse rendering and editing an indoor panorama under natural illumination. In *International Symposium on Visual Computing*, pages 329–342. Springer, 2023. 2

[14] Z. Jiang, Z. Xiang, J. Xu, and M. Zhao. Lgt-net: Indoor panoramic room layout estimation with geometry-aware transformer network. In *Proc. CVPR*, pages 1654–1663, 2022. 2

[15] M. Kazhdan and H. Hoppe. Screened poisson surface reconstruction. *ACM Trans. Graph.*, 32(3), jul 2013. 4

[16] S. Kottner, M. J. Thali, and D. Gascho. Using the iphone's lidar technology to capture 3d forensic data at crime and crash scenes. *Forensic Imaging*, 32:200535, 2023. 4

[17] Z. Li, L. Wang, X. Huang, C. Pan, and J. Yang. Phyir: Physics-based inverse rendering for panoramic indoor images. In *Proc. CVPR*, pages 12713–12723, 2022. 2

[18] Q. Liu, S. Xu, J. Xiao, and Y. Wang. Sharp feature-preserving 3d mesh reconstruction from point clouds based on primitive detection. *Remote sensing*, 15(12), 2023. 6

[19] G. Luetzenburg, A. Kroon, and A. A. Bjørk. Evaluation of the apple iphone 12 pro lidar for an application in geosciences. *Scientific Reports*, 11:22221, 123. 4

[20] K. G. Nalbant and Ş. Uyanık. Computer vision in the metaverse. *Journal of Metaverse*, 1(1):9–12, 2021. 6

[21] H. Oh, S.-Y. Yoon, and C.-R. Shyu. How can virtual reality reshape furniture retailing? *Clothing and Textiles Research Journal*, 26(2):143–163, 2008. 6

[22] C.-H. Peng and J. Zhang. High-resolution depth estimation for 360° panoramas through perspective and panoramic depth images registration. In *Proc. WACV*, pages 3115–3124, 2023. 4

[23] G. Pintore, M. Agus, E. Almansa, and E. Gobbetti. Instant automatic emptying of panoramic indoor scenes. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3629–3639, 2022. 1, 2, 6

[24] G. Pintore, M. Agus, E. Almansa, J. Schneider, and E. Gobbetti. SliceNet: deep dense depth estimation from a single indoor panorama using a slice-based representation. In *Proc. CVPR*, pages 11536–11545, 2021. 4

[25] G. Pintore, M. Agus, and E. Gobbetti. Atlantanet: inferring the 3d indoor layout from a single 360° image beyond the manhattan world assumption. In *Proc. ECCV*, pages 432–448. Springer, 2020. 2

[26] G. Pintore, E. Almansa, M. Agus, and E. Gobbetti. Deep3DLayout: 3d reconstruction of an indoor layout from a spherical panoramic image. *ACM Transactions on Graphics (TOG)*, 40(6):1–12, 2021. 2

[27] G. Pintore, E. Almansa, A. Sanchez, G. Vassena, and E. Gobbetti. Deep panoramic depth prediction and completion for indoor scenes. *Computational Visual Media*, pages 1–20, 2024. 2

[28] G. Pintore, F. Ganovelli, E. Gobbetti, and R. Scopigno. Mobile mapping and visualization of indoor structures to simplify scene understanding and location awareness. In *Proc. ECCV Workshops*, pages 130–145. Springer, October 2016. 2

[29] G. Pintore, A. Jaspe-Villanueva, M. Hadwiger, E. Gobbetti, J. Schneider, and M. Agus. Panoverse: automatic generation of stereoscopic environments from single indoor panoramic images for metaverse applications. In *Proc. ACM Conference on 3D Web Technology*, pages 1–10, 2023. 2, 4

[30] G. Pintore, A. Jaspe-Villanueva, M. Hadwiger, J. Schneider, M. Agus, F. Marton, F. Bettio, and E. Gobbetti. Deep synthesis and exploration of omnidirectional stereoscopic environments from a single surround-view panoramic image. *Computers & Graphics*, 119:103907, March 2024. 1, 2, 3, 4, 7

[31] G. Pintore, C. Mura, F. Ganovelli, L. Fuentes-Perez, R. Pajarola, and E. Gobbetti. State-of-the-art in automatic 3d reconstruction of structured indoor environments. *Computer Graphics Forum*, 39(2):667–699, 2020. 2, 4

[32] I. Polycam. Polycam, 2024. LiDAR scanning iOS application. 5, 6

[33] M. Scorpio, R. Laffi, A. Teimoorzadeh, and S. Sibilio. Immersive virtual reality as a tool for lighting design: applications and opportunities. *Journal of Physics: Conference Series*, 2042(1):012125, 2021. 2

[34] U. Shah, M. Tukur, M. Alzubaidi, G. Pintore, E. Gobbetti, M. Househ, J. Schneider, and M. Agus. MultiPanoWise: holistic deep architecture for multi-task dense prediction from a single panoramic image. In *Proc. OmniCV (CVPRW)*, 2024. 2, 3

[35] Z. Shen, C. Lin, K. Liao, L. Nie, Z. Zheng, and Y. Zhao. PanoFormer: Panorama transformer for indoor 360° depth estimation. In *Proc. ECCV*, pages 195–211. Springer, 2022. 2

[36] Spatial.io. Introducing: The Spatial.io Creator Toolkit, Powered by Unity (Beta), December 19, 2022. 4

[37] L. O. K. Tchomdji, S.-j. Park, and R. Kim. Developing virtual tour content for the inside and outside of a building using drones and matterport. *International Journal of Contents*, 18(3):74–84, 2022. 2

[38] T.-A. Teo and C.-C. Yang. Evaluating the accuracy and quality of an ipad pro's built-in lidar for 3d indoor mapping. *Developments in the Built Environment*, 14:100169, 2023. 1, 7

[39] L. Teppati Losè, A. Spreafico, F. Chiabrando, and F. G. Tonolo. Apple lidar sensor for 3d surveying: Tests and results in the cultural heritage domain. *Remote Sensing*, 14(17):4157, 2022. 4, 5

[40] J. Thomason. Metahealth-how will the metaverse change health care? *Journal of Metaverse*, 1(1):13–16, 2021. 7

[41] M. Tukur, G. Pintore, E. Gobbetti, J. Schneider, and M. Agus. Spider: A framework for processing, editing and presenting immersive high-resolution spherical indoor scenes. *Graphical Models*, 128:101182, 2023. 1, 2, 3, 4, 6, 7

[42] M. Tukur, A. U. Rehman, G. Pintore, E. Gobbetti, J. Schneider, and M. Agus. Panostyle: Semantic, geometry-aware and shading independent photorealistic style transfer for indoor panoramic scenes. In *Proc. CVPR*, pages 1553–1564, 2023. 1, 2, 3, 6

[43] M. Tukur, J. Schneider, M. Househ, A. H. Dokoro, U. I. Ismail, M. Dawaki, and M. Agus. The metaverse digital environments: a scoping review of the challenges, privacy and security issues. *Frontiers in big Data*, 6, 2023. 1

[44] M. Tukur, J. Schneider, M. Househ, A. H. Dokoro, U. I. Ismail, M. Dawaki, and M. Agus. The metaverse digital environments: A scoping review of the techniques, technologies, and applications. *Journal of King Saud University-Computer and Information Sciences*, page 101967, 2024. 1, 6

[45] C. Vazquez, N. Tan, and S. Sadalgi. Home studio: a mixed reality staging tool for interior design. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–5, 2021. 2, 6

[46] X. Wang, L. Xie, C. Dong, and Y. Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proc. ICCV Workshops*, 2021. 1, 3, 4, 6

[47] K. Yadav, R. Ramrakhya, S. K. Ramakrishnan, T. Gervet, J. Turner, A. Gokaslan, N. Maestre, A. X. Chang, D. Batra, M. Savva, et al. Habitat-matterport 3d semantics dataset. In *Proc. CVPR*, pages 4927–4936, 2023. 2, 4

[48] Y. Zhang, H. Liu, M. Zhao, and M. Al-Hussein. User-centered interior finishing material selection: An immersive virtual reality-based interactive approach. *Automation in Construction*, 106:102884, 2019. 2, 6

[49] J. Zheng, J. Zhang, J. Li, R. Tang, S. Gao, and Z. Zhou. Structured3d: A large photo-realistic dataset for structured 3d modeling. In *Proc. ICCV*, pages 519–535. Springer, 2020. 3