STAG: Smart Tools and Applications in Graphics (2022)
D. Cabiddu, T. Schneider, G. Cherchi, and R. Scateni (Editors)

SPIDER: SPherical Indoor DEpth Renderer

M. Tukur[1], G. Pintore[2], E. Gobbetti[2], J. Schneider[1], and M. Agus[1]

[1] College of Science and Engineering, HBKU, Qatar
[2] Visual and Data-intensive Computing, CRS4, Italy

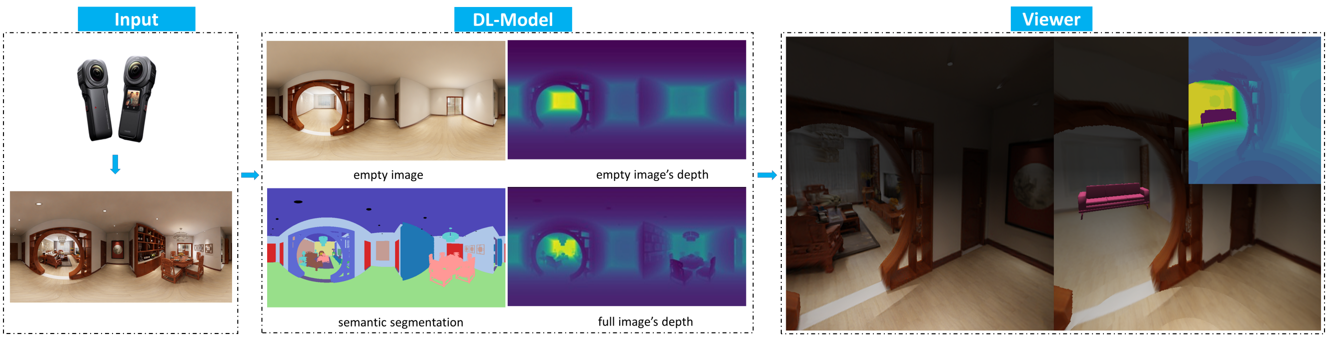

Figure 1: *SPIDER: Our SPherical Indoor Depth Renderer framework consists of three modules: a single RGB panoramic image of an indoor environment snapped using 360° camera (left) is fed into various deep learning models for abstracting depth signal and semantic content, and removing all clutter (middle). The viewer allows users to perform interactive exploration and basic editing operations on the reconstructed indoor scene, including: virtual object insertion and refurnishing (transferring portions of rooms), and lighting effects on various graphical representations, such as triangle meshes and point clouds. These rendering operations can be utilized for several virtual staging applications.*

Abstract

*Today's Extended Reality (XR) applications that call for specific Diminished Reality (DR) strategies to hide specific classes of objects are increasingly using 360° cameras, which can capture entire areas in a single picture. In this work, we present an interactive-based image editing and rendering system named **SPIDER**, that takes a spherical 360° indoor scene as input. The system incorporates the output of deep learning models to abstract the segmentation and depth images of full and empty rooms to allow users to perform interactive exploration and basic editing operations on the reconstructed indoor scene, namely: i) rendering of the scene in various modalities (point cloud, polygonal, wireframe) ii) refurnishing (transferring portions of rooms) iii) deferred shading through the usage of precomputed normal maps. These kinds of scene editing and manipulations can be used for assessing the inference from deep learning models and enable several Mixed Reality (XR) applications in areas such as furniture retails, interior designs, and real estates. Moreover, it can also be useful in data augmentation, arts, designs, and paintings.*

CCS Concepts

• ***Applied computing*** → ***Architecture (buildings);*** • ***Computing methodologies*** → ***Image-based rendering;*** *Reconstruction;*

1. Introduction

Recently, 360° cameras or omnidirectional cameras, that are capable to capture full environments in a single shot, became extremely popular for a variety of applications, ranging from egocentric videos [JMK*22] to virtual reality applications [PD21]. Concurrently with the diffusion of these imaging devices, the explosion of AI-based computer vision technologies have enabled a variety of methods for automatically extracting information about the surrounding environment, with a variety of applications, ranging from object tracking for security to autonomous driving [YZR*21]. Between the various application domains that obtained great benefit from deep learning technologies, one of the most popular is the architectural design of indoor environments, both for entertainment industry (3D games and animation) and for real estate and construc-

DOI: 10.2312/stag.20221267

tion industry. Realistic 3D indoor scenes are in high demand due to the quick development of VR/AR technology and new applications [Ma17; MLZ*16]. However creating and manipulating complex 360° indoor settings takes a lot of time and demands strong 3D modeling and designing abilities [VTS21].

Furthermore, when looking for a potential home or apartment to rent or buy, the spaces are usually packed with furniture and other clutter, making it challenging to envision how it may appear empty or, preferably, with our own stuff in place. Moreover, before buying a new sofa for even our own occupied apartment, we want to see how it would appear in our lounge. Additionally, with the present pandemic-related limitations and many people working from home, house sales are increasing significantly. Remote house shopping is becoming widely attractive, and efficient solutions to support virtual home tours are desperately needed. One such solution is virtual staging: how would furniture fit in a home and what would it look like if specific settings (e.g., object size, and location, sun orientation, flooring) were changed? Panoramas are increasingly being utilized to display properties in order to give good imagery for staging. Panoramas give surround information, but approaches created for images with a narrow field of view cannot be immediately applied [ZCB*22].

In this paper, we present a real time renderer for spherical images representing indoor environments, that we dub SPIDER. The system exploits recent methods for instant removal of clutter from a scene [PAAG22], together with the output of deep learning models to abstract the semantic content [SSC21] and depth information for full [PAA*21] and empty rooms [PAAG22]. The system enables the insertion of virtual objects into a single panoramic photo of an empty room. The objects are inserted as a tessellated meshes that can be interactively moved, scaled, and lit according to simple illumination models and environment mapping. The system also allow users to perform other interactive exploration and basic editing operations on the reconstructed indoor scene, namely: i) rendering of the scene in various modalities (point cloud, polygonal, wireframe) ii) refurnishing (transferring portions of rooms) iii) deferred shading through the usage of normal maps. These kinds of scene editing and manipulations enables several XR applications in areas such as furniture retails, interior designs, and real estates.

In summary, our system provides the following advantages:

- **Interactivity:** we provide an interactive system that allow users to explore and edit 360° indoor scenes in a flexible way. Our system provides an interactive and dynamic 3D representation of the scene, either in form of polygonal representation of the environment through spherical tessellation, or in form of a sparse point cloud.
- **Simplicity:** Our proposed interactive system is simple to the extent that it allows even a non-expert users to intuitively edit indoor layout and greatly reduces the design time cost. Almost all the operations can be achieved by click and drag operations.
- **360° scene manipulation:** contrary with previous frameworks that work mostly on perspective images and geometric representations, our systems uses only 360° images representing indoor scenes and automatically inferred metadata.

To our knowledge, none of the existing systems combine the above mentioned contributions and editing operations on spherical 360° indoor scenes.

2. Related work

This work deals with processing, editing, rendering and interactive exploration of 3D scenes reconstructed from panoramic 360 degree images representing indoor environments. We do not aim to provide here a full review of these topics: we refer readers to recent surveys about deep learning methods for VR content creation [WLLZ20], 3D reconstruction of indoor environments in general [PMG*20; KYYC20], and 3D scene reconstruction from 360 imagery [dSPMJ22]. In the following we will discuss the methods that most closely relate to our framework.

Reconstruction of 3D indoor scenes from single panoramic images Since the explosion in the consumer market of omnidirectional (360 degree cameras), that are able to provide a field of view that covers approximately the entire sphere by composing multiple fish-eye lenses, various computer vision applications have been developed, ranging from immersive videos representing sport actions [FLPH19] to virtual tours for the real estate market [SAB*20]. This imaging technology have been rapidly become very popular for acquisition of indoor environments, because one single image acquired from the center of one room is able in most cases to provide an adequate representation of the scene and it can be used for various applications. In last decade, various groups investigated automatic methods for extracting 3D information from single panoramic image, using various technologies from geometric reasoning up to convolutional neural networks. Pintore et al. [PGG*16] recover 2.5D layout of rooms exhibiting Manhattan world properties from omnidirectional images by using a specialized spatial transform based on catadioptric theory to define a parametric model for a global optimization problem. The same authors extended this technology to reconstruct room 3D layouts on less restrictive Atlanta world scenes [PAG20] through the usage of Recurrent Neural Networks (RNNs) and a customized training strategy based on domain-specific knowledge, and full triangulated 3D meshes representing the room layout without any constraint through Graph Convolutional Networks [PAAG21]. As long as public large datasets containing annotated panoramic images representing indoor environments have been made available to the computer vision community, like Matterport3D [CDF*17] and Structured3D [ZZL*20], various automatic technologies exploiting deep learning have been developed for inferring additional information about the indoor scenes, like depth signals [PAA*21; JSI*22; ZLW*22; YLR22], or semantic labelling for detection of objects and features [SSC21; ZKG20]. Very recently, deep learning technologies have targeted the extraction of geometric information for improving mixed reality applications: for example, Pintore et al. [PAAG22] developed a method for automatic emptying indoor scenes that can be used for interactive editing, while Li et al. [LWH*22] investigated the problem of physics-based inverse lighting for panoramic indoor scenes, and their system is able to insert virtual objects in the indoor scene with realistic lighting effects. Our framework integrates the outputs of various 3D reconstruction models, like depth inference [PAA*21], semantic segmentation [SSC21] and automatic room emptying [PAAG22], for providing an interactive indoor scene editing and exploration system.

Interactive exploration/editing of indoor scenes Recently, various frameworks have been proposed for interactive exploration and modelling of man-made scenes [ZAQW20; TCC*21; MLZ*16; ZCC16; PLWZ19; CK17; IZZE17; QCJK18; WLZ*18; VTS21]. However, some of them are applied only to outdoor images [ZAQW20; PLWZ19; CK17; LLWL20], while others do not support interactivity [TCC*21; QCJK18], or are limited to perspective images [IZZE17]. Recently, Vazquez et al. [VTS21] presented Home Studio, a web-based tool that integrates Matterport scan and generate photorealistic renders of products in the user's context. On the other side, our framework works directly on single panoramic images and inferred pixel-to-pixel representations. It also integrates the outputs of various 3D reconstruction models, like depth inference [PAA*21], semantic segmentation [SSC21] and automatic room emptying [PAAG22], for providing an interactive indoor scene editing and exploration system. In addition, our system allows the insertion and manipulation of virtual objects. Very recently, Zi et al [ZCB*22] used semantic segmentation for appearance decomposition in a way to enable various photorealistic operations, like virtual furniture insertion, floor material replacement. Since their technology is based on panoramic images, their method can be incorporated in our framework.

3. Methods

Overview SPIDER is a deep learning/OpenGL-based graphical application tool that allows users to edit and manipulate 360° indoor scenes. As shown in Figure 1, the tool takes as input, a single spherical indoor scene snapped using a 360° camera (Figure 1-Left). The input RGB image is then used for automatic inference of additional panoramic images (Figure 1-Middle) to support the rendering operations. The images are generated through various deep learning models: [PAA*21] is used to generate the depth for the original image, while [PAAG22] is utilized to automatically generate the emptied panoramic indoor scene in form of RGB and depth image, and finally [SSC21] is incorporated to our framework for generating semantic segmentation of the panoramic image.

The renderer (Figure 1-Right) is implemented in C++ by using Qt and OpenGL libraries inside QtCreator environment. The shaders are written on OpenGL Shading Language (GLSL) and the code is written in the way to ensure full portability with mobile devices (OpenGL ES 3.0) . Several editing operations can be performed using the proposed renderer.

Processing pipeline Our system incorporates the output of various deep learning models for the inference of the following information from single RGB panoramic images:

- **Depth inference:** we use the architecture proposed in SliceNet [PAA*21], that exploits the role of gravity in man-made indoor scenes, for deriving a compact representation of the scene into vertical slices of the sphere. Long- and short-term relationships among slices are then incorporated in a residual Convolutional Neural Network to recover the equirectangular depth map. The output of the model is a 16-bit single channel image containing the depth signal at millimetric resolution in which every pixel is representing the distance between the scene object and the camera.
- **Clutter removal:** we use the method proposed by Pintore et al. [PAAG22] for instant automatic emptying of indoor scenes. The approach computes an attention mask of the clutter in the image based on the geometric difference between full and empty scenes, and propagates it through gated convolutions that drive the generation of the output image and its depth. For obtaining that, it exploits, during supervised training, geometric losses of different orders, including robust pixel-wise geometric losses and high-order 3D constraints typical of indoor structures.
- **Semantic content:** we apply the method proposed in HoHoNet [SSC21] for deriving a labelled representation of the indoor scene, able to separate floor, ceiling, walls and the various furniture items.

Additional signals can be automatically derived from a single panoramic image, like various light contributions [ZCB*22], normal maps, or global effects. Currently, we compute normal maps though differentiation of the depth map, and we use it for creating simple lighting effects through diffuse Lambert model. Fig. 2 shows an example of a portion of a scene with texture mapped the original RGB signal (left), the extracted semantic content (middle), and the computed normal map (right).

Rendering pipeline The renderer takes as input the original RGB equirectangular image together with the inferred images (depth, normal, and semantic content) to construct an interactive 3D representation of the indoor environment, with and without clutter. For doing that, the pipeline is composed by the following steps, to be applied to both the cluttered and empty environment:

- **Geometry generation:** we start from a subdivision sphere, that we use for texturing the panoramic images. We obtain it as iterative subdivision of an icosahedron through midpoint edge splitting. For each subdivision level and for each triangle the scheme creates four new triangles (see Fig. 3). For all results generated in this paper, we considered $L = 8$ as subdivision level, in a way to obtain a tessellated sphere containing $1.3M$ vertices and $1.3M$ triangles, that is comparable to the pixel resolution of input panoramic images ($1024X768$).
- **Depth application:** the depth signal is applied to the spherical dome through a geometry shader that, for each point in the sphere, fetches the corresponding depth from the depth image and computes the world coordinates of the scene point through the ray tracing from the camera position. The recomputed scene can be visualized as rasterized polygons (filled or wireframed) or as a sparse point cloud (see Fig. 4 for some examples).
- **Lighting:** for lighting the reconstructed scene and the inserted virtual objects, we perform deferred shading in the fragment shader, by exploiting the information contained in the normal maps. We apply a simple Lambertian diffuse model with a single point light for the spherical scene, and an additional environment map for virtual inserted objects (see Fig.8 for some examples).
- **Scene modification:** we can perform insertion of virtual objects, as triangular meshes, and we can transfer portion of scenes between the full environment and the empty environment, by ex-

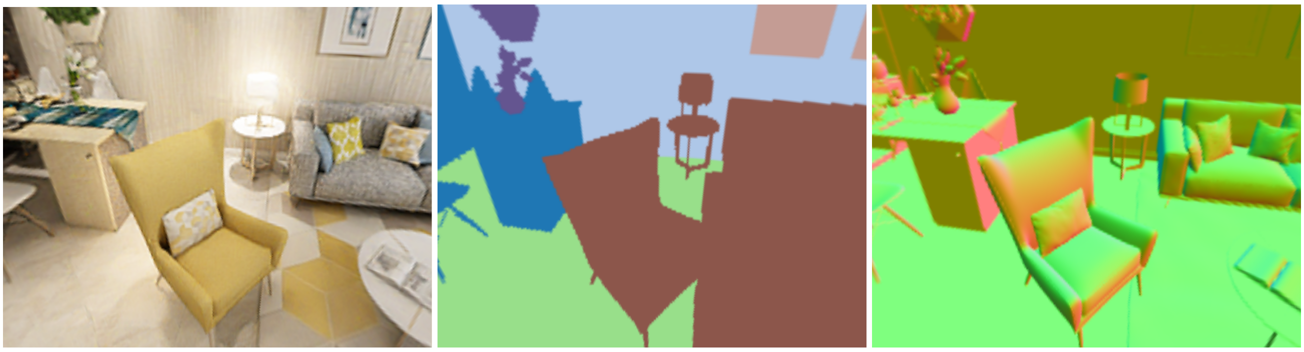

Figure 2: *Signals used in SPIDER: original RGB image(left), semantic segmentation content(middle), normal map(right).*

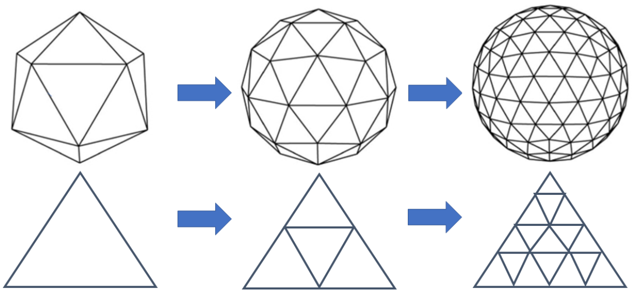

Figure 3: *The panoramic image is texture mapped on a sphere generated through iterative subdivision of an icosahedron (top). At each iteration, triangles are subdivided through midpoint edge splitting (bottom).*

ploiting the semantic map through picking operation. The virtual objects can be modified inside the scene (panning, scaling, rotating) (see Fig. 8).

User interface The current viewer interface is based on simple click and point operations (see Fig. 5). We provide two side views of the indoor scene: on the left the original cluttered scene, and on the right the empty scene, ready to be filled with novel 3D content or with portions of the original scene. For better visual understanding of the scene, we also provide an additional inset containing a view of the scene from the top. Users can transfer semantic portions of the scene through mouse picking operations, and can move the virtual objects through mouse panning and rotations. The operations for changing visual representations are instead attached to keyboard inputs, as well as the operations for creating exploration animations.

4. Applications

The SPIDER prototype can be used for interactive exploration of indoor scenes with a variety of applications.

Assessment of deep learning models The proposed system can be used for checking accuracy, qualitative results, and individuating artifacts, for example in the generation of depth signal. In Fig. 6 we show an error related to depth inference due to refractive properties of the window glass(left), and an example of the limitations in the generation of semantic content of the model proposed in [SSC21].

Figure 4: *Examples of point cloud rendering from various positions and different point size.*

We used extensively the system for qualitative assessment of our recent deep learning framework for automatic emptying indoor environments [PAAG22], especially for generating rendering results. Fig. 7 visualizes examples of real-time emptying of two different panoramic indoor scenes. We will use in the future for testing models for automatic segmentation, computation of normal maps, and light signal separation [ZCB*22].

Virtual staging Using 3D virtual staging, real estate experts can decorate and fill a vacant property without ever moving a muscle (or spending thousands on staging). It begins with the representation of empty rooms, which are then transformed into

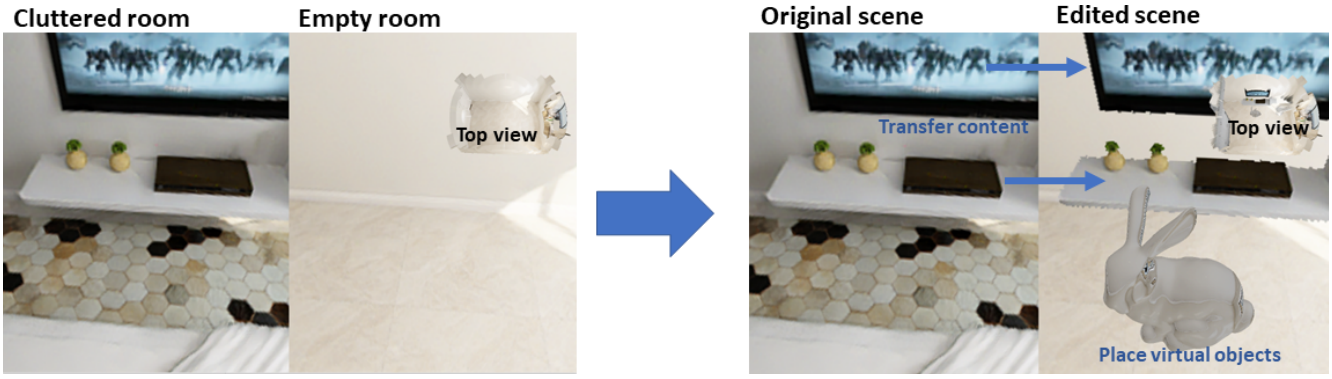

Figure 5: *The user interface of our prototype renderer contains two views: the cluttered scene (left) and the empty room (right). Users can compose a new scene by placing new virtual objects or by transferring content from the full scene and empty scene through picking operations.*

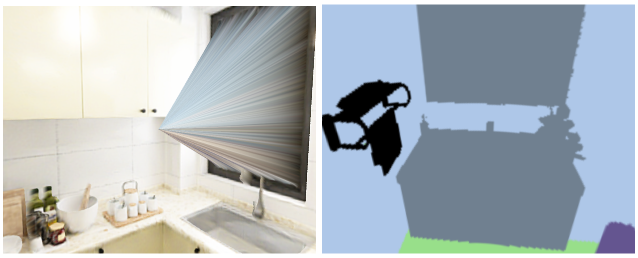

Figure 6: *Inference assessment: SPIDER can be used for finding artifacts during the inference of depth signal (left) or semantic content (right).*

lived-in spaces through virtual furnishing. Styldod [Sty22], Stuccco [Stu22], and PadStyler [Pad22] are a few firms that offer virtual staging services. Users may upload images of their rooms, and the firm will furnish and render them for them. However, the majority of them require significant manual modelling efforts. By contrast, our viewer enables the insertion of virtual objects as tessellated item that can be interactively moved and scaled. Moreover, simple picking operations allow users to transfer semantic content from the full room to the empty room. Fig. 8 shows some examples of insertion of virtual objects in the indoor empty scene, as well as transfer of scene portions from the full original scene to empty one.

Immersive exploration of indoor scenes The viewer is using Qt for Android and can be easily deployed to mobile devices, to be used for stereoscopic visualization of indoor scenes (see Figure 9). The mobile tool exploits IMU sensors and can be used together with VR headsets like Google cardboard for providing immersive exploration of the original scenes, the uncluttered ones and the refurnished ones. In this prototype application, users can explore the stereoscopic edited scene by using two kind of interfaces: the standard touch interface for implementing arcball motion, or the inertial sensors, through the Qt QRotationReading mechanism, that provides the orientation of the device in three-dimensional space. In the future, we will extend to work on HTC Vive for full immersive experiences, also for what concerns the editing process, while the mobile application will be used only for deploying edited scenes for virtual exploration.

5. Conclusions and future work

In this work, we presented a novel OpenGL/deep learning based tool that enables editing and manipulating 360° panoramic indoor scenes. Relying on some deep learning models to abstract the segmentation, empty, and depth images, our method allows several multiple virtual staging tasks comprising object insertion/movement, automatic room emptying, refurnishing (transferring portions of rooms), rendering of the scene in various modalities (point cloud, polygonal, wireframe), and deferred shading through the usage of normal maps. We believe that these can be used for assessing the inference from deep learning models, and enable several XR applications in areas such as furniture retails, interior designs, and real estates. Furthermore, it can also be useful in data augmentation, arts, designs, and paintings. Despite the promising results, our prototype system contains various limitations, that we plan to address in future work:

- **style transfer and material editing:** we plan to investigate methods for automatic transfer of styles between indoor environments, in a way to allow semiautomatic change of appearance of floor, ceiling and furnitures [HJN22].
- **illumination effects and materials:** currently we are using simple shading effects based on environment maps, resulting in lack of shadows for what concerns the insertion of virtual objects. We plan to investigate more complex illumination effects [MRNK21], like global occlusion, automatic models for the inference of light contributions [ZCB*22], and procedural models for material appearance [GHS*22].
- **immersive virtual staging:** we plan to extend the system for the creation of immersive virtual staging applications using Head Mounted Displays, that would enable the fast modelling of indoor scene through the integration of natural interfaces, geometry processing components, deep learning models for automatic scene understanding and data-driven guided exploration of scene and material databases.

Figure 7: *Interactive exploration of automatically emptied rooms. Various examples processed through the method in [PAAG22]*

Acknowledgments Authors GP and EG acknowledge the contribution of Sardinian Regional Authorities for project SVDC, and of the European Union's H2020 research and innovation program under grant 813170 (EVOCATION). We thank Eva Almansa (ESR for project EVOCATION) for the support on developing indoor tools for depth inference and clutter removal.

References

[CDF*17] CHANG, ANGEL, DAI, ANGELA, FUNKHOUSER, THOMAS, et al. "Matterport3D: Learning from RGB-D Data in Indoor Environments". *Proc. 3DV*. 2017 2.

[CK17] CHEN, QIFENG and KOLTUN, VLADLEN. "Photographic image synthesis with cascaded refinement networks". *Proc. ICCV*. 2017, 1511–1520 3.

[dSPMJ22] Da SILVEIRA, THIAGO LT, PINTO, PAULO GL, MURRUGARRA-LLERENA, JEFFRI, and JUNG, CLÁUDIO R. "3D scene geometry estimation from 360 imagery: A survey". *ACM Computing Surveys* (2022) 2.

[FLPH19] FAN, CHING-LING, LO, WEN-CHIH, PAI, YU-TUNG, and HSU, CHENG-HSIN. "A survey on 360 video streaming: Acquisition, transmission, and display". *ACM Computing Surveys* 52.4 (2019), 1–36 2.

[GHS*22] GUERRERO, PAUL, HAŠAN, MILOŠ, SUNKAVALLI, KALYAN, MĚCH, RADOMÍR, BOUBEKEUR, TAMY, and MITRA, NILOY J. "MatFormer: a generative model for procedural materials". *ACM Transactions on Graphics* 41.4 (2022), 1–12 5.

[HJN22] HÖLLEIN, LUKAS, JOHNSON, JUSTIN, and NIESSNER, MATTHIAS. "StyleMesh: Style Transfer for Indoor 3D Scene Reconstructions". *Proc. CVPR*. 2022, 6198–6208 5.

[IZZE17] ISOLA, PHILLIP, ZHU, JUN-YAN, ZHOU, TINGHUI, and EFROS, ALEXEI A. "Image-to-image translation with conditional adversarial networks". *Proc. CVPR*. 2017, 1125–1134 3.

[JMK*22] JANG, HYEONJOONG, MEULEMAN, ANDRÉAS, KANG, DAHYUN, KIM, DONGGUN, RICHARDT, CHRISTIAN, and KIM, MIN H. "Egocentric scene reconstruction from an omnidirectional video". *ACM Transactions on Graphics* 41.4 (2022), 1–12 1.

[JSI*22] JUNAYED, MASUM SHAH, SADEGHZADEH, AREZOO, ISLAM, MD BAHARUL, WONG, LAI-KUAN, and AYDIN, TARKAN. "HiMODE: A Hybrid Monocular Omnidirectional Depth Estimation Model". *Proc. CVPR Workshops*. June 2022, 5212–5221 2.

[KYYC20] KANG, ZHIZHONG, YANG, JUNTAO, YANG, ZHOU, and CHENG, SAI. "A review of techniques for 3D reconstruction of indoor environments". *ISPRS International Journal of Geo-Information* 9.5 (2020), 330 2.

[LLWL20] LEE, CHENG-HAN, LIU, ZIWEI, WU, LINGYUN, and LUO, PING. "Maskgan: Towards diverse and interactive facial image manipulation". *Proc. CVPR*. 2020, 5549–5558 3.

[LWH*22] LI, ZHEN, WANG, LINGLI, HUANG, XIANG, PAN, CIHUI, and YANG, JIAQI. "PhyIR: Physics-Based Inverse Rendering for Panoramic Indoor Images". *Proc. CVPR*. 2022, 12713–12723 2.

[Ma17] MA, RUI. "Analysis and modeling of 3D indoor scenes". *arXiv preprint arXiv:1706.09577* (2017) 2.

[MLZ*16] MA, RUI, LI, HONGHUA, ZOU, CHANGQING, LIAO, ZICHENG, TONG, XIN, and ZHANG, HAO. "Action-driven 3D indoor scene evolution." *ACM Transactions on Graphics* 35.6 (2016), 173–1 2, 3.

[MRNK21] MÜLLER, THOMAS, ROUSSELLE, FABRICE, NOVÁK, JAN, and KELLER, ALEXANDER. "Real-time neural radiance caching for path tracing". *ACM Transactions on Graphics* 40.4 (2021), 1–16 5.

[PAA*21] PINTORE, GIOVANNI, AGUS, MARCO, ALMANSA, EVA, SCHNEIDER, JENS, and GOBBETTI, ENRICO. "SliceNet: deep dense depth estimation from a single indoor panorama using a slice-based representation". *Proc. CVPR*. 2021, 11536–11545 2, 3.

[PAAG21] PINTORE, GIOVANNI, ALMANSA, EVA, AGUS, MARCO, and GOBBETTI, ENRICO. "Deep3DLayout: 3D Reconstruction of an Indoor Layout from a Spherical Panoramic Image". *ACM Transactions on Graphics* 40.6 (Dec. 2021), 250:1–250:12 2.

Figure 8: *Various examples of virtual object insertion and semantic transfer of furnitures between cluttered and empty rooms.*

[PAAG22] PINTORE, GIOVANNI, AGUS, MARCO, ALMANSA, EVA, and GOBBETTI, ENRICO. "Instant Automatic Emptying of Panoramic Indoor Scenes". *IEEE Transactions on Visualization and Computer Graphics* (Nov. 2022), 1–11 2–4, 6.

[Pad22] PADSTYLER. *3D Architectural Rendering*. [Online; accessed 19-Oct-2022]. 2022. URL: https://www.padstyler.com/ 5.

[PAG20] PINTORE, GIOVANNI, AGUS, MARCO, and GOBBETTI, ENRICO. "AtlantaNet: Inferring the 3D Indoor Layout from a Single 360 Image beyond the Manhattan World Assumption". *Proc. ECCV*. Aug. 2020, 432–448 2.

[PD21] PIRKER, JOHANNA and DENGEL, ANDREAS. "The Potential of 360 Virtual Reality Videos and Real VR for Education—A Literature

Figure 9: *SPIDER-mobile application. The stereoscopics renderer is deployed on Android platform, and the indoor scene can be interactively explored through touch interface or by using IMU sensors for simulating head movements and ready for VR headset like Google Cardboards.*

Review". *IEEE computer graphics and applications* 41.4 (2021), 76–89 1.

[PGG*16] Pintore, Giovanni, Garro, Valeria, Ganovelli, Fabio, Gobbetti, Enrico, and Agus, Marco. "Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps". *Proc. WACV*. Feb. 2016, 1–9 2.

[PLWZ19] Park, Taesung, Liu, Ming-Yu, Wang, Ting-Chun, and Zhu, Jun-Yan. "GauGAN: semantic image synthesis with spatially adaptive normalization". *ACM SIGGRAPH 2019 Real-Time Live!* 2019, 1–1 3.

[PMG*20] Pintore, Giovanni, Mura, Claudio, Ganovelli, Fabio, Fuentes-Perez, Lizeth, Pajarola, Renato, and Gobbetti, Enrico. "State-of-the-art in Automatic 3D Reconstruction of Structured Indoor Environments". *Computer Graphics Forum* 39.2 (2020), 667–699 2.

[QCJK18] Qi, Xiaojuan, Chen, Qifeng, Jia, Jiaya, and Koltun, Vladlen. "Semi-parametric image synthesis". *Proc. CVPR*. 2018, 8808–8816 3.

[SAB*20] Sulaiman, Mohamad Zaidi, Aziz, Mohd Nasiruddin Abdul, Bakar, Mohd Haidar Abu, Halili, Nur Akma, and Azuddin, Muhammad Asri. "Matterport: virtual tour as a new marketing approach in real estate business during pandemic COVID-19". *Proc. IMDES*. 2020, 221–226 2.

[SSC21] Sun, Cheng, Sun, Min, and Chen, Hwann-Tzong. "HoHoNet: 360 indoor holistic understanding with latent horizontal features". *Proc. CVPR*. 2021, 2573–2582 2–4.

[Stu22] Stuccco. *Virtual Staging*. [Online; accessed 19-Oct-2022]. 2022. URL: https://www.stuccco.com/ 5.

[Sty22] Styldod. *Real Estate Virtual Home Staging*. [Online; accessed 19-Oct-2022]. 2022. URL: https://www.styldod.com/ 5.

[TCC*21] Tan, Zhentao, Chen, Dongdong, Chu, Qi, et al. "Efficient semantic image synthesis via class-adaptive normalization". *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021) 3.

[VTS21] Vazquez, Christian, Tan, Nicole, and Sadalgi, Shrenik. "Home studio: a mixed reality staging tool for interior design". *Proc. CHI EA*. 2021, 1–5 2, 3.

[WLLZ20] Wang, Miao, Lyu, Xu-Quan, Li, Yi-Jun, and Zhang, Fang-Lue. "VR content creation and exploration with deep learning: A survey". *Computational Visual Media* 6.1 (2020), 3–28 2.

[WLZ*18] Wang, Ting-Chun, Liu, Ming-Yu, Zhu, Jun-Yan, Tao, Andrew, Kautz, Jan, and Catanzaro, Bryan. "High-resolution image synthesis and semantic manipulation with conditional gans". *Proc. CVPR*. 2018, 8798–8807 3.

[YLR22] Yun, Ilwi, Lee, Hyuk-Jae, and Rhee, Chae Eun. "Improving 360 monocular depth estimation via non-local dense prediction transformer and joint supervised and self-supervised learning". *Proc. AAAI Conference on Artificial Intelligence*. Vol. 36. 3. 2022, 3224–3233 2.

[YZR*21] Yang, Kailun, Zhang, Jiaming, Reiss, Simon, Hu, Xinxin, and Stiefelhagen, Rainer. "Capturing omni-range context for omnidirectional segmentation". *Proc. CVPR*. 2021, 1376–1386 1.

[ZAQW20] Zhu, Peihao, Abdal, Rameen, Qin, Yipeng, and Wonka, Peter. "Sean: Image synthesis with semantic region-adaptive normalization". *Proc. CVPR*. 2020, 5104–5113 3.

[ZCB*22] Zhi, Tiancheng, Chen, Bowei, Boyadzhiev, Ivaylo, Kang, Sing Bing, Hebert, Martial, and Narasimhan, Srinivasa G. "Semantically supervised appearance decomposition for virtual staging from a single panorama". *ACM Transactions on Graphics* 41.4 (2022), 1–15 2–5.

[ZCC16] Zhang, Edward, Cohen, Michael F, and Curless, Brian. "Emptying, refurnishing, and relighting indoor spaces". *ACM Transactions on Graphics* 35.6 (2016), 1–14 3.

[ZKG20] Zeng, Wei, Karaoglu, Sezer, and Gevers, Theo. "Pano2Scene: 3D Indoor Semantic Scene Reconstruction from a Single Panorama Image". 2020 2.

[ZLW*22] Zhuang, Chuanqing, Lu, Zhengda, Wang, Yiqun, Xiao, Jun, and Wang, Ying. "ACDNet: Adaptively combined dilated convolution for monocular panorama depth estimation". *Proc. AAAI Conference on Artificial Intelligence*. Vol. 36. 3. 2022, 3653–3661 2.

[ZZL*20] Zheng, Jia, Zhang, Junfei, Li, Jing, Tang, Rui, Gao, Shenghua, and Zhou, Zihan. "Structured3d: A large photo-realistic dataset for structured 3d modeling". *Proc. ECCV*. 2020, 519–535 2.