Athens Journal of Sciences- Volume 12, Issue 3, September 2025 – Pages 205-234

Exploring Architectural Design 3D Reconstruction Approaches through Deep Learning Methods: A Comprehensive Survey

By Mehdi Gorjian[*], *Stephen M. Caffey*[±] *& Gregory A. Luhan*[°]

In recent years, substantial advancements in computer vision have been propelled by the advent of deep learning techniques. The transformative impact of deep learning is evident across various computer vision applications, spanning object detection, image classification, and semantic segmentation. Within the dynamic realm of architectural parametric design, a pivotal force shaping contemporary and future architectural pursuits, designers frequently encounter scenarios necessitating the reshaping, augmentation, or modification of building components and facades. This demand arises from diverse factors, including the integration of human-interactive facades, the implementation of sustainable facade systems, and more. In such instances, the foundational structure must possess malleability, enabling openness to edits. However, the grand scale and intricate details inherent in architectural designs pose a challenge, as a densely detailed mesh comprising millions of elements becomes increasingly unwieldy and challenging to modify for subsequent stages. This challenge underscores the critical need to generate objects capable of seamless evolution in tandem with evolving design requirements. This paper, will focus on categorizing 3D deep learning approaches into Learning-based 3D reconstruction in architecture and delving into their methods.

Keywords: *Algorithm Development, Computational Design, Learning-based Generative Design, Deep Learning, 3D Reconstruction*

Introduction

Architecture is intensively relying on the concept, philosophy, and perception of the environment and the human behavior in the three-dimensional space and over the time. The new realm of the digital fabrication and rapid manufacturing is the field that requires an automated approach to accurately help architects during the design and manufacturing process. Besides the significant transformative impact of the deep learning in object detection, classification, and segmentation, its semantic inference capability makes it more powerful and suitable to analyze and interpret the data. Deep learning in three dimensions has the potential to transform the field of architecture by introducing novel approaches to construction, maintenance, and design. This entails the design and enhancement of neural network architectures algorithms that possess the ability to analyze and extract significant insights from a wide range of three-dimensional data formats, such as point clouds, meshes, volumetric data, and depth maps. 3D deep learning has the potential to greatly augment Building

[*]R&D Software Engineer, Motion View LLC. Chattanooga, USA.
[±]Associate Department Head, Texas A&M University College Station, USA.
[°]Department Head of Architecture, Texas A&M University College Station, USA.

https://doi.org/10.30958/ajs.12-3-4 doi=10.30958/ajs.12-3-4

Information Modeling (BIM) in the discipline of architecture through the automation of model development and refinement, resulting in improved accuracy and efficiency. By facilitating the generation of numerous design options that are optimized for aesthetics, structural integrity, and sustainability, it facilitates generative design and optimization. In addition, it facilitates the development of dynamic and intricate building facades and forms, thereby augmenting both functionality and aesthetic appeal. The convergence of deep learning and interactive augmented reality (AR) en-ables the creation of immersive design experiences by facilitating the visualization and interaction with designs in real-time. Moreover, the utilization of 3D deep learning is of major significance in the digital reconstruction and preservation of historical structures within the 3D dense reconstruction, thereby contributing to the maintenance of cultural heritage. By harnessing these capabilities, architects have the potential to attain elevated levels of ingenuity, accuracy, and productivity, ultimately revolutionizing the architectural domain. The integration of 3D deep learning into digital fabrication, robotic fabrication, and 3D printing for large-scale buildings is revolutionizing the construction industry. These technologies enable greater precision, efficiency, and innovation in building processes. Applying deep learning to the digital fabrication process prepares a context to streamline the generative design to the fabrication step by optimizing material and improving the structure, ensuring the fabricated components pass the required specifications and also monitoring the production line for defect detection and for adaptive path planning in robotics fabrication and assemblies. However, our paper would be the initial one that concentrates on 3D reconstruction methods for architecture, there are other papers that explain the method, such as (Lu et al. 2005, Choy et al. 2016, Seitz et al. 2006, B. Zhao et al. 2018, Samavati & Soryani 2023, Pan et al. 2019, Parket al. 2019).

A significant contribution of this work, compared to existing literature, can be summarized as follows:

1. To the best of our knowledge, this is one of the first comprehensive survey papers covering 3D reconstruction in architecture.
2. The paper addresses both classical and deep learning-based methodologies, providing a broader perspective on 3D reconstruction capabilities.
3. Unlike existing reviews, we specifically focus on 3D reconstruction and its applications in architecture.
4. The paper aims to include the most recent and robust techniques in this field, highlighting their relevance in accurate industry workflows and research.

This paper is structured as follows: Section *ii* provides an overview of 3D reconstruction and its applications in architectural design and digital fabrication. Section *iii* delves into classical 3D re-construction methods, while Section *iv* explores deep learning-based 3D reconstruction techniques. Section *v* presents a discussion, highlighting the technology's impact on architectural design. Finally, Section *vi* offers the conclusion.

Background

3D Reconstruction

Extending deep convolutional networks, the cornerstone of image-based techniques, to the third spatial dimension presents a number of challenges. Because of the higher data dimensionality, these networks have a tendency to multiply in terms of both time and space complexity. Conventional dense surface representations, like quad or triangular meshes, pose more challenges to the training process. These challenges arise from the randomly generated topology and the fluctuating vertex count of these models. These problems have severely limited the deep learning algorithms' efficacy, flexibility, and accuracy. Applying these algorithms to 3D data input analysis or generating object segmentations in 3D space particularly highlights these limitations. Because 3D data is inherently complicated, more advanced models and computing power are required, which might hinder training and reduce algorithmic performance. The inconsistent mesh topologies make it more difficult to create reliable and effective deep learning frameworks. Since every 3D model can have a different structure, algorithms must be able to adjust to a variety of vertex combinations and geometries. This heterogeneity necessitates the use of more sophisticated methods for data augmentation and preprocessing, in addition to increasing the computational load. There are different methods to construct a 3D mesh from an image, such as:

- *Classical 3D Reconstruction:* Traditional reconstruction methods usually adopt discrete representations, such as voxel grids, 3D point clouds, and depth maps. The depth map representation is flexible and appropriate for parallel computation, so it is the most widely used representation (Long et al. 2022). Therefore, traditional techniques often provide noisy and partial reconstructions because they are susceptible to reflections, picture noise, and variable illuminations.
- *Deep Learning-based 3D Reconstruction:* These methods typically represent the surfaces to be reconstructed as Signed Distance Function (SDF) and utilize surface rendering or volume rendering to render the pixel colors of the input images for optimizing the SDF fields (Long et al. 2022).

3D Reconstruction in Architecture

In terms of conceptualizing, visualizing, and realizing forms and structures, 3D reconstruction can have an impact on digital production and architectural design. Architects and designers can produce extremely realistic and detailed models of buildings and other structures by utilizing sophisticated computational tools. This improves planning, analysis, and communication during the design process. 3D reconstruction in architecture makes it possible to create detailed digital representations of real locations.

- *Conceptualization and Visualization:* Using accurate 3D models that offer a realistic image of the proposed structure, architects can develop and revise design concepts. This facilitates the visualization of the design's spatial relationships, proportions, and aesthetics, improving stakeholders' comprehension and assessment of the undertaking.
- *Detailed Element Design:* 3D models allow for accurate planning of structural elements, materials, and construction methods. These models can be used to simulate and optimize various design aspects, such as load distribution, energy efficiency, and acoustics, leading to more informed decision-making.
- *Collaboration and Virtual Interaction:* The use of 3D reconstruction enhances communication among architects, engineers, and clients. Web-based interactive technologies, virtual reality, and augmented reality can be used to instantly exchange and analyze interactive 3D models, guaranteeing that everyone involved has a consistent and clear knowledge of the design concept.

Classical 3D Reconstruction

Structure from Motion (SfM): Snavely et al. (2006) presented a method using camera pose (location, orientation, and field of view) and sparse 3D scene information to create new interfaces for browsing large collections of photographs. Knowing the camera pose enables placing the images into a common 3D coordinate system and allows the user to virtually explore the scene by moving in 3D space from one image to another using our photo explorer. We use morphing techniques to provide smooth transitions between photos. In addition, the re-constructed features provide a sparse but effective 3D visualization of the scene that serves as a backdrop for the photographs themselves. Table. 2 shows popular SfM packages and software.

Multi-View Stereo: Seitz et al. (2006) categorized multi-view stereo algorithms into four classes. The first class operates by first computing a cost function on a 3D volume and then extracting a surface from this volume. A simple example of this approach is the voxel coloring algorithm and its variants, which make a single sweep through the volume, computing costs and reconstructing voxels with costs below a threshold in the same pass. The second class of techniques works by iteratively evolving a surface to decrease or minimize a cost function. This class includes methods based on voxels, level sets, and surface meshes. Space carving and its variants progressively remove inconsistent voxels from an initial volume. Other variants of this approach enable adding as well as deleting voxels to minimize an energy function. Level-set techniques minimize a set of partial differential equations defined on a volume. Like space carving methods, level-set methods typically start from a large initial volume and shrink inward; unlike most space carving methods, however, they can also locally expand if needed to minimize an energy function (Seitz et al. 2006). The third class is image-space methods that compute a set of depth maps. To ensure a single consistent 3D scene interpretation, these methods enforce consistency constraints between depth maps or merge the set of depth maps into a 3D scene as a post process. The final class consists of

algorithms that first extract and match a set of feature points and then fit a surface to the reconstructed features (Seitz et al. 2006). The Multi-View Stereo (MVS) technique utilizes *Structure from Motion (SfM)* output to compute depth and normal data for each pixel within an image. The amalgamation of depth and normal maps from various images in a three-dimensional space subsequently yields a comprehensive point cloud representation of the scene. Table. 1 compares SfM and MVs.

Table 1. *Structure from Motion (SfM) vs. Multi-View Stereo (MVS)*

SfM	**MVS**
Generally employs a collection of photos not arranged in a particular sequence, captured from diverse views and time instances, and frequently exhibiting variations in focal lengths and resolutions.	It often involves a collection of calibrated photos captured from predetermined angles with constant camera settings.
In the context of camera calibration, SfM is a technique used to estimate the calibration parameters, such as the focal length, camera rotation, and camera translation, as an integral component of the overall process.	It often necessitates the use of pre-calibrated cameras that possess predefined characteristics.
Utilizes the process of feature matching be-tween pictures to ascertain the relative cam-era postures and the scene's sparse three-dimensional (3D) structure.	It encompasses several techniques, including direct methods that do not need explicit feature matching and methods that include feature matching.
Initially generates a sparsely reconstructed scene by detecting a restricted set of points. Dense reconstruction can be accomplished as the following procedure by utilizing depth map estimation techniques.	Frequently pursues the objective of achieving dense reconstruction in a direct manner, wherein the complete visible surface of the 3D object or scene is reconstructed.
The technique of inferring the 3-dimensional structure of a scene and the camera's motion is commonly referred to as SfM.	The primary objective of multi-view re-construction is to reconstruct the three-dimensional structure by utilizing numerous views without explicitly considering the camera's motion.
It is often employed in several domains, including photogrammetry, robotics, and computer vision jobs that utilize unordered pictures captured by handheld cameras or drones.	It is a widely employed technique in several fields, including computer graphics, medical imaging, and industrial applications, mainly when calibrated cameras are accessible.

Simultaneous Localization & Mapping: Simultaneous Localization and Mapping (SLAM) is a technique employed to get information on the surrounding environment by utilizing various sensors such as cameras, LIDAR (Light Detection and Ranging), and IMUs (Inertial Measurement Units). The system concurrently calculates the estimates: *Localization,* refers to determining the location and orientation, also known

as the pose, of a robot or device within its surrounding environment. *Mapping,* The process of producing a visual representation, known as a map, that depicts the spatial arrangement of the surrounding physical or conceptual world. SLAM aims to resolve the inherent challenge of the chicken-and-egg dilemma. In order to construct a map, the system needs knowledge of its stance. However, to estimate its pose, the system relies on the existence of a map. The SLAM algorithm effectively addresses this issue through its iterative process of updating both the map and pose estimations in response to incoming input. SLAM methodologies can be categorized into *feature-based, grid-based,* and *topological techniques*. The *feature-based* technique uses recognizable elements like corners for mapping and localization. *Grid-based* techniques discretize the environment into grids, including occupancy information, and use this information to estimate the map and pose. *Topological* techniques represent the environment as a graph. The SLAM system consists of three components: *data association, state estimation, map update,* and *loop closure*. Data association matches observable features in the environment with map features. State estimation infers the robot's current state, including posture and map, based on sensor data. Map update integrates recent data, and loop closure identifies robot return instances for improved accuracy and consistency. There are three popular types of SLAM algorithms *Filter-based SLAM, Graph-based SLAM,* and *Visual SLAM (VSLAM)*. Table. 2 shows popular SLAM packages and software.

Reconstructing with Voxels Using Occupancy Grid: This representation discretizes the three-dimensional space into a grid of uniform cubic parts. Each part has two states: empty or occupied. Due to memory limitations, the grid resolution is often limited to 10243 and lower. Despite this limitation, this type of representation easily fits into the deep learning frameworks. Therefore, models that infer 3D shapes using this representation are very easy to implement (Samavati & Soryani 2023). Choy et al. (2016) introduced a 3D Recurrent Reconstruction Neural Network (3D-R2N2). This network is designed to understand the 3D shape of objects by learning from a wide array of synthetic data and mapping images to their respective 3D structures. The network processes one or multiple images of an object from any angle and then generates a 3D occupancy grid, essentially creating a reconstructed version of the original object.

Table 2. *Popular SfM & SLAM Software Packages*

Package	Description	Method
COLMAP	A 3D reconstruction software that includes both sparse and dense matching modules, which includes Radial and Fish-eye camera models (Schonberger & Frahm, 2016).	SfM
OpenMVG	A well-known sparse reconstruction software that provides Radial, Brown, Fisheye, and Sphere, camera models. It can directly process spherical images (Moulon, Monasse, Per-rot, & Marlet, 2017).	SfM
OpenVSLAM	A visual SLAM system that supports various camera models, including perspective, fisheye, and spherical cameras (Sumikura, Shibuya, & Sakurada, 2022).	SLAM
Cubemap-SLAM	A visual SLAM system that converts fisheye images into perspective images and achieves image orientation based on the ORB-SLAM (Y. Wang et al., 2018).	SLAM

Space Carving: By framing shape reconstruction as a constraint satisfaction problem, Kutu-lakos and Seitz (Kutulakos & Seitz 2000) established an approach to reconstruct 3D scenes from photographs. In their paper, they demonstrated that any collection of photographs of an immutable scene defines a set of image constraints delighted by each scene that can be projected onto those images. They represented "the set of all 3D shapes that satisfy these constraints and use the underlying theory to design a practical reconstruction algorithm, called *Space Carving*, that applies to fully-general shapes and camera configurations" (Kutulakos & Seitz 2000). Space Carving (Kutulakos & Seitz 2000, Broadhurst et al. 2001, Slabaugh et al. 2004, Aykin & Negahdaripour 2016, R. Yang et al. 2003) or Volumetric Reconstruction techniques involve the removal of space from the volume based on the coherence of 2D images captured from multiple perspectives.

Depth Map Fusion: A number of techniques have been developed for reconstructing surfaces by integrating groups of aligned range images. A desirable set of properties for such algorithms includes incremental updating, representation of directional uncertainty, the ability to fill gaps in the reconstruction, and robustness in the presence of outliers (Curless & Levoy 1996). Using depth-sensing cameras to generate a set of depth maps (Merrell et al. 2007, Zach 2008, Zhu et al. 2010) combined to create a 3D model constitutes Depth Map Fusion. Generally speaking, depth map fusion represents a technique applied to computer vision to blend *(fuse)* several depth maps into an individual, precise, or comprehensive depth map. Depth maps are images in which each pixel comprises a value indicating the distance from the object at that image location to the camera.

Light Field Methods: Light field (Gershun 1939, Levoy & Hanrahan 1996, W.C. Chen et al. 2002) imaging has emerged as a technology allowing us to capture richer visual information from our world. As opposed to traditional photography, which captures a 2D projection of the light in the scene integrating the angular domain, light fields collect radiance from rays in all directions, demultiplexing the angular information lost in conventional photography. On the one hand, this higher dimensional representation of visual data offers powerful capabilities for scene

understanding and substantially improves the performance of traditional computer vision problems such as depth sensing, post-capture refocusing, segmentation, video stabilization, material classification, etc. (G. Wu et al. 2017).

Poisson Surface Reconstruction: (Kazhdan et al. 2006, Bolitho 2009, Estellers et al. 2015) Poisson surface reconstruction algorithm is an implicit function method. It combines global and local fitting schemes. Yet, the basis functions are associated with the ambient space rather than the data points, are locally supported, and have a simple hierarchical structure that results in a sparse, well-conditioned system. But this algorithm reconstructs the isosurface using the Marching Cube method (Long et al. 2022, Newman & Yi 2006, Lopes & Brodlie 2003), which will generate a simple smooth surface and waste lots of time on the detecting of non-null cubes (X. Li et al. 2010).

Ball Pivoting Algorithm (BPA): (Bernardini et al. 1999, Digne 2014, Stelldinger 2008) is a method for reconstructing surfaces from unorganized point clouds. The BPA, initially proposed by Bernardini et al. (1999), has been broadly adopted because of its simplicity, effectiveness, and excellent outcome. According to (Bernardini et al. 1999), the method is conceptually simple. Starting with a seed triangle, it pivots a ball around each edge on the current mesh boundary until a new point is hit by the ball. The edge and point define a new triangle, which is added to the mesh, and the algorithm considers a new boundary edge for pivoting. The output mesh is a manifold subset of an alpha shape of the point set. Some of the nice properties of alpha-shapes can also be proved for our reconstruction.

Signed Distance Function (SDF): The Signed Distance Function (SDF) (Zhou, et al. 2016, Bylow et al. 2013, Osher et al. 2003, Zollhöfer et al. 2015) is explained, given its intersection with the categorization of 3D deep learning. In simultaneous localization and mapping, 3D reconstruction and surface tracing are con-ducted, which presents a continuous and differentiable surface representation that is advantageous for camera estimation poses and reconstructing the scene's geometry.

Determining image pixels' depth ranks among the foremost fundamental challenges of 3D re-construction. Deducing the spatial structure of an environment from two-dimensional images is the task of depth estimation. The objective of this task is to recover the crucial spatial data that was lost during the 3D-to-2D projection procedure during image capture. Classical 3D reconstruction approaches determine the depth of pixels using Triangulation and Epipolar Geometry. Nevertheless, these techniques require knowledge of internal and external parameters associated with the cameras, and the images must be captured with minor rotational and translational adjustments to the camera. Based on these algorithms, a variety of methodologies have been established for 3D re-construction algorithms. In addition to Poisson Surface Reconstruction, Ball Pivoting Algorithm, and Signed Distance Function methods, there are other algorithms such as Marching Cube and Truncated Signed Distance (TSDF).

Deep Learning-based 3D Reconstruction

Applying deep learning techniques to the processing and analysis of three-dimensional data is known as 3D deep learning. It involves the creation of neural network designs and algorithms that can comprehend and extract useful information from 3D data, including depth maps, 3D models, point clouds, and volumetric data. The efficient handling of 3D data by deep learning opens up a variety of applications in many industries. These domains cover a wide range of industries, such as architecture, computer-aided design (CAD), robotics, autonomous cars, augmented reality (AR), medical imaging, and game creation.

3D deep learning extends the utilization of deep learning algorithms to 3D data, such as point clouds (Achlioptas et al. 2018, L. Jiang et al. 2018, Fan et al. 2017, K. Li et al. 2018, Thomas et al. 2019, G. Yang et al. 2019), volumes (Brock et al. 2016, Choy et al. 2016, Gadelha et al. 2017 Jimenez Rezende et al. 2016, Riegler et al. 2017, Stutz & Geiger 2018), multi-view images (Xie et al. 2019, Rhodin et al. 2018, Xu et al. 2016, Tatarchenko et al. 2016, B. Zhao et al. 2018), meshes (Zhou & Tuzel 2018, Groueix et al. 2018, Kanazawa et al. 2018, Liao 2018, Pan et al. 2019), etc. as opposed to traditional deep learning, which primarily operates on 2D data. This shift enables machines to interpret and analyze a more comprehensive and detailed representation of physical entities, thereby producing more accurate results and decisions. As a consequence, it is increasingly being utilized in diverse areas such as autonomous vehicles, robotics, virtual reality, and healthcare (Z. Wu et al. 2015). The methodology of 3D deep learning encompasses a multitude of strategies, each serving a different function and possess-ing certain advantages and disadvantages. The main methods include volumetric, multi-view, point cloud-based, and graph-based.

One pioneering 3D deep learning technique involves the volumetric analysis of 3D data. These methods convert 3D data into a volumetric grid and then apply traditional 3D convolutions to capture spatial correlations. An example is VoxNet, a 3D Convolutional Neural Network (3D-CNN) that operates on voxel grids to recognize and classify 3D shapes (Maturana & Scherer 2015). 3D-CNNs are a direct extension of traditional 2D-CNNs to the 3D domain, enabling the processing of 3D data directly (Ji et al. 2012). Various voxel-based approaches have been developed following this work, such as PointNetVLAD by Zu et al. (Uy & Lee 2018), which aggregated local features for global scene representation using an encoding module. Nevertheless, volumetric methods are typically computationally intensive because of the substantial dimensionality of 3D data. Researchers have proposed multi-view methods as a solution to the computational difficulties of volumetric methods. MVCNNs process 3D data by projecting it onto multiple 2D views and then analyze these 2D views using 2D-CNNs (Su 2015). However, this method may fail to capture certain spatial relationships as some information is lost during the projection process. The PointNet architecture developed by Qi et al. (2017) has become among the most widely recognized within this category. It explicitly applies neural networks to point clouds, considering the input an unsorted collection of points. Later architectures, including PointNet++ (Qi et al. 2017) and DGCNN (Y. Wang et al. 2019), enhanced the authentic PointNet by more accurately preserving the

point cloud's local structures. PointNet operates directly on 3D point cloud data, which are unordered sets of points in a 3D coordinate system (Qi et al. 2017). Other notable point-based methods include PointConv by Wu et al. (2019), which introduced learnable convolutional operators on point clouds. Graph-based strategies encode 3D data as graphs, manipulating the raw connectivity data within 3D shapes. The Graph CNN (GCNN) by Bronstein et al. (2017), which extrapolates the convolution layers to non-Euclidean spaces, is a seminal work in this field. These strategies are able to represent complex spatial relationships; however, graph creation from point clouds can be complex. GNNs are designed to process data structured as graphs, a very generic form of data, which allows them to handle a variety of 3D data formats (Bronstein et al. 2017). Graph Attention Networks (GAT) by Veličković et al. (2017) utilized attention mechanisms to weigh the importance of neighboring points during information propagation.

3D Geometry is a subfield of mathematics concerned with studying three-dimensional shapes and figures. Numerous shapes are involved, including cubes, pyramids, spheres, and cylinders, and are widely utilized in computer graphics, simulations, and game design due to the advancement of computer technology. As Kato et al. (2020) stated, "Most 3D estimation methods rely on supervised training regimes and costly annotations, which makes the collection of all properties of 3D observations challenging. Hence, there have been recent efforts towards leveraging easier-to-obtain 2D information and differing levels of supervision for 3D scene understanding. One of the approaches is integrating graphical rendering processes into neural network pipelines. This allows transforming and incorporating 3D estimates into 2D image-level evidence." Rendering in computer graphics is the process of generating images of 3D scenes defined by geometry, materials, scene lights, and camera properties. Rendering is a complex process, and its differentiation is not uniquely defined, which prevents straightforward integration into neural networks (Kato et al. 2020). The technique of Differentiable Rendering (DR) (Loper & Black 2014, Loubet et al. 2019, J. Chen et al. 2019) bridges the divide between 3D graphics and artificial intelligence. It is a process by which the procedure for rendering becomes differentiable, i.e., output gradients can be computed concerning their input which enables backpropagation of gradients, a technique essential for training neural networks.

Differentiable Rendering (DR) constitutes a family of techniques that tackle such integration for end-to-end optimization (Kotary et al. 2021, Ballé et al. 2016, Sitzmann et al. 2018) by obtaining useful gradients of the rendering process. By differentiating the rendering, DR bridges the gap between 2D and 3D processing methods, allowing neural networks to optimize 3D entities while operating on 2D projections. Previously, rendering was a process that generated a 2D image from 3D shapes, textures, illumination, and perspective. This function is typically not differentiable everywhere, making it challenging to use conventional optimization (Reddy et al. 2017, Banos et al. 2011) methods or develop deep learning models that entail a rendering phase. DR algorithms attempt to achieve the rendering operating as differentiable as feasible, allowing gradients to be backpropagated over it with various applications, including training deep neural networks that generate 3D shapes and adjusting the hyper-parameters of a 3D model to suit an image.

Differentiable rendering algorithms can be divided primarily into four main categories based on (Kato et al. 2020). Here, we discuss differentiable rendering in detail:

Mesh-based Methods

Analytical Derivative: According to (Kato et al. 2020), "a mesh represents a 3D shape as a set of vertices and the surfaces that connect them." It is frequently utilized, particularly in computer graphics, due to its ability to represent sophisticated three-dimensional shapes compactly. Given the rendering data as input sources and pixel information, estimating the color consists of two steps. First, defining a triangle around the pixels and, second, calculating the pixel color based on the hues of the vertices from the allocated triangle.

Approximated Gradient: In optimization methods, an approach known as approximated gradient applies to the estimation rather than the accurate calculation of a function's gradient. This method is frequently employed when computing the gradient would be expensive or complex. The gradient may be estimated by making specific assumptions or applying numerical techniques, which enables the use of gradient-based optimization techniques. As Kato et al. (2020) stated, "gradients should be objective-aware for better optimization. Since the objective of OpenDR is not to provide accurate gradients but to provide useful gradients for optimization, a loss-aware gradient flow is required for this purpose."

Approximated Rendering: (Reeves & Blau 1985, Lighthill 1949, Loper & Black 2014) In-stead of approximating the backward pass, other methods approximate the rasterization (or the forward pass) of the rendering in order to be able to compute useful gradients (Kato et al. 2020). Creating a picture from a 3D model is known as approximate rendering, focusing more on speed and efficiency than precise accuracy. These methods can quickly create visuals that are visually near to a correct depiction by making approximations and simplifications. This method is frequently applied in scenarios where rendering in real-time is necessary, as in simulations or video games.

Global Illumination: (H. Jiang et al. 2021, Y. Li et al. 2022, Jensen 1996, Sillion et al. 1991, Landis 2002) is an essential notion in 3D reconstruction that describes how light is reflected and dispersed throughout a scene. Global lighting considers the indirect impacts of light because it bounces and scatters across a scene, in contrast to local or direct illumination, which solely analyzes the direct interaction across lighting sources and surfaces. A reconstructed picture may become much more realistic with accurate modeling of global lighting. Still, it also becomes more complicated since it must replicate all the ways light interacts with the environment's surfaces.

Voxel-based Methods

In three-dimensional space, a voxel, a phrase derived from *volume* and *pixel,* denotes a value on a regular grid. It is a pixel's two-dimensional counterpart in three dimensions. Volumetric in-formation is determined in a Cartesian coordinate system and is contained in each voxel. A voxel carries information, including density, color,

or material attributes, based on the context, much like a pixel has a particular color associated with it. Before delving into the specifics, let's clarify two key concepts:

Occupancy Grid: This is a representation of space in the form of a grid, where each cell denotes its occupancy status—either occupied or free. Typically, cells can have binary values, indicating whether they are occupied or not. Alternatively, they can hold a probability value that signifies the chances of the cell being occupied

Deep Learning-based Signed Distance Function (SDF): This describes space as a continuous scalar field in which every point is assigned a value corresponding to the shortest distance to the closest surface. These values are positive when the point lies outside the surface and negative when it's inside the surface.

It is common practice to encode occupancy (Snow et al. 2018) in voxel occupancy, the scene is represented as a set of 3-dimensional voxels, and the task is to label the individual voxels as filled or empty (Snow et al. 2000). Voxel information using a binary value and transparency using a non-binary one. Shapes can be represented as the shortest path from the center of each voxel to the surface of the object rather than by storing occupancy. The shortest distance between a point and a surface is returned by a computer graphics algorithm called a Signed Distance algorithm (SDF) (Bylow et al. 2013, Osher et al. 2003, Zollhöfer et al. 2015). This distance is positive if the point is outside the surface and negative if it is within, as indicated by the word "signed" in the function's name. SDFs, as implicit surface reconstruction methods, are used extensively in generating and manipulating intricate geometric objects. SDFs have the ability to specify a form in space with just one function, which makes them an effective tool for developing procedural content. SDFs may be used to generate complicated forms with less computational expense than conventional polygon-based models, which is especially beneficial in real-time 3D rendering and graphics.

Point Cloud-based Methods

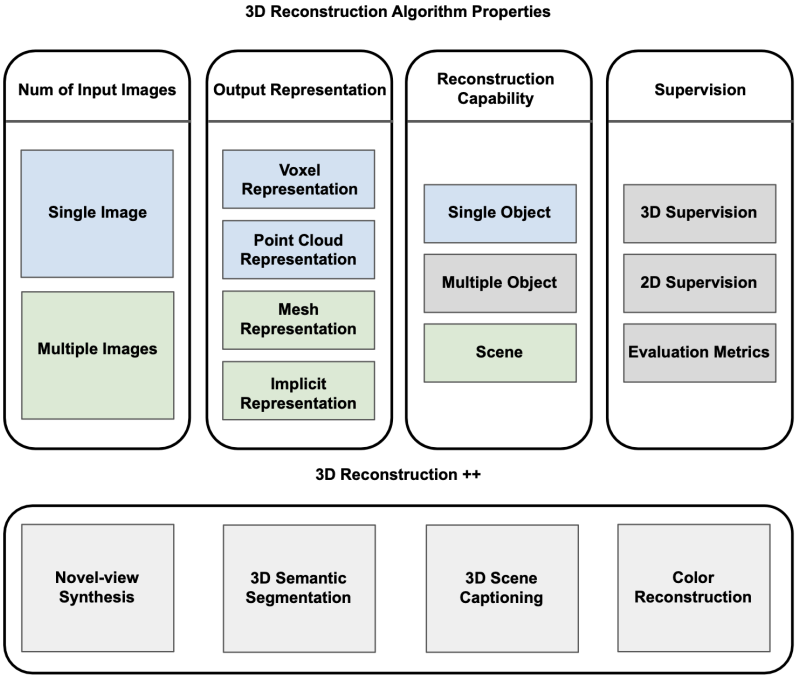
A point cloud is a group of data points arranged in three dimensions. Each cloud point cor-responds to a location where a surface was identified. A considerable amount of points are often measured on the surfaces of nearby objects using 3D scanners or photogrammetry software, which produces a file with these 3D coordinates. There are three primary processes for rendering a point cloud. Initially, matrix multiplication is used to determine a 3D point's screen space coordinates, much like when rendering a mesh. This makes this step differentiable. Secondly, each 3D point's impact on the color of the target pixel is calculated. To get the final color value for each pixel, the points are finally combined based on their influences and z-values. Several techniques have been put forth to successfully handle this last stage of aggregation.

Implicit Methods

(Niemeyer et al. 2020, Liu et al. 2020, Cole 2021) Using functions that implicitly determine a form, implicit representations in differentiable rendering approaches indicate a technique for modeling and rendering 3D shapes. In contrast to traditional

3D modeling, which employs explicit representations such as meshes or point clouds to specify the vertices and edges of a shape, implicit 3D modeling uses a function to specify whether a particular point in 3D space is located within or outside the shape. Unlike voxel-based methods, in implicit representations, memory usage is constant with respect to spatial resolution. Therefore, a surface can be reconstructed at infinite resolution without excessive memory footprint. In implicit representations, memory utilization is constant with respect to spatial resolution, unlike voxel-based techniques. This allows for the recreation of a surface at an infinite resolution and with a minimal memory footprint.

Figure 1. *3D Reconstruction Methods' Properties Overview*



3D Reconstruction in Architectural Design Discussion

Deep Learning-based 3D Reconstruction Model Selection Criteria

Architectural forms can be found in a wide range of styles, from the elaborate intricacies of Gothic designs to the sleek minimalism of Modern architecture and even experimenting with avant-garde shapes. The extreme range of architectural entities in terms of scale and nature is a significant difficulty in this field. The difficulty of creating a complete dataset encompassing all these styles stems from the scattered and frequently insufficient data for individual architectural disciplines.

Because of these limitations, choosing a model that can produce reliable findings without excessively depending on large datasets was essential. Architectural entities are highly detailed stylistically, and their actual forms vary widely. They can be as large as skyscrapers or lightweight structures that resemble shelters and

are sometimes referred to as non-watertight structures. Because of this variety, a method that is adaptable enough to allow reconstructions regardless of the structural complexity of the form is required. Based on (Samavati & Soryani 2023), "as voxel grids are discrete, the output resolution must be kept at low levels due to resource constraints. Thus, the reconstructed shapes are low-detailed, and the reconstruction accuracy is low. Additionally, this type of representation cannot be directly used in some applications, such as game and movie industries, as it requires a different algorithm to obtain more flexible and natural outputs." Therefore, a middle-tier method that can enhance voxel outputs and give them a more flexible and realistic appearance is required. Adding to these difficulties is that voxels (3D-R2N2 (Choy et al. 2016), (Xie et al. 2019)) naturally yield results with lower resolution, primarily because of memory usage limitations. There is, however, a bright side. Based on the observations made by Kato (2020), it is clear that voxels can be effectively converted into useful informational resources by using specific methods. Using the Signed Distance Function (SDF) (Takikawa et al. 2021, Y. Jiang 2020) is one technique that expands the possible uses of voxel-based reconstructions; however, SDF is useful when used for watertight shapes.

Point Cloud-based approach does a more accurate job with a better resolution. According to (Samavati & Soryani 2023), "A point cloud is simply a set of unordered 3D points in a coordinate system. Compared to ordered structures like meshes that store combinational connectivity patterns, predicting 3D shapes in point-cloud representation is less challenging." It may use different types of networks, mainly the Encoder-Decoder model (Zhang 2020). However, there are advances to the primary models using Multilayer Perceptron (MLP), especially for the shape completion (Qi et al. 2017). Point Cloud-based methods mainly rely on AE models, and they require training steps in which the decoder knowledge is confined to the 3D data shapes and categories used in the training phase.

Typically, conventional signal representations exhibit discreteness. For example, when rep-resenting 3D shapes, they are commonly specified as grids consisting of voxels, point clouds, or meshes. Conventional signal representations employ a discrete methodology. Consider 3D shapes as an illustrative example, which are frequently depicted through the utilization of voxel grids, point clouds, or mesh structures. In contrast, implicit neural representations adopt another approach. The researchers define a signal as a continuous function that establishes a connection between the domain of the signal and the corresponding value or entity associated with that coordinate. This value could represent either an occupancy probability or an SDF (Park et al. 2019) (also see Table. 3) value. Instead of employing explicit representations, implicit neural techniques approximate the continuous function by utilizing an MLP network. MLP possesses the capability to replicate any function to a notable extent accurately. One notable advantage of these strategies is their high level of memory efficiency. This methodology enables the neural network to accurately infer complex 3D shapes at different resolutions. One prominent limitation of the SDF models is "limited to objects with closed surfaces since they adopt Signed Distance Function (SDF) as a surface representation which requires the target shape to be divided into inside and outside", (Long et al. 2022). Additionally, methods such

as NeuralUDF (Long et al. 2022) (Unsigned Distance Function) try to rectify this issue; a comprehensive training dataset is inevitable.

In the past, SfM, multi-view stereo, and depth-map fusion-based methods were widely used for active 3D acquisition tasks. However, as alluded to above, these classical approaches have limitations in the dense 3D reconstruction of an object (Lee et al. 2022). Additionally, classical methods like multi-view have been integrated with deep learning for better performance and results, such as the work by Niemeyer et al. (2020). According to (Lee et al. 2022) "NeRF learns an implicit function to represent complex 3D shapes such that the novel poses of the object can be synthesized. The implicit function is learned by an MLP (Samavati & Soryani 2023). NeRF is a simple yet effective volume rendering approach. It represents the continuous static scene as 5D neural radiance fields, parameterized by multi-layer perceptron (MLP). It demonstrated that regressing density and light fields via an MLP could yield photo-realistic rendering. Due to its remarkable ability to capture complex geometric details, it gathered significant interest from the community (Deng 2022). Although NeRF requires more views than other models and the training will be expensive; however, according to (Deng et al. 2022) "Neural rendering with implicit representations has become a widely-used technique for solving many vision and graphics tasks ranging from view synthesis to re-lighting, to pose and shape estimation, to 3D-aware image synthesis and editing, to modeling dynamic scenes. On the other hand, NeRF is built upon perspectives of particular objects rather than relying on extensive collections of diverse data. This methodology places greater emphasis on the visual representation of the object as opposed to the quantity or diversity of data points, which can be adjustable for a variety of architectural styles. The seminal work of Neural Radiance Fields (NeRF) demonstrated impressive view synthesis results by using implicit functions to encode volumetric density and color observations.'' Overall, it is important to consider that every solution has its drawbacks. However, we pursued a solution offering greater efficiency, enhanced flexibility, and reliance on cases. Table. 3 compares the SDF-based methods with NeRF architecture.

Table 3. *Comparison between SDF and NeRF*

SDF	NeRF
Is a mathematical representation that depicts a three-dimensional shape by assigning a scalar value to every point in space, representing the shortest distance from that point to the shape's surface. The scalar's sign indicates whether the point is located inside *(negative)* or outside *(positive)* the geometric shape.	Employs a neural network architecture to effectively model and describe a continuous volumetric scene function. The process involves mapping a three-dimensional coordinate and a viewing direction to determine the radiance value and volume density.
Is employed in form representation, reconstruction, and generation.	Is designed to generate original perspectives of a particular scene based on a collection of two-dimensional photographs captured from predetermined vantage points.
Suitable for shape completion, shape interpolation, 3D modeling, and more.	Novel view synthesis, 3D scene representation from 2D images, AR/VR.
Focusing on representing the shape or boundary of objects.	Focusing on representing the entire volumetric scene, including color, lighting, and density.
Using 3D data (like point clouds or meshes) for training.	Using 2D images of a scene for training.
	Computationally intensive (particularly during training).

With the introduction of NeRF (Mildenhall et al. 2021), a completely new direction for novel-view synthesis has been established. Given a sparse set of images capturing an object of interest from different viewpoints, NeRF learns an implicit function to represent complex 3D shapes such that the novel poses of the object can be synthesized (Samavati & Soryani 2023). Meanwhile, NeRF was established on scene training, using sparse views of an object. As stated by Samavati et al. (2023), "the training and inference times of the original NeRF are high; therefore, instead of representing the entire scene as a single implicit field, Neural Sparse Voxel Fields (NSVF) (Liu et al. 2020) organize the scene as a sparse voxel octree and bound a set of implicit functions to its voxels. This speeds up rendering by a factor of ten. While NeRF only operates on static scenes, there are some other methods proposed to handle dynamic ones." This gives the opportunity to employ architectural-related works as it is independent of a massive, diverse training dataset. Figure. 1 shows the more advanced 3D reconstruction methods that NeRF could be categorized under *Novel-View Synthesis* class.

3D Reconstruction in Architecture vs. Other Fields

3D Reconstruction is a commonly utilized technique for simplifying the process of object observation from images and generating 3D meshes. By employing

this approach, any person can obtain a three-dimensional object without requiring sophisticated technologies or 3D scanners. Its reliability and precision have been established through extensive use to the extent that it has been integrated into various sectors such as medicine, robotics, and autonomous vehicles. However, the rationale for this transformation from 2D to 3D varies among academic fields. In computer vision, research on structure from motion (SFM) and multi-view stereo (MVS) has produced many compelling results, in particular accurate camera tracking and sparse reconstructions, and increasingly reconstruction of dense surfaces. However, much of this work was not motivated by real-time applications (Newcombe et al. 2011). Here, we will discuss the application of the 3D Reconstruction in robotics, medicine, and architecture.

In order to perform real-time data processing duties such as localization and grasping, robots must have an exhaustive understanding of their immediate environs. "Real-time 3D semantic re-construction is required in many robotics applications, such as autonomous navigation or grasping and manipulation. While a variety of well-known methods can reconstruct accurate 3D maps at real-time frame rates, the resulting point clouds contain no semantic-level understanding of the observed scenes. Hence, the problem of 3D semantic reconstruction is attracting increasing attention in the robotics research community. Recent methods not only generate a 3D point cloud map but also simultaneously assign a semantic label to each point in the cloud'', (C. Zhao et al. 2017). The utilization of semantic observation and reconstruction extends to numerous environmental attributes, including materials, due to its critical application in disciplines such as robotic grasping and rescue. According to Endres et al. (2013), "Manipulation robots, for example, require a detailed model of their workspace for collision-free motion planning, and aerial vehicles need detailed maps for localization and navigation.'' Moreover, in robotic systems, active vision involves adeptly adjusting the camera pose to gather the maximum amount of information from the scene. Existing methods that calculate the information yield for active robotic 3D Reconstruction focus on explicit 3D models like point clouds and voxels, using structure from motion (SfM). Using well-posed multi-view images, NeRF can provide an object's dense 3D reconstruction with favorable accuracy, overcoming the inherent limitations of SfM and depth fusion methods (Lee et al., 2022).

Medical images have been used in disease diagnosis and therapy since their discovery. Image processing techniques have been used to improve the quality of the images through tasks such as contrast enhancement and also in image analysis to aid in the interpretation of the images by clinicians (Ahishakiye et al. 2021). According to (Singh et al. 2020), 3D Deel Learning has multiple applications in 3D medical imaging such as *segmentation* (for brain lesions, skull stripping, etc.), *classification* (detecting dementia from MRI), *detection and localization* (detecting cerebral micro-hemorrhages in brain tissue), and *registration*. As Wang et al. (2021) explained in, "medical 3D reconstruction refers to computer processing of sampled single or sequence images from human organs by computerized tomography (CT), magnetic resonance imaging (MRI), and Ultrasonic imaging (UI). It includes data acquisition, pre-processing, point cloud splicing, and feature analysis. In medical images, 3D reconstruction of tissue sections, CT sections, and autoradiography

sections have become an important part of biomedical research. Through the restoration of the 3D surface shape of the organ, the human-computer interaction can be further used to carry out various realistic operations on the reconstructed object, such as rotation, scaling, and cutting.'' Architecture spans a broad spectrum of applications, from the detailed conservation of historical sites to the dynamic realms of parametric and algorithmic designs. When focusing on historic preservation, there is a critical emphasis on achieving utmost accuracy, ensuring every intricate detail is retained. In this context, computer vision emerges as a powerful toolset, offering sophisticated techniques to aid in the preservation and restoration of historical treasures. Leveraging advanced algorithms and methodologies, such as object classification and feature detection, architects and preservationists can analyze and interpret collections of data for a specific element or site. Through object classification, computer vision algorithms can identify and categorize a variety of architectural elements and features present within historical structures, ranging from decorated facades to intricate interior details. By automatically recognizing and cataloging these components, preservation efforts can be streamlined, allowing for more efficient documentation and analysis of the site's architectural heritage. Moreover, feature detection techniques enable the identification of distinctive characteristics within historical imagery and architectural drawings. By establishing correspondences between these features across different data sources, computer vision facilitates the creation of accurate digital reconstructions and 3D models, enabling architects to visualize and analyze the site's evolution over time. By harnessing the capabilities of computer vision in historic preservation work, architects can enhance their ability to document, analyze, and protect our cultural heritage. Through the precise identification and preservation of architectural elements, these advanced technologies can ensure that the legacy of our past remains intact for generations to come. Computers and algorithms altered the architectural design in three major fields:

- *Parametric and algorithmic design*
- *Remodeling and renovating existing buildings*
- *Expanding the building by adding newly designed forms and more*

One of the most notable uses of 3D reconstruction is its ability to enable quick and precise re-modeling and building enlargement. This technology is especially useful for projects that include adding contemporary additions to old or existing buildings. As an example, Figure. 2 presents two well-known instances of building extensions in which traditional buildings have been modernized with computer-aided design. Because 3D reconstruction produces accurate digital models of the existing structures, it assists designers in such situations. Using these precise models as a starting point helps designers and architects combine new ideas with the existing structure in an easy-to-use manner. All aspects of the existing building, including its size, geometrical features, and structural details, are taken into account thanks to the high degree of detail obtained through 3D reconstruc-tion. 3D reconstruction makes the design process more effective and efficient. It makes it possible to analyze and visualize in great detail how new components will interact with the current structure, which can help spot any problems and improve integration.

Figure 2. *Royal Ontario Museum in Toronto, by Daniel Libeskind, Coop Himmelb(l)au's Rooftop Extension on Falkestrasse in Vienna*

Sources: weburbanist.com, dezeen.com

This capacity is especially helpful in maintaining the original building's historical significance and architectural integrity while adding cutting-edge, practical modern additions. Conversely, in parametric design, the requirements fluctuate based on the design's intent and underlying concept. In the dynamic world of parametric design, which is pivotal for contemporary and envisioned future architectural endeavors, designers often find themselves in situations demanding the reshaping, augmentation, or modification of building components or facades. This could be for integrating human-interactive facades, sustainable facade systems, and more. In such scenarios, the foundational structure needs to be shapeable and open to edits. With the grand scale and intricate details associated with architectural designs, a densely detailed mesh comprising millions of elements becomes less manageable, if possible, to modify for subsequent stages. This issue underscores the importance of generating objects that can seamlessly evolve in pair with design requirements.

3D Reconstruction Effects in Architectural Education & Professional Field

Creativity in architectural design compositions can be viewed as an emergence of new forms and shapes or relationships between forms and shapes from which new concepts can be discovered (Reffat 2003). Architects and designers are thus interested in situations where distinctions matter. For this purpose, architects may use or modify existing forms as a basis for their designs. As Leimer et al. (2017) stated, "with online repositories for 3D models like 3D Warehouse becoming more prevalent and growing ever larger, new possibilities have opened up for both experienced and inexperienced users alike. These large collections of shapes can provide inspiration for designers or make it possible to synthesize new shapes by combining different parts from already existing shapes, which can be both easy to learn and a fast way of creating new shapes." However, architects are inspired by a

specific image and incorporate the concept into their design to produce a more inventive structure.

Current methodologies enable arduous manual form modification through the utilization of documents or images. Architects often draw inspiration from specific visual depictions or conceptual frameworks, which they then incorporate into their design paradigms to produce more innovative architectural structures. Conventional approaches have predominantly enabled the labor-intensive alteration of forms, frequently relying on manual modifications informed by visual imagery or documentary evidence. On the contrary, the emergence and widespread adoption of Machine Learning technologies have caused a profound transformation in the obligations and functionalities of architects. According to (Penney & Chen 2019), "traditionally, the relationship between computer architecture and ML has been relatively imbalanced, focusing on architectural optimizations to accelerate ML algorithms. In fact, the recent resurgence in AI research is, at least partly, attributed to improved processing capabilities". These emergent technologies facilitate rapid creativity and innovative solutions and accelerate the overall design lifecycle. Moreover, they introduce a forward-thinking approach that enables immediate alterations and the retracing of design decisions, thereby increasing efficiency and adaptability within the architectural process.

An outstanding attribute of these computational methods is their ability to algorithmically and iteratively manipulate geometry. Vale et al. (2003) stated, "engineering design is increasingly being influenced by the proliferation of computer tools in the design environment. CAD systems that previously supported human efforts through design representation and visualization are being supplemented with optimization algorithms that automate various routine design tasks." Instead of utilizing static, immutable designs, students are now capable of creating dynamic, adaptable models that can develop in reaction to various inputs or criteria. The utilization of structure from surface techniques empowers the fabrication of intricate and organic forms, whereas cladding techniques facilitate the seamless incorporation of recurring and modular design components. In addition, deformation permits the twisting, distorting, and transformation of simple shapes, which produces novel and complex design outcomes. As a result, designers may be required to interact with an existing form in numerous scenarios of the design computation process in order to create or redesign structures. A foundational framework is necessary for this to be modifiable throughout the design process. Professional practice, conversely, pertains to punctuality, expenditures, and precision. In circumstances where a redesign aims to modify, enlarge, or renovate the preexisting structure or to incorporate emerging technologies (such as interactive and smart facades, interiors, and so forth), the architect must possess the existing structure, which may be adapted to accommodate the new design. "The main reasons are that architecture design involves complex products, consisting of many distinct parts, and creating and editing these 3D models in 3D space is difficult. Hence, architectural design continues to be done on 2D drawings. However, for various construction applications, such as quantity surveying, inventory, construction, and visualization, it is necessary to convert the widely used 2D architectural drawings to accurate 3D models", according to (Lu et al. 2005). The requirement for a foundational framework that considers labor and

time generates a need for the automation of laborious and time-consuming duties using artificial intelligence and geometric algorithms. In conjunction with scholarly and professional developments in deep learning, the algorithm expands the limits of computational and geometric methodologies within architectural research.

Virtual Reality Integration with 3D Reconstruction in Architecture

The research method could be applied in developing accurate, optimized 3D mesh structures that integrate into VR applications, ensuring a seamless and realistic presentation of architectural designs. These immersive experiences transcend traditional static presentations, enabling users to explore, interact, and modify architectural elements in real time. This process offers a human-centric interactive experience in the VR setting, allowing users to seamlessly transition from a 2D representation to an editable and amendable 3D object within the virtual environment. This level of interactivity empowers users to manipulate architectural elements, experiment with different design configurations, and acquire practical insights into the design's spatial relationships and aesthetic qualities. These VR architectural presentations also serve as helpful tools for architects and designers, improving the design process by providing a platform for collaborative exploration and iterative refinement. Design decisions can be made in real time, informed by user feedback from virtual environment experiences.

Furthermore, the adaptability of this approach allows for the integration of additional features and functionalities, such as real-time lighting simulations, material selection, and environmental modeling. This augments the influential experience and gives architects and designers a comprehensive understanding of how their designs will perform in different lighting conditions and contexts. Finally, the integration of the NeRF and 3D optimization methods into VR architectural presentations represents a paradigm shift in architectural visualization, offering outstanding interactivity and versatility. By adopting this innovative approach, architects and designers can collaborate more effectively, make informed design decisions, and ultimately create spaces that resonate with users on a profoundly immersive level.

Evolution of the Computer Vision in Architectural Design

Computer vision plays a multifaceted role in architectural design, influencing different stages from initial conception to final presentation. Beginning with the design strategy phase, computer vision tools facilitate the transition from early sketches to precise lines, enabling architects to trans-late their creative vision into tangible plans more efficiently and accurately. One of the significant contributions of computer vision lies in 3D Reconstruction, which assists in creating detailed 3D models of architectural structures from images or point clouds. This technology allows architects to visualize their designs in a realistic manner and make informed decisions about spatial arrangements and aesthetics. Moreover, object detection and classification, particularly within the context of Building Information Modeling (BIM), revolutionize the design process by automating the identification and categorization of architectural elements and regions. By leveraging these capabilities,

both students and professional architects can explore multiple design alternatives efficiently, saving valuable time and resources that would otherwise be spent on manual labor. Furthermore, computer vision enhances the rendering and post-processing phases of architectural design by stream-lining the visualization process and improving the quality of final presentations. The accuracy and efficiency of design workflows are significantly enhanced through auto-detection and auto-classification functionalities, leading to more precise and visually compelling architectural representations. Therefore, the integration of computer vision technologies into architectural design practices not only empowers designers to unleash their creativity but also facilitates the generation of innovative solutions, streamlines workflow procedures, and eventually enhances the quality of architectural outcomes.

Conclusions

The integration of architectural principles with inspirations from philosophy, nature, and digital technology has ushered in a revolutionary period of architectural creativity characterized by cutting-edge innovation, precision, and ecological sustainability. In the future, continuous learning, adaptability, and cross-disciplinary collaboration will be essential to harnessing the full potential of these methodologies, anticipating a future in which architecture parallels our evolving global environment with an emphasis on sustainability. Deciphering three-dimensional data presents both opportunities and complexities within the 3D deep learning domain. The transition from 2D methodologies to 3D deep learning techniques, advantageous across multiple industries, exemplifies the field's dynamic evolution. This paper has explored various 3D reconstruction strategies, each offering distinct advantages for extracting depth from 2D images. With the rise of learning-driven 3D representations and the bridging potential of differentiated visualization, the future of this field appears bright. Given the complexity and immense potential of 3D data analysis, ongoing research, particularly in the field of architecture, is essential. As our journey to comprehend the 3D domain progresses, it becomes clear that the horizons for innovation and understanding are limitless. This promises a transformative future in architectural interactions with the world of three-dimensional deep learning.

References

Achlioptas P, Diamanti O, Mitliagkas I, Guibas L (2018) *Learning representations and generative models for 3d point clouds*. In International conference on machine learning (pp. 40–49). PMLR.

Ahishakiye E, Bastiaan Van Gijzen M, Tumwiine J, Wario R, Obungoloch J (2021) A survey on deep learning in medical image reconstruction. *Intelligent Medicine*, *1*(03), 118–127. (Publisher: Chinese Medical Association Publishing House Co., Ltd 69 Dongheyan Street …)

Aykin MD, Negahdaripour S (2016) Three-dimensional target reconstruction from multiple 2-d forward-scan sonar views by space carving. *IEEE Journal of Oceanic Engineering*, *42*(3), 574589. (Publisher: IEEE)

Ballé J, Laparra V, Simoncelli EP (2016) *End-to-end optimized image compression*. ArXiv preprint arXiv:1611.01704.

Banos R, Manzano-Agugliaro F, Montoya F, Gi C, Alcayde A, Gómez J (2011) Optimization methods applied to renewable and sustainable energy: A review. *Renewable and sustainable energy reviews*, *15*(4), 1753–1766. (Publisher: Elsevier)

Bernardini F, Mittleman J, Rushmeier H, Silva C, Taubin G (1999) The ball-pivoting algorithm for surface reconstruction. *IEEE transactions on visualization and computer graphics*, 5(4), 349–359. (Publisher: IEEE)

Bolitho M, Kazhdan M, Burns R, Hoppe H (2009) *Parallel poisson surface reconstruction*. In Advances in Visual Computing: 5th International Symposium, ISVC 2009, Las Vegas, NV, USA, November 30-December 2, 2009, Proceedings, Part I 5 (pp. 678–689). Springer.

Broadhurst A, Drummond TW, Cipolla R (2001) *A probabilistic framework for space carving*. In Proceedings eighth IEEE international conference on computer vision. ICCV 2001 (Vol. 1, pp. 388–393). IEEE.

Brock A, Lim T, Ritchie JM, Weston N (2016) *Generative and discriminative voxel modeling with convolutional neural networks*. arXiv preprint arXiv: 1608.04 236.

Bronstein MM, Bruna J, LeCun Y, Szlam A, Vandergheynst P (2017) Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, *34*(4), 18–42. (Publisher: IEEE)

Bylow E, Sturm J, Kerl C, Kahl F, Cremers D (2013) Real-time camera tracking and 3D reconstruction using signed distance functions. In *Robotics: Science and Systems* (Vol. 2, p. 2).

Chen J, Kira Z, Cho YK (2019) Deep learning approach to point cloud scene understanding for automated scan to 3D reconstruction. *Journal of Computing in Civil Engineering*, *33*(4), 04019027. (Publisher: American Society of Civil Engineers)

Chen W-C, Bouguet J-Y, Chu MH, Grzeszczuk R (2002) Light field mapping: Efficient representation and hardware rendering of surface light fields. *ACM Transactions on Graphics* (TOG), 21(3), 447–456. (Publisher: ACM New York, NY, USA)

Choy CB, Xu D, Gwak J, Chen K, Savarese S (2016) *3d-r2n2: A unified approach for single and multi-view 3d object reconstruction*. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14 (pp. 628–644). Springer.

Cole F, Genova K, Sud A, Vlasic D, Zhang Z (2021) *Differentiable surface rendering via non-differentiable sampling*. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 6088–6097).

Curless B, Levoy M (1996) *A volumetric method for building complex models from range images.* Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, 303–312. Retrieved 2023-07-17, from Conference Name: SIGGRAPH96: 23rd active Techniques ISBN: 9780897917469 Publisher: ACM) doi: 10.1145/237170.237269

Deng K, Liu A Zhu J-Y, Ramanan D (2022) *Depth-supervised nerf: Fewer views and faster training for free*. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12882–12891).

Digne J (2014) An analysis and implementation of a parallel ball pivoting algorithm. *Image Processing On Line*, *4*, 149–168.

Endres F, Hess J, Sturm J, Cremers D, Burgard W (2013) 3-D mapping with an RGB-D camera. *IEEE transactions on robotics*, *30*(1), 177–187. (Publisher: IEEE)

Estellers V, Scott M, Tew K, Soatto S (2015) *Robust poisson surface reconstruction*. In Scale Space and Variational Methods in Computer Vision: 5th International Conference,

SSVM 2015, Lège-Cap Ferret, France, May 31-June 4, 2015, Proceedings 5 (pp. 525–537). Springer.

Fan H, Su H, Guibas LJ (2017) *A point set generation network for 3d object reconstruction from a single image*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 605–613).

Gadelha M, Maji S, Wang R (2017) *3d shape induction from 2d views of multiple objects.* In 2017 International Conference on 3D Vision (3DV) (pp. 402–411). IEEE.

Gershun A (1939) The light field. *Journal of Mathematics and Physics*, *18*(1-4), 51–151.

Groueix T, Fisher M, Kim VG, Russell BC, Aubry M (2018) *A papier-mâché approach to learning 3d surface generation*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 216–224).

Jensen HW (1996) *Global illumination using photon maps*. In Rendering Techniques' 96: Proceedings of the Eurographics Workshop in Porto, Portugal, June 17–19, 1996 7 (pp. 21–30). Springer.

Ji S, Xu W, Yang M, Yu K (2012) 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 221–231. (Publisher: IEEE)

Jiang H, Li Y, Zhao H, Li X, Xu Y (2021) Parallel single-pixel imaging: A general method for direct–global separation and 3d shape reconstruction under strong global illumination. *International Journal of Computer Vision*, *129*, 1060–1086. (Publisher: Springer)

Jiang L, Shi S, Qi X, Jia J (2018) *Gal: Geometric adversarial loss for single-view 3d-object reconstruction*. In Proceedings of the European conference on computer vision (ECCV) (pp. 802–816).

Jiang Y, Ji D, Han Z, Zwicker M (2020) *Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1251–1261).

Jimenez Rezende D, Eslami S, Mohamed S, Battaglia P, Jaderberg M, Heess N (2016) *Unsu-pervised learning of 3d structure from images.* Advances in neural information processing systems, 29.

Kanazawa A, Tulsiani S, Efros AA, Malik J (2018) *Learning category-specific mesh reconstruction from image collections.* In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 371–386).

Kato H, Beker D, Morariu M, Ando T, Matsuoka T, Kehl W, Gaidon A (2020) *Differentiable rendering: A survey*. arXiv preprint arXiv:2006.12057.

Kazhdan M, Bolitho M, Hoppe H (2006) *Poisson surface reconstruction.* In Proceedings of the fourth Eurographics symposium on Geometry processing (Vol. 7, p. 0).

Kotary J, Fioretto F, Van Hentenryck P, Wilder B (2021) *End-to-end constrained optimization learning: A survey*. arXiv preprint arXiv:2103.16378.

Kutulakos KN, Seitz SM (2000) A theory of shape by space carving. *International journal of computer vision*, *38*, 199–218. (Publisher: Springer)

Landis H (2002) *Production-ready global illumination*. Siggraph course notes, 16(2002), 11. (Publisher: sn)

Lee S, Chen L, Wang J, Liniger A, Kumar S, Yu F (2022) Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields. *IEEE Robotics and Automation Letters*, *7*(4), 12070–12077. (Publisher: IEEE)

Leimer K, Gersthofer L, Wimmer M, Musialski P (2017) Relation-based parametrization and exploration of shape collections. *Computers & Graphics*, 67, 127–137. (MAG ID: 2737903340) doi: 10.1016/j.cag.2017.07.001

Levoy M, Hanrahan P (1996) *Light field rendering.* In Proceedings of the 23rd annual conference on Computer graphics and interactive techniques (pp. 31–42).

Li K, Pham T, Zhan H, Reid I (2018). *Efficient dense point cloud object reconstruction using deformation vector fields.* In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 497–513).

Li X, Wan W, Cheng X, Cui B (2010) *An improved Poisson surface reconstruction algorithm*. In 2010 International Conference on Audio, Language and Image Processing (pp. 1134–1138). IEEE.

Li Y, Zhao H, Jiang H, Li X (2022) *Projective Parallel Single-pixel Imaging to Overcome Global Illumination in 3D Structure Light Scanning*. In European Conference on Computer Vision (pp. 489–504). Springer.

Liao Y, Donne S, Geiger A (2018) *Deep marching cubes: Learning explicit surface representations*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2916–2925).

Lighthill MJ (1949) CIX. A Technique for rendering approximate solutions to physical problems uniformly valid. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *40*(311), 1179–1201. (Publisher: Taylor & Francis)

Liu S, Zhang Y, Peng S, Shi B, Pollefeys M, Cui Z (2020) Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2019–2028).

Long X, Lin C, Liu L, Liu Y Wang P, Theobalt C, … Wang W (2022) *Neural UDF: Learning Unsigned Distance Fields for Multi-view Reconstruction of Surfaces with Arbitrary Topologies.* arXiv. Retrieved 2023-05-20, from (arXiv: 2211.14173 [cs]) doi: 10.48550/arXiv.2211.14173

Long X, Lin C Liu, Lingjie L, Yuan W, Peng Theobalt C, … Wang W (2022) *NeuralUDF: Learning Unsigned Distance Fields for Multi-view Reconstruction of Surfaces with Arbitrary Topologies*. *C*ornell University - arXiv. (ARXIV_ID: 2211.14173 MAG ID: 4310285368 S2ID: 01ce9e 739a5b12f492b40cf56fa01bf80e1e6327) doi: 10.48550/arxiv.2211.14173

Loper MM, Black MJ (2014) OpenDR: *An approximate differentiable renderer*. In Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII 13 (pp. 154–169). Springer.

Lopes A, Brodlie K (2003) Improving the robustness and accuracy of the marching cubes algorithm for isosurfacing. *IEEE Transactions on Visualization and Computer Graphics*, *9*(1), 16–29. (Publisher: IEEE)

Loubet G, Holzschuch N, Jakob W (2019) Reparameterizing discontinuous integrands for differentiable rendering. *ACM Transactions on Graphics (TOG)*, *38*(6), 1–14. (Publisher: ACM New York, NY, USA)

Lu T, Tai C-L, Bao L, Su F, Cai S (2005) 3D reconstruction of detailed buildings from architectural drawings. *Computer-Aided Design and Applications*, *2*(1-4), 527–536. (Publisher: Taylor & Francis)

Maturana D, Scherer S (2015) *Voxnet: A 3d convolutional neural network for real-time object recognition*. In 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS*)* (pp. 922–928). IEEE.

Merrell P, Akbarzadeh A, Wang L, Mordohai P, Frahm J-M, Yang R, … Pollefeys M (2007) *Real-time visibility-based fusion of depth maps.* In 2007 IEEE 11th International Conference on Computer Vision (pp. 1–8). Ieee.

Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R (2021) Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, *65*(1), 99–106. (Publisher: ACM New York, NY, USA)

Moulon P, Monasse P, Perrot R, Marlet R (2017) *Openmvg: Open multiple view geometry*. In Reproducible Research in Pattern Recognition: First International Workshop,

RRPR 2016, Cancún, Mexico, December 4, 2016, Revised Selected Papers 1 (pp. 60–74). Springer.

Newcombe RA, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison AJ, … Fitzgibbon A (2011) *Kinectfusion: Real-time dense surface mapping and tracking.* In 2011 10th IEEE international symposium on mixed and augmented reality (pp. 127–136). Ieee.

Newman TS, Yi H (2006) A survey of the marching cubes algorithm. *Computers & Graphics*, *30*(5), 854–879. (Publisher: Elsevier)

Niemeyer M, Mescheder L, Oechsle M, Geiger A (2020) *Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3504–3515).

Osher S, Fedkiw R, Osher S, Fedkiw R (2003) *Constructing signed distance functions.* Level set methods and dynamic implicit surfaces, 63–74. (Publisher: Springer)

Pan J, Han X, Chen W, Tang J, Jia K (2019) *Deep Mesh Reconstruction From Single RGB Images via Topology Modification Networks.* In 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 9963–9972). Seoul, Korea (South): IEEE. Retrieved 2023-05-20, from doi: 10.1109/ICCV.2019.01006 Park, J. J., F019, January). DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation. arXiv. Retrieved 2023-05-19, from (arXiv:1901.05103 [cs]) doi: 10.48550/arXiv.1901.05103 Pennef machine learning applied to computer architecture design. arXiv preprint arXiv: 1909.12373.

Qi CR, Su H, Mo K, Guibas LJ (2017) *Pointnet: Deep learning on point sets for 3d classification and segmentation.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 652–660).

Reddy SS, Sandeep V, Jung, C-M (2017) *Review of stochastic optimization methods for smart grid*. Frontiers in Energy, 11, 197–209. (Publisher: Springer)

Reeves WT, Blau R (1985) Approximate and probabilistic algorithms for shading and rendering structured particle systems. *ACM siggraph computer graphics*, *19*(3), 313–322. (Publisher: ACM New York, NY, USA)

Reffat RM (2003) *Architectural Exploration and Creativity using Intelligent Design Agents.* eCAADe proceedings. (MAG ID: 2784370890 S2ID: 22e762ee0ef302f766b23325fe485803a0138ddd) doi: 10.52842/conf.ecaade.2003.181

Rhodin H, Spörri J, Katircioglu I, Constantin V, Meyer F, Müller E, … Fua P (2018. *Learning monocular 3d human pose estimation from multi-view images.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8437–8446).

Riegler G, Osman Ulusoy A, Geiger A (2017) *Octnet: Learning deep 3d representations at high resolutions.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3577–3586).

Samavati T, Soryani M (2023) *Deep learning-based 3D reconstruction: a survey.* Artificial Intelligence Review. Retrieved 2023-07-16, from doi: 10.1007/s10 462-023-10399-2

Schauer J, Nüchter A (2018) The peopleremover—removing dynamic objects from 3-d point cloud data by traversing a voxel occupancy grid. *IEEE robotics and automation letters*, *3*(3), 1679–1686. (Publisher: IEEE)

Schonberger JL, Frahm J-M (2016) *Structure-from-motion revisited.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4104–4113).

Seitz S, Curless B, Diebel J, Scharstein D, Szeliski R (2006) *A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms*. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06), 1, 519–528. Retrieved 2023-07-17, from (Conference Name: 2006 IEEE Compute cognition - Volume 1 (CVPR'06) ISBN: 9780769525976 Place: New York, NY, USA Publisher: IEEE) doi: 10.1109/CVPR.2006.19

Sillion FX, Arvo JR, Westin SH, Greenberg DP (1991) *A global illumination solution for general reflectance distributions.* In Proceedings of the 18th annual conference on Computer graphics and interactive techniques (pp. 187–196).

Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyás B (2020) 3D deep learning on medical images: a review. *Sensors*, *20*(18), 5097. (Publisher: MDPI)

Sitzmann V, Diamond S, Peng Y, Dun X, Boyd S, Heidrich W, … Wetzstein G (2018) End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, *37*(4), 1–13. (Publisher: ACM New York, NY, USA)

Slabaugh GG, Culbertson WB, Malzbender T, Stevens MR, Schafer RW (2004) Methods for volumetric reconstruction of visual scenes. *International Journal of Computer Vision*, *57*, 179–199. (Publisher: Springer)

Snavely N, Seitz SM, Szeliski R (2006) Photo tourism: exploring photo collections in 3D. *ACM SIGGRAPH 2006 Papers on - SIGGRAPH '06*, 835. Retrieved 2023-07-17, from (Conference Name: ACM S husetts Publisher: ACM Press) doi: 10.1145/1179352.1141964

Snow D, Viola P, Zabih R (2000) *Exact voxel occupancy with graph cuts*. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) (Vol. 1, pp. 345–352). IEEE.

Stelldinger P (2008) Topologically correct surface reconstruction using alpha shapes and relations to ball-pivoting. In 2008 19th International Conference on Pattern Recognition (pp. 1–4). IEEE.

Stutz D, Geiger A (2018) Learning 3d shape completion from laser scan data with weak supervision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1955–1964).

Su H, Maji S, Kalogerakis E, Learned-Miller E (2015) Multi-view convolutional neural networks for 3d shape recognition. In Proceedings of the IEEE international conference on computer vision (pp. 945–953).

Sumikura S, Shibuya M, Sakurada K (2022) OpenVSLAM: a versatile visual SLAM framework. *ACM SIGMultimedia Records*, *11*(4), 1–1. (Publisher: ACM New York, NY, USA)

Takikawa T, Litalien J, Yin K, Kreis K, Loop C, Nowrouzezahrai D, … Fidler S (2021) *Neural geometric level of detail: Real-time rendering with implicit 3d shapes.* In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11358–11367).

Tatarchenko M, Dosovitskiy A, Brox T (2016) *Multi-view 3d models from single images with a convolutional network.* In Computer Vision–ECCV 2016:14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14 (pp. 322–337). Springer.

Thomas H, Qi CR, Deschaud JE, Marcotegui B, Goulette F, Guibas LJ (2019) *Kpconv: Flexible and deformable convolution for point clouds*. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 6411–6420).

Uy MA, Lee GH (2018) *Pointnetvlad: Deep point cloud-based retrieval for large-scale place recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4470–4479).

Vale CA, Shea K et al. (2003) *A machine learning-based approach to accelerating computational design synthesis*. In DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm (pp. 183–184).

Velickovic P, Cucurull G, Casanova A, Romero A, Lio P, Bengio Y et al. (2017) Graph attention networks. *stat*, *1050*(20), 10–48550.

Wang J-J, Chen S-W, Shao J-Q, Gu X-W, Zhu H-S (2021) *Medical 3D reconstruction based on deep learning for healthcare*. In Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion (pp. 1–5).

Wang Y, Cai S, Li S-J, Liu Y, Guo Y, Li T, Cheng M-M (2018) *CubemapSLAM: A piecewise-pinhole monocular fisheye SLAM system.* In Asian Conference on Computer Vision (pp. 34–49). Springer.

Wang Y, Sun Y, Liu Z, Sarma SE, Bronstein MM, Solomon JM (2019) Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, *38*(5), 1–12. (Publisher: ACM New York, NY, USA)

Wu G, Masia B, Jarabo A, Zhang Y, Wang L, Dai Q, … Liu Y (2017) Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, *11*(7), 926–954. (Publisher: IEEE)

Wu W, Qi Z, Fuxin L (2019) *Pointconv: Deep convolutional networks on 3d point clouds.* In Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (pp. 9621–9630).

Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J (2015) *3d shapenets: A deep representation for volumetric shapes.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1912–1920).

Xie H, Yao H, Sun X, Zhou S, Zhang S (2019) *Pix2vox: Context-aware 3d reconstruction from single and multi-view images.* In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2690–2698).

Xu M, Li M, Xu W, Deng Z, Yang Y, Zhou K (2016) Interactive mechanism modeling from multi-view images. *ACM Transactions on Graphics (TOG)*, *35*(6), 1–13. (Publisher: ACM New York, NY, USA)

Yang G, Huang X, Hao Z, Liu M-Y, Belongie S, Hariharan B (2019) *Pointflow: 3d point cloud generation with continuous normalizing flows.* In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4541–4550).

Yang R et al. (2003) *Dealing with textureless regions and specular highlights-a progressive space carving scheme using a novel photo-consistency measure.* In Proceedings Ninth IEEE International Conference on Computer Vision (pp. 576–584). IEEE.

Zach C (2008) *Fast and high-quality fusion of depth maps*. In Proceedings of the international symposium on 3D data processing, visualization and transmission (3DPVT) (Vol. 1). Citeseer. (Issue: 2)

Zhang D, Lu X, Qin H, He Y (2020) Pointfilter: Point cloud filtering via encoder-decoder modeling. *IEEE Transactions on Visualization and Computer Graphics*, *27*(3), 2015–2027. (Publisher: IEEE)

Zhao B, Wu X, Cheng Z-Q, Liu H, Jie Z, Feng J (2018) *Multi-view image generation from a single-view*. In Proceedings of the 26th ACM international conference on Multimedia (pp. 383–391).

Zhao C, Sun L, Stolkin R (2017) *A fully end-to-end deep learning approach for real-time simultaneous 3D reconstruction and material recognition*. In 2017 18th International Conference on Advanced Robotics (ICAR) (pp. 75–82). IEEE.

Zhou Y, Tuzel O (2018) *Voxelnet: End-to-end learning for point cloud based 3d object detection.* In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4490–4499).

Zhou Y, Zhang W, Zhu J, Xu Z (2016) Feature-driven topology optimization method with signed distance function. *Computer Methods in Applied Mechanics and Engineering*, *310*, 1–32. (Publisher: Elsevier)

Zhu J, Wang L, Yang R, Davis JE, et al. (2010) Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps. *IEEE transactions on pattern analysis and machine Intelligence*, *33*(7), 1400–1414. (Publisher: IEEE)

Zollhöfer M, Dai A, Innmann M, Wu C, Stamminger M, Theobalt C, Nießner M (2015) Shading-based refinement on volumetric signed distance functions. *ACM Transactions on Graphics (ToG)*, *34*(4), 1–14. (Publisher: ACM New York, NY, USA)