Visual Intelligence

REVIEW Open Access

Recent advances in implicit representation-based 3D shape generation

Jia-Mu Sun[1,2], Tong Wu[1,2] and Lin Gao[1,2*]

Abstract

Various techniques have been developed and introduced to address the pressing need to create three-dimensional (3D) content for advanced applications such as virtual reality and augmented reality. However, the intricate nature of 3D shapes poses a greater challenge to their representation and generation than standard two-dimensional (2D) image data. Different types of representations have been proposed in the literature, including meshes, voxels and implicit functions. Implicit representations have attracted considerable interest from researchers due to the emergence of the radiance field representation, which allows the simultaneous reconstruction of both geometry and appearance. Subsequent work has successfully linked traditional signed distance fields to implicit representations, and more recently the triplane has offered the possibility of generating radiance fields using 2D content generators. Many articles have been published focusing on these particular areas of research. This paper provides a comprehensive analysis of recent studies on implicit representation-based 3D shape generation, classifying these studies based on the representation and generation architecture employed. The attributes of each representation are examined in detail. Potential avenues for future research in this area are also suggested.

Keywords: Generative models, 3D shape representations, Geometry learning, Deep learning

1 Introduction

In recent years, the demand for three-dimensional (3D) content has reached an unprecedented level as virtual reality (VR) and augmented reality (AR) applications have become increasingly influential. Fancy terms such as Metaverse and digital human are being created and used in different contexts. However, it is a challenge to acquire the vast amount of 3D content that is needed to build these applications. Traditional approaches to creating 3D shapes rely heavily on trained artists and are struggling to keep up with the growing demand. To solve this problem, various methods have been proposed to generate 3D shapes, making 3D content generation an active area of computer graphics and computer vision.

However, the inherent complexity and variety of 3D data makes 3D content generation a difficult task. Unlike two-dimensional (2D) data, which can be effectively represented by an array, a number of representations have been proposed for 3D content generation. These representations include meshes, voxels, point clouds, structures (or primitives), deformation-based representations, multi-view images, and implicit representations [1, 2]. Many methods and architectures for 3D content generation have been built on top of these representations. Traditionally, researchers have focused on explicit representations such as meshes, voxels, and point clouds [3–5] because they are easy to render and edit. With the rapid development of deep learning and neural networks, function-based implicit representations have become popular [6–9] since neural networks can be flawlessly transferred to an implicit function. It has been observed that these methods enhanced by deep learning outperform traditional methods. However, these methods omit the appearance of 3D shapes, and they often need abundant ground truth 3D

*Correspondence: gaolin@ict.ac.cn
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China

data. Led by the pioneering work [10], neural radiance fields (NeRFs) are rapidly gaining attention for their ability to learn and generate appearance along with geometry from just a few multi-view images [11, 12]. Furthermore, EG3D [13] shows the possibility of compressing the 3D representation of NeRF into three feature planes (triplanes). More recently, Dreamfusion [14] and a series of follow-up works have taken advantage of the power of 2D diffusion models [15] and generated NeRFs from multimodal conditions. These studies have contributed to the increasing popularity of 3D shape generation using implicit representation. Existing surveys [1, 2] usually involve generating implicit shapes along with other types of representations such as meshes and point clouds, and they are generally based on works published before 2022. However, recent developments in the above-mentioned methods for generating 3D content have led to numerous studies, achieving high-quality generation results. The wealth of work can also be confusing for researchers attempting to get involved. Therefore, a comprehensive survey of recent work is needed.

In this survey, we focus on recently proposed implicit representation-based 3D shape generation methods. We categorize the implicit representations actively used in the literature into three types: signed distance fields, radiance fields, and triplanes. In Sect. 2, we will first introduce these popular representations, and then we describe various architectures used to generate geometry from these representations. In Sect. 3, we will list and analyze works according to these representations and architectures. In Sect. 4, we present some open problems and future research directions and finally draw conclusions.

2 Background

In this section, we briefly introduce the preliminary knowledge of implicit 3D shape generation. Section 2.1 describes three different implicit representations: the signed distance field, radiance field, and triplane, and Sect. 2.2 covers some commonly used deep learning methods used to generate 3D data.

2.1 Implicit representation of 3D shapes

2.1.1 Signed distance fields (*SDFs*)

Signed distance fields (SDFs) are essentially functions defined in 3D space: $f(\boldsymbol{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$. The level set of this function $\mathcal{S} = \{\boldsymbol{s} | \boldsymbol{s} \in \mathbb{R}^3, f(\boldsymbol{s}) = 0\}$ is the surface of the underlying 3D geometry, and $|f(\boldsymbol{x})|$ for any other points represents the minimum distance from $\boldsymbol{x}$ to $\mathcal{S}$. The sign of $f(\boldsymbol{x})$ is positive if $\boldsymbol{x}$ lies inside $\mathcal{S}$ and negative otherwise. As a function, SDF is more flexible than the common explicit representations such as point clouds or meshes, and inherently allows topology manipulations such as constructive solid geometry (CSG) operations. Moreover, SDF allows the use of a technique known as "sphere tracing" [16], which can accelerate the rendering of path tracing. Owing to these properties, SDF is popular in several areas of computer graphics literature. SDF can be easily transformed to meshes using algorithms such as Marching Cubes [17], and this process is performed by means of deep marching tetrahedra (DMTet) [18]. DMTet can be used as a bridge between SDF and mesh, enabling much previous work such as neural rendering built on meshes to be transferred to SDF. We include this type of work in signed distance field-based 3D content generation, since the underlying representation is SDF.

2.1.2 Radiance fields (*RFs*)

Radiance fields (RFs) are a pair of functions: $c(\boldsymbol{x}) : \mathbb{R}^3 \rightarrow [0, 1]^3$ and $d(\boldsymbol{x}) : \mathbb{R}^3 \rightarrow [0, +\infty]$. c and d are the radiance color and density of a point, respectively. They can be rendered by volume rendering and ray marching algorithms [19]. RFs are often accompanied by a positional encoding $e(\boldsymbol{x}) = (\sin(2^0\pi\boldsymbol{x}, \cos(2^0\pi\boldsymbol{x}), \ldots, \sin(2^{L-1}\pi\boldsymbol{x}, \cos(2^{L-1}\pi\boldsymbol{x}))$, where L is a hyperparameter controlling the dimension of the embedding layer. Positional encoding is the key for RFs to reconstruct high-frequency features of the geometry. The use of the radiance field as a representation was introduced in Ref. [10]. Although recently proposed, it has gained surprising popularity due to its ability to accurately reconstruct 3D geometry from only a few sparse multi-view images. However, the density field used by vanilla RFs struggles to define a clear surface for the geometry, limiting the fidelity of RFs as a representation. To solve this problem, VolSDF [20] and NeuS [21] use SDFs as the geometry and propose algorithms to transfer SDF values to the weight of volume rendering. Although they use SDFs, we regard works built upon them as radiance field-based works because they preserve the volume rendering and positional encoding techniques of RFs.

2.1.3 Triplanes

Triplanes are three 2D feature planes, each of which is represented by a $N \times N \times C$, where N is the resolution of the feature planes, and C is the channel number of the feature planes. The three planes can be denoted by F_{xy}, F_{xz} and F_{yz} since they are placed perpendicular to each other in 3D space, aligning to xy, xz and yz planes. The rendering of triplane-based geometry also uses ray marching. In contrast to RFs, the sample points are directly projected to F_{xy}, F_{xz} and F_{yz} to sample the features via bilinear interpolation. Three sampled features are then concatenated and fed to a small multi-layer perceptron (MLP). The output of the MLP is often density or SDF values and color values, following the convention of RFs. Triplane was introduced by EG3D [13] in 2022, and the initial purpose of the triplane is shape generation. Given that triplanes and RFs often appear together, one can easily cat-

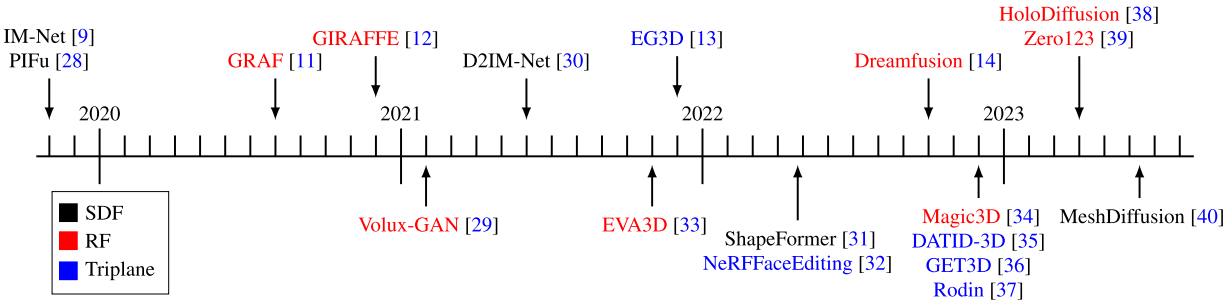

Figure 1 A brief timeline of implicit representation based 3D shape generation

egorize triplanes into another "trick" in RFs such as positional encoding, but we decide to list these triplanes independently as a representation because: 1) Triplanes can be directly generated from random noise or latent vectors by utilizing methods like StyleGAN [22]. 2) Triplanes can be transferred to other function-based implicit representations such as occupancy fields by modifying the head of the MLP. 3) A great deal of recent work is built based on triplanes, so it makes sense to create a category for them when reviewing. Note that the SDF-based generation method is only capable of generating shapes, while the RF- or triplane-based methods can generate both shapes and appearances simultaneously due to their ability to combine the color with the geometry in an implicit neural field.

2.2 Architectures for 3D shape generation

2.2.1 Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) [23] consist of a pair of neural networks. One is called the generator and the other is called the discriminator. The generator produces a data sample from a random noise or a condition, while the discriminator takes the sample and tries to distinguish it from the data taken from the real distribution. The generator and the discriminator are trained simultaneously, which is why they are called adversarial networks. GANs are flexible and can be used to generate both 2D and 3D data under various conditions.

2.2.2 Variational autoencoders (VAEs)

Variational autoencoders (VAEs) [24] are encoder-decoder structures, in which the encoder compresses data samples to a latent vector $\boldsymbol{z}$, and the decoder maps it back to the sample. A good feature of the VAEs is the space that contains the latent vectors. When VAE is trained, the latent space is naturally obtained and can be used as an embedding of the original data distribution, supporting operations such as interpolation. In this context, VAEs are more controllable than GANs.

2.2.3 Diffusion models (DMs)

Diffusion models (DMs) [25] generate data by assuming a noise of the same dimension as the data and iteratively "denoising" it using the same network. During training, Gaussian noise is added to the real data, and the network is supervised to recover the data from the noisy data by predicting the added noise. Since performance of the original model in the data space is slow, the latent diffusion model [15] runs the DM in the latent space and uses an encoder-decoder structure to link the latent space to the data space.

2.2.4 2D-to-3D models

2D-to-3D models are special kind of models introduced by Dreamfusion [14]. It takes advantage of the pre-trained large latent diffusion models such as stable diffusion [15] available on the Internet and RFs that can reconstruct 3D shapes from a few 2D images. These models typically utilize a kind of loss like the score distillation sampling (SDS) loss [14] to obtain the gradient from the frozen latent diffusion models (LDMs), using it to update the weight of the NeRF.

3 Generation of implicit shapes

In this section, we review in detail recent work on implicit representation-based 3D shape generation. We categorize the works according to the type of representation they use, including signed distance fields (Sect. 3.1), radiance fields (Sect. 3.2), and triplanes (Sect. 3.3). A brief timeline of the generation of implicit shapes is shown in Fig. 1.

3.1 Signed distance field-based shape generation

Signed distance fields implicitly represent shapes by predicting the distance values for sample points in the 3D space. The distance values' signs provide inside-outside information indicating which points are inside the surface and which are outside. These distance values can also be fed to the Marching Cubes algorithm [17] to extract an explicit triangle mesh. In Table 1, we briefly categorize shape-generating methods based on SDF according to the generation architecture and type of the generated results.

Table 1 Overview of works based on SDFs according to the generator and the generated results. GAN stands for generative adversarial network, VAE for variational autoencoders, and DM for diffusion models

Generated results	Generator architectures	
	GAN & VAE	DM
General objects	[6, 7, 9, 26]	[27–33]
Human bodies	[34, 35]	–
Human faces	[36]	–

3.1.1 GANs and VAEs

With the emergence of SDF, several pioneering works started to represent and generate 3D shapes using the SDF representation. Two concurrent works [8, 9] model the 3D space with occupancy grids where 1 represents inside and 0 represents outside. Later, a convolution-based occupancy network [37] predicts learnable features defined in the volume space or on multiple planes from an input point cloud, and the signed distance value of a sample point is determined by the interpolated features and a decoder network. Instead of modeling the whole shape at once, BAE-Net [38] and BSP-Net [26] learn to segment and reconstruct shape parts in an unsupervised manner based on IM-Net [9]. RIM-Net [39] also decomposes shapes into multiple parts but further predicts hierarchical structure. With the availability of fine-grained segmentation datasets [40, 41], researchers [42] have started to model shape geometry at the part level to capture more details where each part is represented by a latent vector and an occupancy decoder and when generating shapes, these parts are generated sequentially by an RNN network. SDF is a continuous version of occupancy grids that can represent more geometry details. DeepSDF [6] was the first to use SDF to represent a 3D shape and it uses an auto-decoder architecture to jointly optimize the latent vectors and the decoder network. To improve the generation and reconstruction quality, PIFu [34] and PIFuHD [35] propose extracting pixel-aligned features from human body images as extra outputs for the decoder network. DISN [7] shares a similar idea but focuses on general 3D objects. D2IM-Net [43] reconstructs and generates more geometric details by separating the signed distance field learning as a base signed distance field and a displacement value. SDF-StyleGAN [44] extends the 2D StyleGAN [22] to 3D and both global and local discriminators are deployed to ensure the generation quality. To reduce the flexibility of generated shapes, template-based methods [45, 46] have been proposed for modeling shapes in a specific category with a template signed distance field and a displacement field. With the discrete encoding [47] becoming popular in data compression, ShapeFormer [48] learns to generate 3D shapes from incomplete point cloud data by first encoding the incomplete data as incomplete discrete indices using VQ-VAE [47] and a transformer network [49] to fill the missing indices. AutoSDF [50] shares a similar idea but splits the whole 3D space into multiple 3D grids and encodes these grids as indices of the codebook in VQ-VAE [47]. A transformer network later takes the indices sequence as input and generates them auto-regressively. AutoSDF [50] allows not only the completion of shapes from incomplete data, but also the generation of random shapes. Apart from geometry generation, TextureFields [51] was the first to explore texture generation based on image and shape conditions, but it can synthesize only the rendered results of a given shape because both geometry and texture are implicit. To obtain explicit geometry and texture, DVR [52] represents geometry and texture with a single latent vector and uses an occupancy function and a texture function to predict the occupancy value and texture color for a sample point. AUV-Net [53] moves a step forward to learn an aligned UV parameterization network for shapes in the same category to allow seamless texture generation and transfer.

3.1.2 Diffusion

With recent advances in generative modeling with diffusion models, LION [27] applies the diffusion model to the 3D domain but takes point clouds as its 3D representation. MeshDiffusion [28] was the first to extend the diffusion model to an implicit representation. It represents the 3D shape with the Deep Marching Tetrahedra [18]. Li et. al. [30] proposed encoding local patches of 3D shapes into the voxel grids and training the diffusion model on a 3D grid. SDF-Diffusion [29] reduces the learning difficulty by first training a diffusion model on a low-resolution grid and performing a patch-based super-resolution to introduce additional geometric details. NeuralWaveletDiffusion [31] and NeuralWaveletDiffusion++ [32] instead convert the 3D signed distance volume into coefficient volumes using multi-scale wavelet decomposition. The diffusion model first generates a coarse coefficient volume and a detailed predictor models the geometric details. LAS-Diffusion [33] also uses a coarse-to-fine generation paradigm. It first trains an occupancy diffusion network to generate a sparse voxel grid and subdivide it into a voxel grid with higher resolution. Later, an SDF diffusion network is optimized to generate local details. Diffusion-SDF [54] first encodes shapes into triplane features and further compresses them into a compact latent vector. The diffusion model learns to generate 3D shapes by generating the latent vector. As regular grids and global latent code are proven less expressive in geometric modeling, 3DShape2VecSet [55] proposes representing a 3D shape with a set of latent vectors distributed irregularly in the 3D space. The feature of a sample point used to predict the occupancy value is determined by querying features from the set of latent vectors with a cross-attention layer. The diffusion model takes multiple sets of latent vectors

as training data to generate a plausible set of latent vectors, which is decoded as a 3D shape. HyperDiffusion [56] overfits each 3D shape in a training set with an occupancy network and represents one with the optimized parameters in the occupancy network. Later, the parameters are taken as the training data for the diffusion model, which also means that the diffusion model operates in a "hyper" space.

3.1.3 Summary

SDF is a fundamental implicit representation of 3D shapes. Many works take SDF as its representation for geometry reconstruction and generation tasks. One advantage of SDF is that it provides clear inside-outside information that can be used to extract an explicit surface. This property can be well integrated into the geometry reconstruction pipeline. However, such a representation also has limitations. For example, it is difficult to represent thin structures or open surfaces using the SDF representation. Hence, introducing a more flexible and general representation like unsigned distance field [57] is a possible direction.

3.2 Radiance field-based shape generation

Radiance fields model the appearance and geometry of 3D shapes by using volume rendering (ray marching) algorithms and positional encoding [10]. This technique can efficiently reconstruct geometry given only a few multi-view posed images. To achieve higher geometry quality, NeuS [21] and VolSDF [20] extend radiance fields by using SDF instead of density as the geometry. In Table 2, we briefly categorize shape generation methods based on NeRF according to the generation architecture and the type of results generated.

3.2.1 GANs and VAEs

The first attempt to generate radiance fields is GRAF [11]. The generator of the GRAF is simply a "conditional" NeRF, which takes a random noise in addition to the positional encoding to render a random patch of the full image, and the patch is fed into the discriminator. NeRF-VAE [58] bases its generative model on VAE and uses multiple scenes to train the VAE. On the other hand, pi-GAN [59] replaces the activation function with trigonometric functions as proposed by SIREN [97], while avoiding convolution-based networks. GIRAFFE [12] extends GRAF by implementing the disentanglement of geometry and appearance features and by considering the scene in a compositional manner.

Wang et al. [60] utilized template and deformation fields on geometry to control the shape generation. VolumeGAN [61] introduces a feature volume and uses volume rendering to map it into the image space. GIRRAFFE HD [62] further improves the resolution of the generated image by using the super-resolution module [22]. Xu et al. [63] proposed another extension to the GRAF. To improve the fidelity of the generated geometry, they added a progressive sampling strategy to the GRAF. ShadeGAN [64] uses the consistency under multiple lighting conditions as a further constraint on the generation of shadows. StyleNeRF adopts the super-resolution of StyleGAN [22] to improve 3D consistency. This method uses a novel regularization loss and upsampler. StyleSDF [36] replaces the original density field of NeRF with SDF, and simultaneously renders a low-resolution image and a 2D feature map. The feature map is transferred to a high-resolution image with the 2D generator. Persistent Nature [93] participates in the terrain with a grid and uses an upsampler to generate fine-grained geometry. Discoscene [94] generates large scenes from a layout prior that consists of labeled bounding boxes and generates radiance fields in the boxes. GRAM [65] combines the primitive-based method and radiance fields, generating multiple manifolds and their radiance. Volume rendering is modified by directly integrating the radiance of these manifolds. GVP [66] is based on the same idea as GRAM, predicting multiple primitives with radiance fields defined in them. Apart from the works above that focus on generating "general" 3D shapes, generating human faces or bodies (or called "avatars") is also an active topic. MultiNeuS [84] attempts to directly generate 3D heads represented by SDF field of NeuS. Tewari et al. [85] further disentangled the face geometry and appearance by predicting a deformation field and an appearance network. Tang et al. [86] used an explicit parametric face model for better control of the generated faces. Volux-GAN [87] incorporates lighting in 3D face synthesis by using an environment map and decomposing the material, achieving relighting in the synthesized models. AnifaceGAN [88] generates a movable 3D face, using different codes to generate template and deformation fields and an imitation loss. EVA3D [82] introduces the SMPL human body prior, segments the human body into multiple bounding boxes, and subsequently generates radiance fields inside them. MetaHead [89] and GANHead [90] introduce additional priors such as semantic labels and FLAME representations to generate 3D human heads. GeneFace [91] and GeneFace++ [92] control talking faces directly with audio by controlling facial landmarks, and then the landmarks are used to control 3D faces.

Table 2 Overview of works based on NeRFs according to the generator and the generated results

Generated results	Generator architectures		
	GAN & VAE	DM	2D-to-3D
General objects	[11, 12, 36, 58–66]	[67–70]	[14, 71–81]
Human bodies	[82]	–	[83]
Human faces	[84–92]	–	–
Scenes	[93, 94]	–	[95, 96]

3.2.2 *Diffusion models*

Recently, diffusion models have rapidly gained popularity since the proposal of LDM [15]. However, it is not easy to directly apply DMs to generate radiance fields: as a function, it is difficult to directly add noise to the RF. One of the methods for applying DMs is to use voxel-based radiance fields [67]. It utilizes the base NeRF model following the voxel-grid-based representations such as DVGO [98]. Voxel-based radiance field representations are fast to render, and 3D UNets can be used to implement the diffusion process. Holodiffusion [68] also adopts voxel grid representation for diffusion. They use a single 2D image as a prior and utilize a diffusion model to generate 3D shapes. Another type of diffusion model for RFs involves diffusion on latent vectors and the use of a conditional NeRF that maps a latent code to 3D shapes. 3D-CLFusion [69] combines the CLIP encoder and diffusion model. This approach enables the model to use both images and text as input conditions. Neuralfield-LDM [70] modifies the diffusion model to be "hierarchical", taking 1D, 2D, and 3D latent features simultaneously. It also trains an autoencoder to obtain these features from NeRFs.

3.2.3 2*D-to-3D*

2D-to-3D models are a "special" category of models, which basically take advantage of the publicly available large pretrained diffusion models. All of these methods can inherit all the features of the large DMs such as multi-modal information and generate almost all types of objects. The pioneering work of 2D-to-3D models is Dreamfusion [14]. To distill 2D generation models to 3D, Dreamfusion needs to train a NeRF for every generated object by minimizing the SDS loss. The SDS loss takes the gradient of the U-Net out of the original diffusion loss, which is proven to be effective. However, multiple problems associated with SDS loss have been observed: 1) The method needs to optimize a NeRF every time it generates a shape, limiting the generation speed. 2) The underlying stable diffusion does not have pose priors, introducing ambiguity to the generated geometry. This may cause multiple artifacts such as the "Janus Problem", low-detailed geometry, or over-smooth geometry. 3) The appearance problem. In order to make the network with SDS loss converge, the guidance weight of the SDS is set high. This can lead to overly saturated colors in the generated shapes, making them look "cartoonish" and unrealistic. To address these problems, several improvements have been proposed. Latent-NeRF [71] uses a feature space NeRF instead of an image space to better connect stable-diffusion [15], which is the state-of-the-art guidance model used in 2D-to-3D methods. Perp-Neg [99] attempts to solve the "Janus problem" by using negative prompts in the diffusion model to make it faithfully produce images with desired views. SJC [72] has proposed another loss term that is similar but not identical to SDS and provides a clearer deduction of the loss. ProlificDreamer [73] replaces the SDS loss with the variational score distillation (VSD) loss and proves that the VSD loss is more generic and produces high-quality shapes. This modified version of SDS loss can partially solve the low-detail or over-smooth problem of dreamfusion. Magic3D [74] extends Dreamfusion to a two-stage coarse-to-fine approach using a mesh prior, improving the generation quality. Fantasia3D [75] also leverages a mesh prior and a two-stage pipeline, and further decomposes the material into PBR components. The two-stage methods can serve as a solution to the unrealistic appearance problem mentioned earlier, as they can individually optimize the appearance, reducing the need for 3D guidance. It can also potentially solve the geometry problem since the extracted mesh can serve as a template and a strong prior, making it easy for the diffusion model of refinement stages to guide optimization. Apart from the pure text-guided version, a single image or a few images' prior conditions can be applied. NeRDi [76] generates 3D shapes from diffusion prior and single image, but it relies on an "inverse process" by narrowing priors from visual cues and textual descriptions. Dream3D [77] uses both CLIP and diffusion model priors for generation. Dreambooth3D [78] uses a three-stage strategy that combines text-to-image and text-to-3D methods to gradually refine the generated NeRF. Zero-1-to-3 [79] tries to solve the ambiguity pose problem by controlling the camera pose of generated views via diffusion models. Make-It-3D [100] also uses a two-step strategy: First, it transforms the single image with an estimated depth predicted by off-the-shelf methods into a radiance field and then uses a diffusion model prior to refine the geometry. 3DFuse [80] improves the 3D consistency of generated shapes by feeding the diffusion model with a generated depth map. Apart from generating a single shape, scenes containing multiple shapes can be generated by 2D-to-3D methods. Po and Wetzstein [95] considered the text input and bounding boxes of multiple objects at the same time. The bounding boxes are used as masks in rendering, and every object is "merged" using the masks, and the whole image is used to compute the SDS loss. CompoNeRF [96] also utilizes a bounding-box-based scene composition convention, but it applies SDS loss to both the local (inside bounding box) and global geometry. The local radiance fields are "projected" to a global MLP in the joint training process. Other works have focused on generating dynamic scenes or human avatars using 2D-to-3D methods. DreamTime [81] generates dynamic scenes using timestep sampling with non-increasing functions when optimizing NeRF from the SDS. DreamAvatar [83] utilizes SMPL [101] parameters and text input, and uses SDS loss for both canonical space and observation space to generate a NeRF model.

3.2.4 Summary

The NeRF algorithm was the origin of the recent "explosion" of 3D shape generation work. It is easy to see that NeRF can fuse the geometry and appearance of objects into neural networks while preserving quality. Due to its simplicity, one can combine NeRF with various generator architectures and obtain decent results. However, NeRF also has some disadvantages. Even with multiple acceleration methods such as instant-ngp [102], it is still difficult to render NeRF completely as a mesh in real time, which affects the performance of various 2D-to-3D generation methods. The all-MLP architecture of NeRF also poses challenges in the case of editing, filtering, and post-processing. This difficulty also prevents subtle control when generating them. Finally, the relatively high dimensionality (5D) of the NeRF input can cause artifacts like floaters due to overfitting.

3.3 Triplane-based shape generation

As a representation, triplanes are newly introduced by EG3D [13], which is dedicated to the generation of high quality human head geometry. EG3D proves that the 3D data of radiance fields can be effectively compressed into three 2D feature maps, which can be directly generated by StyleGAN [22]. During rendering, three features from the maps are sampled from feature planes, concatenated, and fed into downstream networks. Exploring the potential of triplane representation has recently been a popular topic. In Table 3, we briefly categorize shape-generating methods based on triplanes according to the generating architecture and types of the generated results.

3.3.1 GANs, VAEs and 2D-to-3D

In the pioneering work on triplanes, EG3D [13] uses a StyleGAN [22] structure to generate the triplane and standard NeRF-like volume rendering techniques. Noguchi [115] extends EG3D and its GAN+triplane method to articulated humans. Avatargen [116] leverages SMPL [101] prior to generate controllable humans. EpiGRAF [103] removes the upsampler of EG3D and trains the GAN in patches to improve the fidelity of generation. IDE-3D [118] extends EG3D by utilizing different geometry and texture codes, and employs a sophisticated generator, encoder, and GAN inversion techniques. NeRFFaceEditing [119] further enhances IDE-3D. This approach enables fine-grained editing of generation results by utilizing appearance codes and semantic masks. DATID-3D [120] aims to transfer EG3D to another domain, e.g., animation. They use a pre-trained text-to-image model to generate a new dataset and use the refined dataset to transfer the underlying EG3D model. PODIA-3D [121] extends DATID-3D. They modify the diffusion model to make it pose-aware and use a debiasing module based on text. GET3D [104] generates a triplane over a GAN and then replaces the MLP to make it predict texture values, and the geometry is generated via a tetrahedron-based proxy mesh. Finally, they use DMTet [18] to extract triangular mesh from the SDF. SinGRAF [130] generates a 3D shape from the pattern of a specific scene, providing a single image of the scene. Next3D [122] extends the EG3D to generate animated faces. It uses two triplanes, one of which is used to deform the static geometry. Moreover, PV3D [123] generates dynamic videos. It extends triplanes, and separates appearance codes and motion codes, ensuring consistent motion. MAV3D [105] extends 2D-to-3D generation to dynamic scenes. They use the "hexplane" representation, incorporating the time axis. SDS loss is applied to both the static image and the dynamic video. TAPS3D [106] extends GET3D to allow text-to-3D generation via a caption generation module and CLIP. Geometry and texture are modeled with different triplanes. Skorokhodov et al. [107] considered arbitrary cameras and utilized depth priors. This approach can generate more diverse and challenging datasets such as ImageNet [131]. LumiGAN [124] generates geometry and albedo, specular tint, and visibility at the same time using triplane representation, and uses SH light to render the head instead of using the original NeRF method. NeRFFaceLighting [125] uses separate shading, geometry and albedo triplanes, in which the shading triplane is conditioned on SH lighting. The lighting and rendering are separately fed into discriminators. A regularization method is used to enhance generalizability of the algorithm. PanoHead [126] generates 3D human heads from 360° full head images, using a self-adaptive image alignment and a tri-grid volume to solve the "mirrored face" artifact of EG3D. Head3D [127] utilizes a teacher-student distillation technique and a dual-discriminator structure to solve the front-back gap for full-head generation present in EG3D-based methods. GINA-3D [108] decouples representation learning and generation, and uses VAE to map input images to latent feature represented by triplanes using quantization, cross-attention, and neural rendering. Trevithick et al [128] generated 3D heads with a single image prior at real-time speed, eliminating the costly generator at inference time. Additionally, they used encoders and Vit modules to generate triplanes. AG3D [117] separately generates canonical humans and poses (via deformation), and uses multiple discriminators of normal and rendered images (there are also different discriminators of

Table 3 Overview of works based on triplanes according to the generator and the generated results

Generated results	Generator architectures	
	GAN & VAE & 2D-to-3D	DM
General objects	[103–109]	[110–114]
Human bodies	[115–117]	–
Human faces	[13, 118–128]	[129]
Scenes	[130]	–

whole body and face). Zhu et al [109] based their method on GET3D with the aim of applying the trained model to another domain using silhouette images.

3.3.2 Diffusion models

Recently, diffusion models have received considerable attention because of their ability to generate high-quality 2D data. On the other hand, triplanes can compress 3D data into 2D data, making the combination of DMs and triplanes natural. RenderDiffusion [110] uses the diffusion model as the backbone and a single image as a condition. For every denoising step in the diffusion model, the image is encoded into a triplane and rendered back to a denoised image via volume rendering. Rodin [129] is another pioneering study that proposes a roll-out diffusion network that can perform 3D-aware diffusion and take advantage of multi-modal conditions. 3DGen [111] is an extension of Rodin that uses a VAE to obtain a latent space and a diffusion model to generate latent features. NerfDiff [112] uses a camera-aligned triplane to solve the ambiguity in depth. The rendered images are fed from the generated triplane back into the diffusion model to improve the generation quality. SSDNeRF [113] directly applies the diffusion model to triplane representations. They jointly learn the diffusion model and a decoder that can render the triplane into a NeRF, and the joint learning process enables single-view reconstruction. Gu et al. [114] combined components of VAE, GAN, and diffusion models, using a GAN to learn a latent code triplane and train a diffusion model on this triplane. The model can use both "condition" via the encoder and "guidance" via the diffusion model.

3.3.3 Summary

In contrast to the 3D nature of SDF and NeRF, the triplane compresses 3D data into three 2D planes and proves that these planes contain most of the information needed to reconstruct high-quality 3D shapes. This approach enables various methods to leverage 2D generators for 3D data, increasing generation speed and simplifying the design of the pipeline. However, in addition to the high quality of generation in the areas of human faces and avatars, which come with strong priors, the quality of general objects generated by triplanes seems slightly lower. There is also a lack of work dedicated to the generation of large-scale and multi-object scenes based on triplanes. It may take more time for researchers to realize the full potential of the triplane as an individual representation.

4 Discussion

After reviewing recent work on implicit representation-based 3D shape generation, we will now discuss some of the open problems and future directions in this area.

4.1 3D shape generation with higher quality

Implicit representations of 3D shapes often utilize a learned function to cover all the details of the geometry. However, this pipeline still struggles to generate high-fidelity fine-grained geometry. This is caused by both the limitations of the representation itself (the low-frequency pass feature of the MLP network and the inherent ambiguity of implicit representation) and the design of the architecture (the difficulty of designing the discriminator of the GAN, the "blurry" generation of VAE and the high computational cost of diffusion). Despite the problem of geometric quality, the appearance of the generated shapes also needs improvement. Radiance fields or 2D-to-3D generation architectures seem to be good choices for generating appearances along with geometry, but methods to thoroughly control the appearance of generation are still lacking.

4.2 Faster 3D shape generation

Radiance fields have various advantages as a representation. However, the rendering speed of RFs is not very satisfactory: it takes minutes to render a view of vanilla NeRF. In terms of generation speed, early work directly generating RFs or SDFs by using GAN or VAE can perform tasks at a relatively high speed, but the quality is undesirable. Recent work that is based on a 2D-to-3D generation pipeline usually needs to optimize a radiance field for each generated shape, taking approximately 30 minutes for a single shape. The use of three planes may provide a good balance between speed and quality, but rendering triplane-based shapes also rquires ray marching at a high cost. Generating and rendering implicit 3D shapes at real-time speed is still an open problem.

4.3 3D shape generation on a larger scale

The difficulty of generating large-scale implicit 3D shapes is twofold: 1) Implicit representation is obviously not a satisfactory choice for large scenes. They often require some sort of "range for variables" to make the function easier to learn, making it difficult to balance between the range and the scene scale. 2) The generation pipeline for large scenes is difficult to design, since it requires a holistic understanding of the scene, preventing access to the usual local patch-based method. It is useful to generate large scenes since downstream applications such as robotics or autonomous driving require this type of scene. Splitting the large scenes into smaller scenes [96] may be a solution, and we believe that there is research to be done in this area.

4.4 Combination with other representations

This survey focuses on implicit representations, but other representations such as mesh, point cloud, or structure-based and "procedural" [132] representations also have significant advantages. Recently, several studies such as

Neumesh [133] or DE-NeRF [134] have combined implicit and explicit representations to improve editing. These works requires a mesh for input, which can be obtained by off-the-shelf reconstruction methods [21] or from priors such as SMPL mesh [101]. This combination can provide implicit representations with topological priors, and utilize both higher geometry quality and better local editing. In the field of 3D shape generation, works like GET3D [104] also try to utilize traditional "mesh and texture" for generation and achieve good performance by transforming implicit representations to explicit mesh using differentiable middlewares such as deep marching tetrahedra [18]. In general, the combination of multiple representations for 3D content creation is a topic that is well worth exploration.

5 Conclusion

This survey has reviewed recent advances in 3D shape generation methods based on implicit representations. We begin with an introduction to the most commonly used implicit representations and generation architectures. We then review the recent work on implicit representation based 3D shape generation in detail. We categorize these studies according to the type of 3D representation they use, including signed distance fields, radiance fields, and triplanes. We have also included a brief timeline of the development of 3D shape generation based on implicit representations and highlighted key work in the literature. Finally, we discuss some of the aspects of current work that need to be improved in future research. It is hoped that this survey will provide some insights for other researchers and inspire future work in this area.

Acknowledgements
The authors thank Yu-Jie Yuan for his help in the literature collection and suggestions for improving this paper.

Funding
This work was supported by National Natural Science Foundation of China (No. 62322210), Beijing Municipal Natural Science Foundation for Distinguished Young Scholars (No. JQ21013), and Beijing Municipal Science and Technology Commission (No. Z231100005923031).

Abbreviations
AR, augmented reality; CSG, constructive solid geometry; GAN, generative adversial network; (L)DM, (latent) diffusion model; MLP, multi layer perceptron; (Ne)RF, (neural) radiance field; RNN, recurrent neural network; SDF, signed distance field; SDS, score distillation sampling; VAE, variational autoencoder; VR, virtual reality.

Data availability
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Author contributions
All authors contributed in the literature collect and review process. JS and TW perpared the manuscript, and then all authors participated in the comment and modification of the paper. All authors read and approved the final manuscript.

Received: 5 September 2023 Revised: 7 March 2024
Accepted: 10 March 2024 Published online: 25 March 2024

References

1. Xu, Q., Mu, T., & Yang, Y. (2023). A survey of deep learning-based 3D shape generation. *Computational Visual Media*, *9*(3), 407–442.
2. Xiao, Y., Lai, Y., Zhang, F., Li, C., & Gao, L. (2020). A survey on deep geometry learning: from a representation perspective. *Computational Visual Media*, *6*(2), 113–133.
3. Wu, J., Zhang, C., Xue, T., Freeman, B., & Tenenbaum, J. (2016). Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In D. Lee, M. Sugiyama, U. Luxburg, et al. (Eds.), *Proceedings of the 29th international conference on neural information processing systems*. (pp. 82–90). Red Hook: Curran Associates.
4. Fan, H., Su, H., & Guibas, L. J. (2017). A point set generation network for 3D object reconstruction from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2463–2471). Piscataway: IEEE.
5. Tan, Q., Gao, L., Lai, Y., & Xia, S. (2018). Variational autoencoders for deforming 3D mesh models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5841–5850). Piscataway: IEEE.
6. Park, J. J., Florence, P., Straub, J., Newcombe, R. A., & Lovegrove, S. (2019). DeepSDF: learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 165–174). Piscataway: IEEE.
7. Xu, Q., Wang, W., Ceylan, D., Mech, R., & Neumann, U. (2019). DISN: deep implicit surface network for high-quality single-view 3D reconstruction. In H. Wallach, H. Larochelle, A. Beygelzimer, et al. (Eds.), *Proceedings of the 32nd international conference on neural information processing systems*. (pp. 490–500). Red Hook: Curran Associates.
8. Mescheder, L. M., Oechsle, M., Niemeyer, M., Nowozin, S., & Geiger, A. (2019). Occupancy networks: learning 3D reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4460–4470). Piscataway: IEEE.
9. Chen, Z., & Zhang, H. (2019). Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5939–5948). Piscataway: IEEE.
10. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: representing scenes as neural radiance fields for view synthesis. In A. Vedaldi, H. Bischof, T. Brox, et al. (Eds.), *Proceedings of the 17th European conference on computer vision* (pp. 405–421). Cham: Springer.
11. Schwarz, K., Liao, Y., Niemeyer, M., & Geiger, A. (2020). GRAF: generative radiance fields for 3D-aware image synthesis. In H. Larochelle, M. Ranzato, R. Hadsell, et al. (Eds.), *Proceedings of the 33rd international conference on neural information processing systems*. (pp. 1254–1267). Red Hook: Curran Associates.
12. Niemeyer, M., & Geiger, A. (2021). GIRAFFE: representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11453–11464). Piscataway: IEEE.
13. Chan, E. R., Lin, C. Z., Chan, M. A., Nagano, K., Pan, B., Mello, S. D., et al. (2022). Efficient geometry-aware 3D generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16102–16112). Piscataway: IEEE.
14. Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2023). DreamFusion: text-to-3D using 2D diffusion. In *Proceedings of the 11th international conference on learning representations*. Retrieved January 25, 2024, from https://openreview.net/pdf?id=FjNys5c7VyY.
15. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10674–10685). Piscataway: IEEE.
16. Hart, J. C. (1996). Sphere tracing: a geometric method for the antialiased ray tracing of implicit surfaces. *Visual Computing*, *12*(10), 527–545.

17. Lorensen, W. E., & Cline, H. E. (1987). Marching cubes: a high resolution 3D surface construction algorithm. In *Proceedings of the 14th annual conference on computer graphics and interactive techniques* (pp. 163–169). New York: ACM.
18. Shen, T., Gao, J., Yin, K., Liu, M., & Fidler, S. (2021). Deep marching tetrahedra: a hybrid representation for high-resolution 3D shape synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, et al. (Eds.), *Proceedings of the 34th international conference on neural information processing systems*. (pp. 6087–6101). Red Hook: Curran Associates.
19. Weiskopf, D. (2007). *GPU-based interactive visualization techniques*. Berlin: Springer.
20. Yariv, L., Gu, J., Kasten, Y., & Lipman, Y. (2021). Volume rendering of neural implicit surfaces. In M. Ranzato, A. Beygelzimer, Y. Dauphin, et al. (Eds.), *Proceedings of the 34th international conference on neural information processing systems*. (pp. 4805–4815). Red Hook: Curran Associates.
21. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., & Wang, W. (2021). NeuS: learning neural implicit surfaces by volume rendering for multi-view reconstruction. In M. Ranzato, A. Beygelzimer, Y. Dauphin, et al. (Eds.), *Proceedings of the 34th international conference on neural information processing systems*. (pp. 27171–27183). Red Hook: Curran Associates.
22. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4401–4410). Piscataway: IEEE.
23. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, et al. (Eds.), *Proceedings of the 27th international conference on neural information processing systems*. (pp. 2672–2680). Red Hook: Curran Associates.
24. Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. In *Proceedings of the 2nd international conference on learning representations*. Retrieved February 25, 2024, from https://iclr.cc/archive/2014/conference-proceedings/.
25. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd international conference on machine learning* (pp. 2256–2265). Stroudsburg: International Machine Learning Society.
26. Chen, Z., Tagliasacchi, A., & Zhang, H. (2020). BSP-Net: generating compact meshes via binary space partitioning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 42–51). Piscataway: IEEE.
27. Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., et al. (2022). LION: latent point diffusion models for 3D shape generation. In S. Koyejo, S. Mohamed, A. Agarwal, et al. (Eds.), *Proceedings of the 35th international conference on neural information processing systems*. (pp. 2132–2142). Red Hook: Curran Associates.
28. Liu, Z., Feng, Y., Black, M. J., Nowrouzezahrai, D., Paull, L., & Liu, W. (2023). MeshDiffusion: score-based generative 3D mesh modeling. In *Proceedings of the 11th international conference on learning representations*. Retrieved February 1, 2024, from https://iclr.cc/media/iclr-2023/Slides/11403_yiX8XSq.pdf.
29. Shim, J., Kang, C., & Joo, K. (2023). Diffusion-based signed distance fields for 3D shape generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20887–20897). Piscataway: IEEE.
30. Li, M., Duan, Y., Zhou, J., & Lu, J. (2023). Diffusion-SDF: text-to-shape via voxelized diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12642–12651). Piscataway: IEEE.
31. Hui, K., Li, R., Hu, J., & Fu, C. (2022). Neural wavelet-domain diffusion for 3D shape generation. In S. K. Jung, J. Lee, & A. W. Bargteil (Eds.), *ACM SIGGRAPH Asia 2022 conference proceedings* (pp. 1–9). New York: ACM.
32. Hu, J., Hui, K., Liu, Z., Li, R., & Fu, C. (2023). Neural wavelet-domain diffusion for 3D shape generation, inversion, and manipulation. arXiv preprint. arXiv:2302.00190.
33. Zheng, X. Y., Pan, H., Wang, P. S., Tong, X., Liu, Y., & Shum, H. Y. (2023). Locally attentional SDF diffusion for controllable 3D shape generation. *ACM Transactions on Graphics*, *42*(4), 1–13.
34. Saito, S., Huang, Z., Natsume, R., Morishima, S., Li, H., & Kanazawa, A. (2019). PIFu: pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2304–2314). Piscataway: IEEE.
35. Saito, S., Simon, T., Saragih, J. M., & Joo, H. (2020). PIFuHD: multi-level pixel-aligned implicit function for high-resolution 3D human digitization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 81–90). Piscataway: IEEE.
36. Or-El, R., Luo, X., Shan, M., Shechtman, E., Park, J. J., & Kemelmacher-Shlizerman, I. (2022). StyleSDF: high-resolution 3D-consistent image and geometry generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 13503–13513). Piscataway: IEEE.
37. Peng, S., Niemeyer, M., Mescheder, L. M., Pollefeys, M., & Geiger, A. (2020). Convolutional occupancy networks. In A. Vedaldi, H. Bischof, T. Brox, et al. (Eds.), *Proceedings of the 17th European conference on computer vision* (pp. 523–540). Cham: Springer.
38. Chen, Z., Yin, K., Fisher, M., Chaudhuri, S., & Zhang, H. (2019). BAE-NET: branched autoencoder for shape co-segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8489–8498). Piscataway: IEEE.
39. Niu, C., Li, M., Xu, K., & Zhang, H. (2022). RIM-Net: recursive implicit fields for unsupervised learning of hierarchical shape structures. In *Proceedings of the IEEE/CVF conference of computer vision and pattern recognition* (pp. 11769–11778). Piscataway: IEEE.
40. Gao, L., Yang, J., Wu, T., Yuan, Y., Fu, H., Lai, Y., et al. (2019). SDM-NET: deep generative network for structured deformable mesh. *ACM Transactions on Graphics*, *38*(6), 1–15.
41. Mo, K., Zhu, S., Chang, A. X., Yi, L., Tripathi, S., Guibas, L. J., et al. (2019). PartNet: a large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 909–918). Piscataway: IEEE.
42. Wu, R., Zhuang, Y., Xu, K., Zhang, H., & Chen, B. (2020). PQ-NET: a generative part seq2seq network for 3D shapes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 826–835). Piscataway: IEEE.
43. Li, M., & Zhang, H. (2021). D2IM-Net: learning detail disentangled implicit fields from single images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10246–10255). Piscataway: IEEE.
44. Zheng, X., Liu, Y., Wang, P., & Tong, X. (2022). SDF-StyleGAN: implicit sdf-based stylegan for 3D shape generation. *Computer Graphics Forum*, *41*(5), 52–63.
45. Zheng, Z., Yu, T., Dai, Q., & Liu, Y. (2021). Deep implicit templates for 3D shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1429–1439). Piscataway: IEEE.
46. Deng, Y., Yang, J., & Tong, X. (2021). Deformed implicit field: modeling 3D shapes with learned dense correspondence. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10286–10296). Piscataway: IEEE.
47. Van den Oord, A., Vinyals, O., & Kavukcuoglu, K. (2017). Neural discrete representation learning. In I. Guyon, U. V. Luxburg, S. Bengio, et al. (Eds.), *Proceedings of the 30th international conference on neural information processing systems* (pp. 6306–6315). Red Hook: Curran Associates.
48. Yan, X., Lin, L., Mitra, N. J., Lischinski, D., Cohen-Or, D., & Huang, H. (2022). ShapeFormer: transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF conference of computer vision and pattern recognition* (pp. 6229–6239). Piscataway: IEEE.
49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, et al. (Eds.), *Proceedings of the 30th international conference on neural information processing systems*. (pp. 5998–6008). Red Hook: Curran Associates.
50. Mittal, P., Cheng, Y., Singh, M., & Tulsiani, S. (2022). AutoSDF: shape priors for 3D completion, reconstruction and generation. In *Proceedings of the IEEE/CVF conference of computer vision and pattern recognition* (pp. 306–315). Piscataway: IEEE.
51. Oechsle, M., Mescheder, L. M., Niemeyer, M., Strauss, T., & Geiger, A. (2019). Texture fields: learning texture representations in function space. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4530–4539). Piscataway: IEEE.
52. Niemeyer, M., Mescheder, L. M., 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. 3501–3512). Piscataway: IEEE.

53. Chen, Z., Yin, K., & Fidler, S. (2022). AUV-Net: learning aligned UV maps for texture transfer and synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1455–1464). Piscataway: IEEE.
54. Chou, G., Bahat, Y., & Heide, F. (2023). Diffusion-SDF: conditional generative modeling of signed distance functions. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2262–2272). Piscataway: IEEE.
55. Zhang, B., Tang, J., Nießner, M., & Wonka, P. (2023). 3DShape2VecSet: a 3D shape representation for neural fields and generative diffusion models. *ACM Transactions on Graphics*, *42*(4), 1–16.
56. Erkoç, Z., Ma, F., Shan, Q., Nießner, M., & Dai, A. (2023). HyperDiffusion: generating implicit neural fields with weight-space diffusion. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 14254–14264). Piscataway: IEEE.
57. Liu, Y. T., Wang, L., Yang, J., Chen, W., Meng, X., Yang, B., et al. (2023). NeUDF: leaning neural unsigned distance fields with volume rendering. In *Proceedings of the IEEE/CVF conference of computer vision and pattern recognition* (pp. 237–247). Piscataway: IEEE.
58. Kosiorek, A. R., Strathmann, H., Zoran, D., Moreno, P., Schneider, R., Mokrá, S., et al. (2021). NeRF-VAE: a geometry aware 3D scene generative model. In T. Z. M. Meila (Ed.), *Proceedings of the 38th international conference on machine learning* (pp. 5742–5752). Stroudsburg: International Machine Learning Society.
59. Chan, E. R., Monteiro, M., Kellnhofer, P., Wu, J., & Wetzstein, G. (2021). Pi-GAN: periodic implicit generative adversarial networks for 3D-aware image synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5799–5809). Piscataway: IEEE.
60. Wang, Z., Deng, Y., Yang, J., Yu, J., & Tong, X. (2022). Generative deformable radiance fields for disentangled image synthesis of topology-varying objects. *Computer Graphics Forum*, *41*(7), 431–442.
61. Xu, Y., Peng, S., Yang, C., Shen, Y., & Zhou, B. (2022). 3D-aware image synthesis via learning structural and textural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18430–18439). Piscataway: IEEE.
62. Xue, Y., Li, Y., Singh, K. K., & Lee, Y. J. (2022). GIRAFFE hd: a high-resolution 3D-aware generative model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18440–18449). Piscataway: IEEE.
63. Xu, X., Pan, X., Lin, D., & Dai, B. (2021). Generative occupancy fields for 3D surface-aware image synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, et al. (Eds.), *Proceedings of the 34th international conference on neural information processing systems* (pp. 20683–20695). Red Hook: Curran Associates.
64. Pan, X., Xu, X., Loy, C. C., Theobalt, C., & Dai, B. (2021). A shading-guided generative implicit model for shape-accurate 3D-aware image synthesis. In M. Ranzato, A. Beygelzimer, Y. Dauphin, et al. (Eds.), *Proceedings of the 34th international conference on neural information processing systems*. (pp. 20002–20013). Red Hook: Curran Associates.
65. Deng, Y., Yang, J., Xiang, J., & Tong, X. (2022). GRAM: generative radiance manifolds for 3D-aware image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10663–10673). Piscataway: IEEE.
66. Mallikarjun, B.R., Pan, X., Elgharib, M., & Theobalt, C. (2023). GVP: generative volumetric primitives. arXiv preprint. arXiv:2303.18193.
67. Müller, N., Siddiqui, Y., Porzi, L., Bulò, S. R., Kontschieder, P., & Nießner, M. (2023). DiffRF: rendering-guided 3D radiance field diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4328–4338). Piscataway: IEEE.
68. Karnewar, A., Vedaldi, A., Novotný, D., & Mitra, N. J. (2023). HOLODIFFUSION: training a 3D diffusion model using 2D images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 18423–18433). Piscataway: IEEE.
69. Li, Y., & Kitani, K. (2023). 3D-CLFusion: fast text-to-3D rendering with contrastive latent diffusion. arXiv preprint. arXiv:2303.11938.
70. Kim, S. W., Brown, B., Yin, K., Kreis, K., Schwarz, K., Li, D., et al. (2023). NeuralField-LDM: scene generation with hierarchical latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8496–8506). Piscataway: IEEE.
71. Metzer, G., Richardson, E., Patashnik, O., Giryes, R., & Cohen-Or, D. (2022). Latent-NeRF for shape-guided generation of 3D shapes and textures. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12663–12673). Piscataway: IEEE.
72. Raj, A., Kaza, S., Poole, B., Niemeyer, M., Ruiz, N., Mildenhall, B., et al. (2023). DreamBooth3D: subject-driven text-to-3D generation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2349–2359). Piscataway: IEEE.
73. Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., et al. (2023). ProlificDreamer: high-fidelity and diverse text-to-3D generation with variational score distillation. In A. Oh, T. Neumann, A. Globerson, et al. (Eds.), *Proceedings of the 37th international conference on neural information processing systems* (pp. 8406–8441). Red Hook: Curran Associates.
74. Lin, C., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., et al. (2023). Magic3D: high-resolution text-to-3D content creation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 300–309). Piscataway: IEEE.
75. Chen, R., Chen, Y., Jiao, N., & Jia, K. (2023). Fantasia3D: disentangling geometry and appearance for high-quality text-to-3D content creation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 22189–22199). Piscataway: IEEE.
76. Deng, C., Jiang, C. M., Qi, C. R., Yan, X., Zhou, Y., Guibas, L. J., et al. (2022). NeRDi: single-view NeRF synthesis with language-guided diffusion as general image priors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20637–20647). Piscataway: IEEE.
77. Xu, J., Wang, X., Cheng, W., Cao, Y., Shan, Y., Qie, X., et al. (2023). Dream3D: zero-shot text-to-3D synthesis using 3D shape prior and text-to-image diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20908–20918). Piscataway: IEEE.
78. Raj, A., Kaza, S., Poole, B., Niemeyer, M., Ruiz, N., Mildenhall, B., et al. (2023). DreamBooth3D: subject-driven text-to-3D generation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2349–2359). Piscataway: IEEE.
79. Liu, R., Wu, R., Hoorick, B. V., Tokmakov, P., Zakharov, S., & Vondrick, C. (2023). Zero-1-to-3: zero-shot one image to 3D object. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9264–9275). Piscataway: IEEE.
80. Seo, J., Jang, W., Kwak, M., Ko, J., Kim, H., Kim, J., et al. (2023). Let 2D diffusion model know 3D-consistency for robust text-to-3D generation. arXiv preprint. arXiv:2303.07937.
81. Huang, Y., Wang, J., Shi, Y., Qi, X., Zha, Z., & Zhang, L. (2023). DreamTime: an improved optimization strategy for text-to-3D content creation. arXiv preprint. arXiv:2306.12422.
82. Hong, F., Chen, Z., Lan, Y., Pan, L., & Liu, Z. (2022). EVA3D: compositional 3D human generation from 2D image collections. In *Proceedings of the 11th international conference on learning representations* (pp. 1–15). Retrieved February 1, 2024, from https://openreview.net/pdf?id=g7U9jD_2CUr.
83. Cao, Y., Cao, Y., Han, K., Shan, Y., & Wong, K. K. (2023). DreamAvatar: text-and-shape guided 3D human avatar generation via diffusion models. arXiv preprint. arXiv:2304.00916.
84. Burkov, E., Rakhimov, R., Safin, A., Burnaev, E., & Lempitsky, V. (2023). Multi-NeuS: 3D head portraits from single image with neural implicit functions. *IEEE Access*, *11*, 95681–95691.
85. Tewari, A., Mallikarjun, B. R., Pan, X., Fried, O., Agrawala, M., & Theobalt, C. (2022). Disentangled3D: learning a 3D generative model with disentangled geometry and appearance from monocular images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1516–1525). Piscataway: IEEE.
86. Tang, J., Zhang, B., Yang, B., Zhang, T., Chen, D., Ma, L., et al. (2022). Explicitly controllable 3D-aware portrait generation. arXiv preprint. arXiv:2209.05434.
87. Tan, F., Fanello, S., Meka, A., Orts-Escolano, S., Tang, D., Pandey, R., et al. (2022). VoLux-GAN: a generative model for 3D face synthesis with HDRI relighting. In M. Nandigjav, N. J. Mitra, & A. Hertzmann (Eds.), *SIGGRAPH '22: special interest group on computer graphics and interactive techniques conference* (pp. 1–9). New York: ACM.
88. Wu, Y., Deng, Y., Yang, J., Wei, F., Chen, Q., & Tong, X. (2022). AniFaceGAN: animatable 3D-aware face image generation for video avatars. In S. Koyejo, S. Mohamed, A. Agarwal, et al. (Eds.), *Proceedings of the 35th international conference on neural information processing systems* (pp. 1245–1255). Red Hook: Curran Associates.
89. Zhang, D., Zhong, C., Guo, Y., Hong, Y., & Zhang, J. (2023). MetaHead: an engine to create realistic digital head. arXiv preprint. arXiv:2304.00838.
90. Wu, S., Yan, Y., Li, Y., Cheng, Y., Zhu, W., Gao, K., et al. (2023). GANHead: towards generative animatable neural head avatars. In *Proceedings of the*

IEEE/CVF conference on computer vision and pattern recognition (pp. 437–447). Piscataway: IEEE.
91. Ye, Z., Jiang, Z., Ren, Y., Liu, J., He, J., & Zhao, Z. (2023). GeneFace: generalized and high-fidelity audio-driven 3D talking face synthesis. In *Proceedings of the 11th international conference on learning representations*. Retrieved February 25, 2024, from https://openreview.net/pdf?id=YfwMIDhPccD.
92. Ye, Z., He, J., Jiang, Z., Huang, R., Huang, J., Liu, J., et al. (2023). GeneFace++: generalized and stable real-time audio-driven 3D talking face generation. arXiv preprint. arXiv:2305.00787.
93. Chai, L., Tucker, R., Li, Z., Isola, P., & Snavely, N. (2023). Persistent nature: a generative model of unbounded 3D worlds. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20863–20874). Piscataway: IEEE.
94. Xu, Y., Chai, M., Shi, Z., Peng, S., Skorokhodov, I., Siarohin, A., et al. (2023). DisCoScene: spatially disentangled generative radiance fields for controllable 3D-aware scene synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4402–4412). Piscataway: IEEE.
95. Po, R., & Wetzstein, G. (2023). Compositional 3D scene generation using locally conditioned diffusion. arXiv preprint. arXiv:2303.12218.
96. Lin, Y., Bai, H., Li, S., Lu, H., Lin, X., Xiong, H., et al. (2023). CompoNeRF: text-guided multi-object compositional NeRF with editable 3D scene layout. arXiv preprint. arXiv:2303.13843.
97. Sitzmann, V., Martel, J. N. P., Bergman, A. W., Lindell, D. B., & Wetzstein, G. (2020). Implicit neural representations with periodic activation functions. In H. Larochelle, M. Ranzato, R. Hadsell, et al. (Eds.), *Proceedings of the 33rd international conference on neural information processing systems*. (pp. 1456–1476). Red Hook: Curran Associates.
98. Sun, C., Sun, M., & Chen, H. (2022). Direct voxel grid optimization: super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5449–5459). Piscataway: IEEE.
99. Armandpour, M., Zheng, H., Sadeghian, A., Sadeghian, A., & Zhou, M. (2023). Re-imagine the negative prompt algorithm: transform 2D diffusion into 3D, alleviate Janus problem and beyond. arXiv preprint. arXiv:2304.04968.
100. Tang, J., Wang, T., Zhang, B., Zhang, T., Yi, R., Ma, L., et al. (2023). Make-It-3D: high-fidelity 3D creation from a single image with diffusion prior. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 22762–22772). Piscataway: IEEE.
101. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., & Black, M. J. (2015). SMPL: a skinned multi-person linear model. *ACM Transactions on Graphics*, *34*(6), 1–16.
102. Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, *41*(4), 1–15.
103. Skorokhodov, I., Tulyakov, S., Wang, Y., & Wonka, P. (2022). EpiGRAF: rethinking training of 3D GANs. In S. Koyejo, S. Mohamed, A. Agarwal, et al. (Eds.), *Proceedings of the 36th international conference on neural information processing systems* (pp. 24487–24501). Red Hook: Curran Associates.
104. Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., et al. (2022). GET3D: a generative model of high quality 3D textured shapes learned from images. In S. Koyejo, S. Mohamed, A. Agarwal, et al. (Eds.), *Proceedings of the 36th international conference on neural information processing systems* (pp. 31841–31854). Red Hook: Curran Associates.
105. Singer, U., Sheynin, S., Polyak, A., Ashual, O., Makarov, I., Kokkinos, F., et al. (2023). Text-To-4D dynamic scene generation. In A. Krause, E. Brunskill, K. Cho, et al. (Eds.), *Proceedings of the 40th international conference on machine learning* (pp. 31915–31929). Stroudsburg: International Machine Learning Society.
106. Wei, J., Wang, H., Feng, J., Lin, G., & Yap, K. (2023). TAPS3D: text-guided 3D textured shape generation from pseudo supervision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16805–16815). Piscataway: IEEE.
107. Skorokhodov, I., Siarohin, A., Xu, Y., Ren, J., Lee, H., Wonka, P., et al. (2023). 3D generation on imagenet. In *Proceedings of the 110th international conference on learning representations*. Retrieved February 25, 2024, from https://openreview.net/pdf?id=U2WjB9xxZ9q.
108. Shen, B., Yan, X., Qi, C. R., Najibi, M., Deng, B., Guibas, L. J., et al. (2023). GINA-3D: learning to generate implicit neural assets in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4913–4926). Piscataway: IEEE.
109. Zhu, J., Ma, H., Chen, J., & Yuan, J. (2023). Few-shot 3D shape generation. arXiv preprint. arXiv:2305.11664.
110. Anciukevicius, T., Xu, Z., Fisher, M., Henderson, P., Bilen, H., Mitra, N. J., et al. (2023). RenderDiffusion: image diffusion for 3D reconstruction, inpainting and generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12608–12618). Piscataway: IEEE.
111. Gupta, A., Xiong, W., Nie, Y., Jones, I., & Oguz, B. (2023). 3DGen: triplane latent diffusion for textured mesh generation. arXiv preprint. arXiv:2303.05371.
112. Gu, J., Trevithick, A., Lin, K., Susskind, J. M., Theobalt, C., Liu, L., et al. (2023). NerfDiff: single-image view synthesis with NeRF-guided distillation from 3D-aware diffusion. In A. Krause, E. Brunskill, K. Cho, et al. (Eds.), *Proceedings of the 40th international conference on machine learning* (pp. 11808–11826). Stroudsburg: International Machine Learning Society.
113. Chen, H., Gu, J., Chen, A., Tian, W., Tu, Z., Liu, L., et al. (2023). Single-stage diffusion NeRF: a unified approach to 3D generation and reconstruction. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2416–2425). Piscataway: IEEE.
114. Gu, J., Gao, Q., Zhai, S., Chen, B., Liu, L., & Susskind, J. M. (2023). Learning controllable 3D diffusion models from single-view images. arXiv preprint. arXiv:2304.06700.
115. Noguchi, A., Sun, X., Lin, S., & Harada, T. (2022). Unsupervised learning of efficient geometry-aware neural articulated representations. In S. Avidan, G. Brostow, M. Cissé, et al. (Eds.), *Proceedings of the 17th European conference on computer vision* (pp. 597–614). Cham: Springer.
116. Zhang, J., Jiang, Z., Yang, D., Xu, H., Shi, Y., Song, G., et al. (2022). AvatarGen: a 3D generative model for animatable human avatars. In S. Avidan, G. J. Brostow, M. Cissé, et al. (Eds.), *Proceedings of the 17th European conference on computer vision workshops* (pp. 668–685). Cham: Springer.
117. Dong, Z., Chen, X., Yang, J., Black, M. J., Hilliges, O., & Geiger, A. (2023). AG3D: learning to generate 3D avatars from 2D image collections. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 14870–14881). Piscataway: IEEE.
118. Sun, J., Wang, X., Shi, Y., Wang, L., Wang, J., & Liu, Y. (2022). IDE-3D: interactive disentangled editing for high-resolution 3D-aware portrait synthesis. *ACM Transactions on Graphics*, *41*(6), 1–10.
119. Jiang, K., Chen, S. Y., Liu, F. L., Fu, H., & Gao, L. (2022). NeRFFaceEditing: disentangled face editing in neural radiance fields. In S. K. Jung, J. Lee, & A. W. Bargteil (Eds.), *Proceedings of the ACM SIGGRAPH Asia 2022* (pp. 1–9). New York: ACM.
120. Kim, G., & Chun, S. Y. (2023). DATID-3D: diversity-preserved domain adaptation using text-to-image diffusion for 3D generative model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 14203–14213). Piscataway: IEEE.
121. Kim, G., Jang, J. H., & Chun, S. Y. (2023). PODIA-3D: domain adaptation of 3D generative model across large domain gap using pose-preserved text-to-image diffusion. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 22546–22555). Piscataway: IEEE.
122. Sun, J., Wang, X., Wang, L., Li, X., Zhang, Y., Zhang, H., et al. (2023). Next3D: generative neural texture rasterization for 3D-aware head avatars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20991–21002). Piscataway: IEEE.
123. Xu, E. Z., Zhang, J., Liew, J. H., Zhang, W., Bai, S., Feng, J., et al. (2023). PV3D: a 3D generative model for portrait video generation. In *Proceedings of the 11th international conference on learning representations*. Retrieved February 25, 2024, from https://openreview.net/pdf?id=o3yygm3lnzS.
124. Deng, B., Wang, Y., & Wetzstein, G. (2023). LumiGAN: unconditional generation of relightable 3D human faces. arXiv preprint. arXiv:2304.13153.
125. Jiang, K., Chen, S., Fu, H., & Gao, L. (2023). NeRFFaceLighting: implicit and disentangled face lighting representation leveraging generative prior in neural radiance fields. *ACM Transactions on Graphics*, *42*(3), 1–18.
126. An, S., Xu, H., Shi, Y., Song, G., Ogras, Ü. Y., & Luo, L. (2023). PanoHead: geometry-aware 3D full-head synthesis in 360°. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 20950–20959). Piscataway: IEEE.
127. Cheng, Y., Yan, Y., Zhu, W., Pan, Y., Pan, B., & Yang, X. (2023). Head3D: complete 3D head generation via tri-plane feature distillation. arXiv preprint. arXiv:2303.15892.

128. Trevithick, A., Chan, M. A., Stengel, M., Chan, E. R., Liu, C., Yu, Z., et al. (2023). Real-time radiance fields for single-image portrait view synthesis. *ACM Transactions on Graphics*, *42*(4), 1–15.
129. Wang, T., Zhang, B., Zhang, T., Gu, S., Bao, J., Baltrusaitis, T., et al. (2023). RODIN: a generative model for sculpting 3D digital avatars using diffusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4563–4573). Piscataway: IEEE.
130. Son, M., Park, J. J., Guibas, L. J., & Wetzstein, G. (2023). SinGRAF: learning a 3D generative radiance field for a single scene. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8507–8517). Piscataway: IEEE.
131. Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Li, F. (2009). ImageNet: a large-scale hierarchical image database. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 248–255). Piscataway: IEEE.
132. Raistrick, A., Lipson, L., Ma, Z., Mei, L., Wang, M., Zuo, Y., et al. (2023). Infinite photorealistic worlds using procedural generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12630–12641).
133. Bao, C., Yang, B., Junyi, Z., Hujun, B., Yinda, Z., Zhaopeng, C., et al. (2022). NeuMesh: learning disentangled neural mesh-based implicit field for geometry and texture editing. In S. Avidan, G. Brostow, M. Cissé, et al. (Eds.), *Proceedings of the 19th European conference on computer vision* (pp. 597–614). Cham: Springer.
134. Wu, T., Sun, J., Lai, Y., & Gao, L. (2023). DE-NeRF: DEcoupled neural radiance fields for view-consistent appearance editing and high-frequency environmental relighting. In J. Kim & M. C. Lin (Eds.), *ACM SIGGRAPH 2023 conference proceedings* (pp. 1–11). New York: ACM.

Publisher's Note