USING NEURAL NETWORK METHODS FOR GENERATION AND RECONSTRUCTION OF 3D IMAGES

Oksana Udovytska
USA

Abstract
This article examines modern neural network methods for generating and reconstructing three-dimensional images, applied in computer graphics, virtual and augmented reality, medicine, and digitalization of objects. Particular attention is paid to implicit neural network representations (such as distance and occupancy functions), neural radial density fields (Neural Radiance Fields , NeRF), as well as approaches using pretrained diffusion models for text-driven 3D content generation. A comparative analysis of architectures, training data requirements, computational complexity, and the quality of the resulting 3D models is conducted. The key advantages and limitations of existing solutions are identified, including issues of scalability, the lack of labeled 3D data, and the accuracy of geometric reconstruction.

Keywords: Neural network methods, three-dimensional images, 3D reconstruction, 3D content generation, deep learning. and implicit representations, Neural Radiance Fields (NeRF), diffusion models, computer graphics.

Introduction

In recent years, the fields of computer vision and computer graphics have undergone significant changes thanks to the application of deep learning methods to problems previously solved by classical algorithms. One such area is the creation and reconstruction of three-dimensional (3D) images—a process by which complete three-dimensional representations of objects and scenes are reconstructed or generated from two-dimensional (2D) data (images, video sequences, etc.). Traditional 3D reconstruction methods, such as photogrammetry and stereovision , rely on rigid geometric models and require precise camera poses and a large number of images to achieve high accuracy. On the other hand, approaches based on deep neural network models can directly learn to represent the shape and appearance of objects, significantly expanding the capabilities of automation and improving the quality of results in situations with limited data or complex scene geometry .

One of the key advances in this area was the emergence of Neural Radiance NeRF (NeRF) is a neural network approach to representing a scene as a continuous function of density and radiation, parameterized by a deep neural network . Introduced in 2020, the NeRF model allows for the reconstruction of the original scene and the synthesis of images from new perspectives using only a set of heterogeneous 2D images and camera pose information [1]. This approach is fundamentally different from classical methods because it does not require manually

Web of Scientists and Scholars: Journal of Multidisciplinary Research

specified 3D structures or depth, but instead optimizes network parameters to match the prediction with multi-view observations.

In a parallel direction, methods of implicit representations of surfaces are being developed, such as DeepSDF or Occupancy Networks that model the shape of an object as a level-zero signed function distance (SDF) or spatial occupancy probabilities. These networks are trained on a large set of 3D models and allow for compact encoding of geometric information with high accuracy and surface smoothness for various classes of objects [2].

Of particular interest are approaches using pixel- aligned neural network functions , such as PIFu and its derivatives, which are capable of reconstructing detailed 3D human models from one or more images. These methods utilize density features from 2D images to interpret 3D structure and texture, which is particularly relevant for the digitalization of people and interactive applications [3].

Moreover, the development of powerful 2D diffusion models has opened up new possibilities for generating 3D content with text or conditionally oriented cues. Methods such as DreamFusion and subsequent modifications use pre-trained 2D diffusion models to optimize 3D representations, enabling the creation of complex 3D objects with a high level of visual quality and controllability, even in the absence of extensive 3D annotations in the training set.

Thus, the combination of deep neural network models, implicit representations, and modern learning strategies opens new perspectives in 3D image generation and reconstruction tasks, offering flexible, powerful, and often more accurate alternatives to traditional algorithms.

The scientific novelty lies in the comprehensive and comparative analysis of modern neural network methods for generating and reconstructing three-dimensional images, identifying their key advantages, limitations, and development prospects.

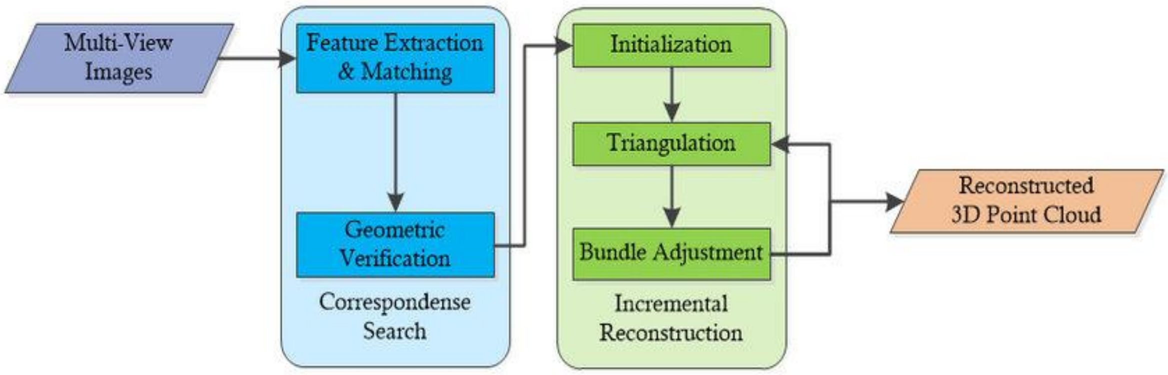

Figure 1 – General scheme of neural network reconstruction of three-dimensional images from two-dimensional data

Overview of approaches:

1. Neural radiance fields (NeRF) represent a scene as a continuous function of volume density and color radiance, parameterized by MLP : the input is a 5D coordinate (x , y , z , viewing

direction), and the output is density and RGB color. Using differentiable volume rendering, the model is optimized on a set of multi-view images with known camera poses, resulting in photorealistic synthesis of new views. The main limitations are the high computational cost of training and rendering, and the dependence of quality on the coverage of the scene by the original views [4].

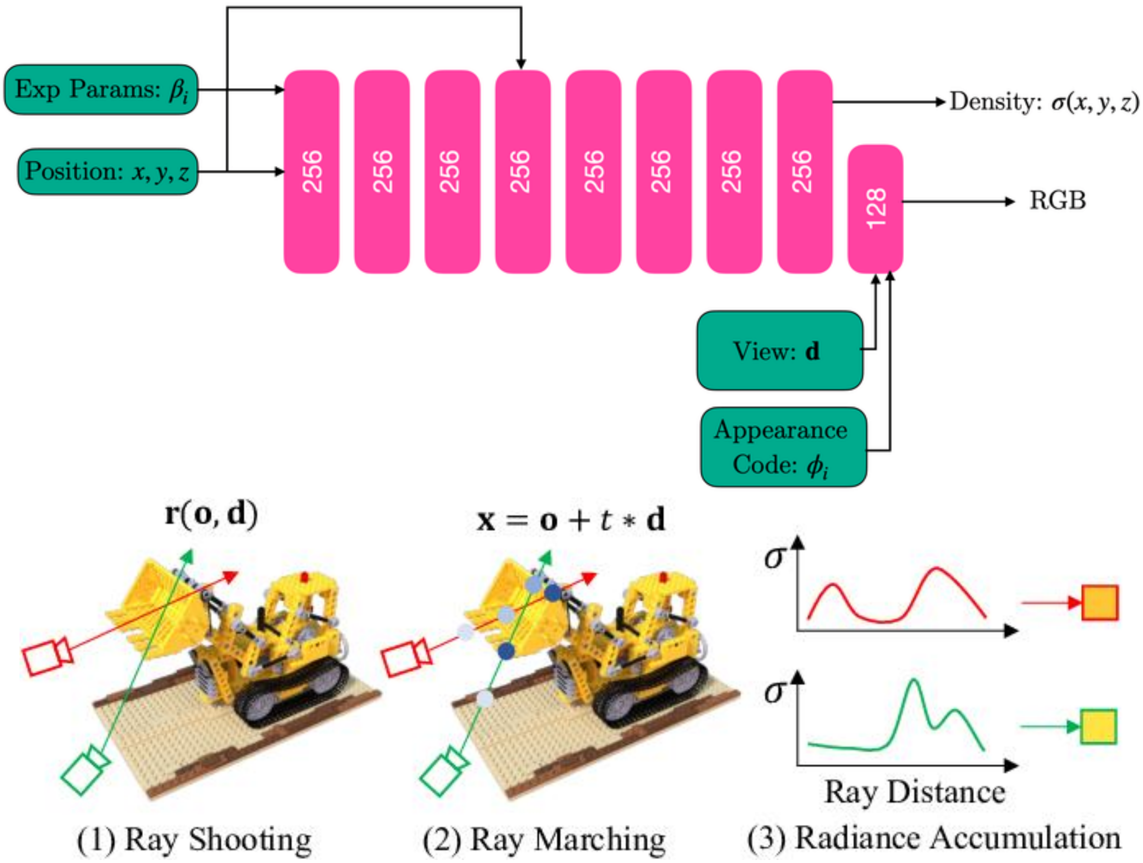

Figure 2 – The operating principle of the Neural method Radiance Fields (NeRF):
a – architecture of the neural radiation field;
b – an example of synthesizing images of a scene from new angles.

2. Implicit methods encode the surface as the level of the zero function:

DeepSDF trains a network to predict signed distance (SDF) is a scalar value of the distance to the nearest surface with a sign, which gives a smooth surface resolving to arbitrary accuracy and good properties for interpolation and a priori generation of shapes [5].

Occupancy Networks model the probability of a point's "occupancy" (whether a point belongs to an object's body). This allows for representing objects "at infinite resolution" without storing a large voxel grid and for efficiently reconstructing shapes from different input signals [6].

Both groups provide a compact, accurate representation of geometry and are well suited for reconstruction tasks and generation of class sets of shapes; weaknesses include the need for large sets of 3D models for training and sensitivity to noise in the training data [5].

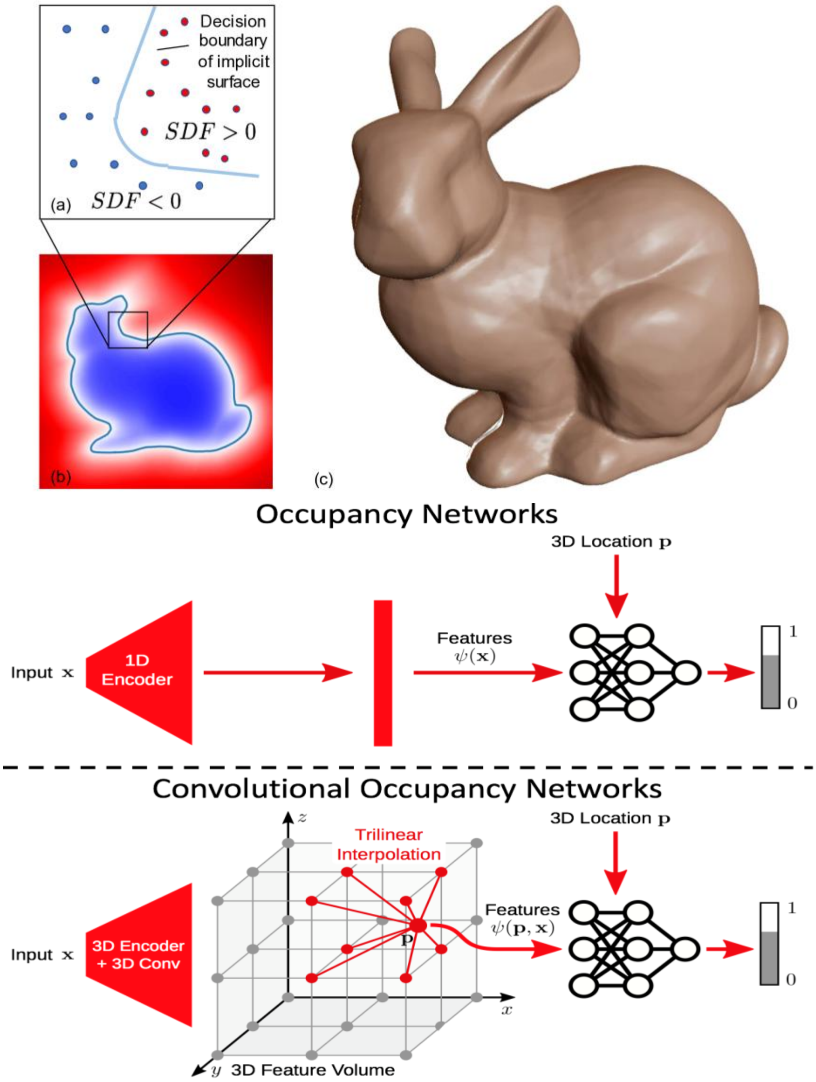

Figure 3 – Implicit neural network representations of three-dimensional objects:
a – representation of the surface using signed distance function (DeepSDF);
b – restoration of the object's shape using the Occupancy method Networks .

3. Pixel - aligned implicit functions (PIFu and derivatives) introduces the idea of aligning 2D pixels with local 3D representations (pixel - aligned features), which allows for the efficient use of image information to reconstruct detailed geometry (in particular clothed human digitization) from one or more images. PIFu and PIFuHD are particularly effective for digitalizing people, where high-detail clothing and fine reliefs are essential. Limitations include their specialization for people/clothing and the difficulty of generalizing to arbitrary categories without additional training [7].

ISSN (E): 2938-3811

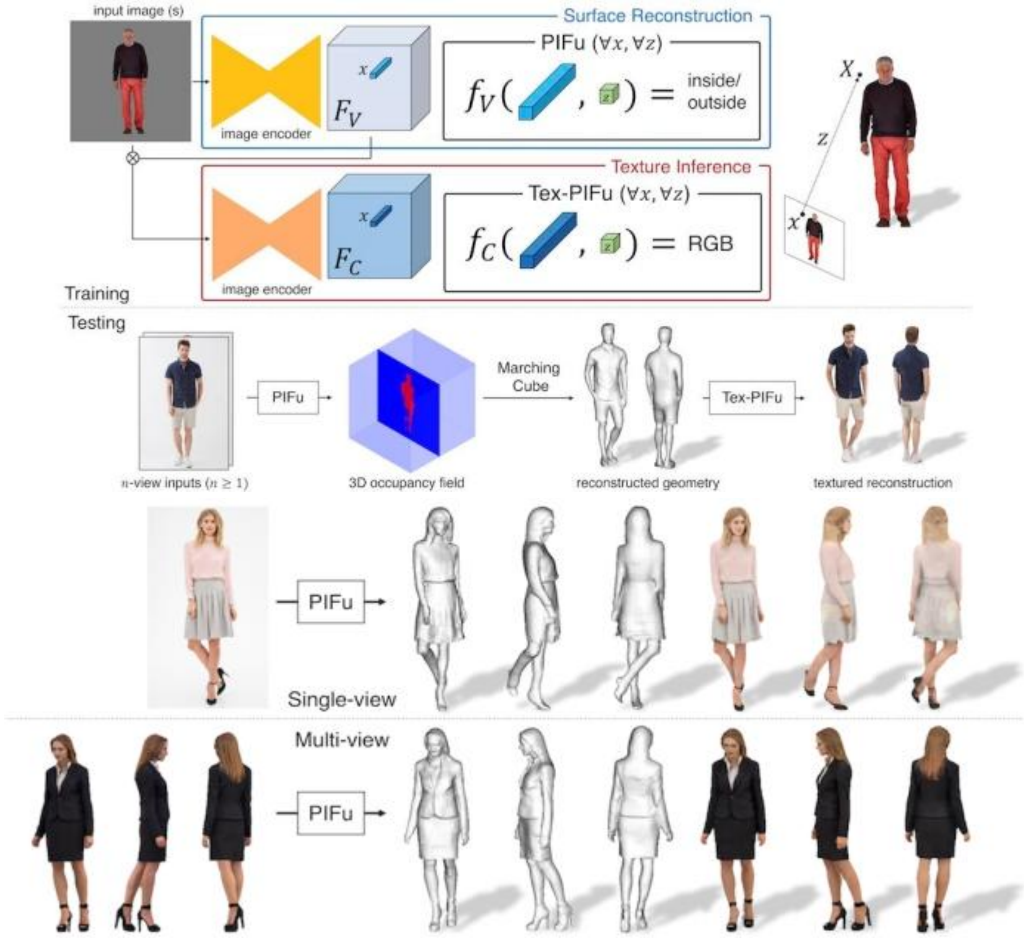

Figure 4 – Reconstruction of a 3D human model using the PIFu method based on a single input image

4. Early generative approaches for 3D used GAN architectures to synthesize voxel grids or other discrete representations (e.g., 3D - GAN) . These methods laid the foundations for generating 3D shapes from latent space, but face problems with memory (for voxel-based representations), resolution, and training robustness; modern reviews summarize both the achievements and limitations of 3D - GANs and their subsequent modifications [8].
5. To overcome the shortage of labeled 3D data , approaches using pre-trained 2D diffusion models as priors for 3D representation optimization have been proposed. In DreamFusion , the idea is to optimize a parameterized 3D object (e.g., NeRF) so that its 2D renders have high probability according to a pre-trained 2D diffusion model; a score - distillation technique is used to transfer signals from the 2D model into 3D parameter updates. This enables text- driven 3D generation without explicit 3D annotations, but issues of optimization stability, artifacts, and high computational cost remain [9].

Given a caption, DreamFusion generates relightable 3D objects with high-fidelity appearance, depth, and normals. Objects are represented as a Neural Radiance Field and leverage a pretrained text-to-image diffusion prior such as Imagen.

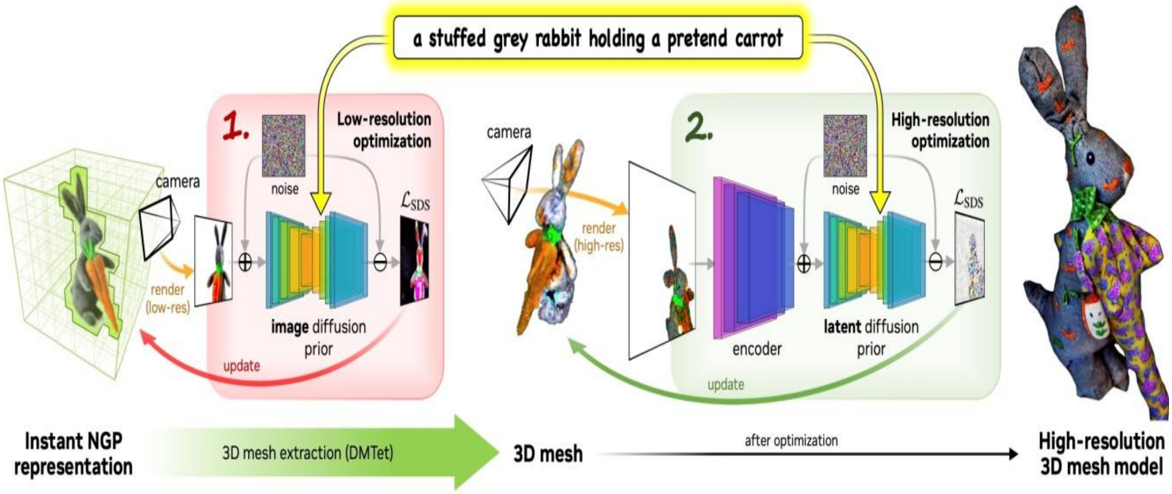

Figure 5 – An example of generating a three-dimensional object from a text description using a diffusion model (Dream Fusion)

6. As NeRF evolved, methods for accelerating it (structures for fast field querying, sampling , approximations) and hybrid approaches (e.g., combining NeRF /implicit representations with raster or splatting techniques) emerged. Current research aims to reduce training/rendering time and improve scalability to large scenes [10].

Table 1 - Comparison of modern methods for generating and reconstructing 3D images

Approach	Performance	Advantages	Restrictions
NeRF (Neural Radiance Fields)	Continuous radiation/volume field (MLP)	Photorealistic novel-view synthesis ; good for static scenes	Slow training/rendering; requires multiple input views
DeepSDF (Signed Distance Functions)	Implicit SDF function	Smooth, compact geometry; good for interpolation/shape generation	3D models needed for training; does not directly provide appearance/textures
Occupancy Networks	Occupancy function	Infinite resolution representation; memory efficient	Surface extraction threshold; requires 3D data
PIFu / PIFuHD (pixel-aligned)	Pixel -aligned implicit function	Highly detailed human forms from 1–2 images	Specialized in people; demanding of input quality
3D-GAN and 3D generative models	Voxels / point clouds / mesh (implementation dependent)	Generation from latent space; early successes in 3D generation	Resolution and stability limitations; high memory consumption
DreamFusion (text-to-3D via 2D-diffusion)	3D optimization (eg NeRF) for the 2D diffusion model	Text-to-3D without 3D annotations; flexibility in styles	Unstable optimization; computationally expensive; artifacts possible

Data requirements and computational aspects. Neural network methods for generating and reconstructing 3D images place high demands on both input data and computing resources. The quality and type of training data significantly impact the accuracy of geometric reconstruction and the visual fidelity of the resulting 3D models.

Methods based on implicit representations (DeepSDF , Occupancy Networks require large sets of labeled 3D objects or dense scans, which limits their use in data-poor tasks. NeRF family approaches use multi-view 2D images with known camera parameters, and the density and diversity of camera angles directly impact the quality of new view synthesis and training stability.

Modern text-to-3D methods using pre-trained 2D diffusion models reduce the dependence on 3D annotations, but require significant computational resources to optimize 3D representations and are characterized by high time and memory costs.

From a computational standpoint, most of the approaches under consideration remain resource-intensive: training and rendering NeRF -like models require GPUs and significant processing time, especially for high-resolution scenes. Therefore, current research areas include accelerating neural network architectures, developing more efficient sampling schemes , and using hybrid representations to reduce computational complexity without significantly compromising quality.

Quality and metric evaluations. Assessing the quality of 3D image generation and reconstruction results is an important step in analyzing neural network methods and includes both quantitative and qualitative indicators. The choice of metrics depends on the type of 3D representation (voxels , meshes, implicit functions, radiation fields) and the target task.

To assess the geometric accuracy of reconstruction, distance metrics between surfaces, such as the Chamfer, are widely used. Distance and Earth Mover's Distance (EMD), as well as the Intersection coefficient over Union (IoU) for voxel representations. These metrics allow us to quantify the degree of conformity of the reconstructed shape to the reference model.
PSNR, SSIM, and LPIPS metrics, which characterize the degree of similarity to reference images, are used in tasks such as synthesizing new camera angles and assessing visual image quality. However, these metrics do not always fully reflect the subjective perception of realism, and are therefore often supplemented by visual expert assessment.
For generative models, statistical metrics of shape diversity and plausibility are also used, and for text-to-3D tasks, additional heuristics assessing the consistency of geometry and visual appearance are used. Taken together, the use of multiple metrics allows for a more objective and comprehensive assessment of the quality of neural network methods for generating and reconstructing 3D images.

Areas of application. Neural network methods for generating and reconstructing three-dimensional images are widely used in various fields of science and industry due to their ability to automatically reconstruct the geometry and appearance of objects from a limited set of data. In computer graphics and the gaming industry, such methods are used for automated 3D model creation, accelerating the development of digital scenes and characters. In virtual and augmented reality, neural network-based 3D approaches enable the creation of realistic, interactive scenes and objects that adapt to the user.
In medical applications, 3D reconstruction methods are used to reconstruct anatomical structures from tomography and imaging data of complex biological objects, which helps improve the accuracy of diagnosis and treatment planning. In architecture, industrial design, and engineering, neural network models are used for digital prototyping and shape analysis.
A separate area is related to the digitalization of people (human Digitalization , including the reconstruction of body shape and clothing from images, is in demand in the fashion industry, online fittings, and social platforms.

Table 2 - Areas of application of neural network methods for 3D generation and reconstruction

Scope of application	**Main tasks**	**Benefits of using neural networks**
Computer graphics and games	Creation of 3D models, scenes and characters	Reduced manual labor, high realism
Virtual and augmented reality	Generation of interactive 3D scenes	Flexibility and adaptability of content
Medicine	Reconstruction of anatomical structures	Improving visualization accuracy
Architecture and engineering	Digital prototyping , form analysis	Fast modeling of complex objects
Digitalization of people	Restoration of body, clothing, appearance	Detailing of limited data

Problems and Limitations. Despite significant progress in neural network- based 3D image generation and reconstruction, existing methods have a number of significant limitations that affect their practical application.
One of the main challenges is the scarcity and high cost of obtaining high-quality 3D training data. Implicit models and NER -like approaches require either large datasets of labeled 3D objects or multi-view images with precisely known camera parameters, which limits the scalability of these solutions.
A significant limitation remains high computational complexity. Training and optimizing neural network 3D representations requires significant computational resources and time, especially at high spatial resolution and with complex scene geometry.
Furthermore, many methods exhibit limited generalization: models trained on specific object classes or scenes often transfer poorly to new domains. Text-to-3D approaches also face the challenge of optimization instability and insufficient geometric correctness, despite the high visual quality of the results.

Table 3 - Main problems and limitations of neural network methods for 3D generation and reconstruction

Problem category	**Manifestation**	**Impact on application**
Data	Lack of labeled 3D sets, difficulty of collection	Limitations in scalability and versatility
Computing resources	High cost of training and rendering	GPU/cluster requirements, long calculation time
Generalizing ability	Poor transfer to new scenes and classes	The need for additional training for the task
Geometric accuracy	Artifacts, shape inaccuracies	Reduced suitability for engineering and medical tasks
Stability of training	Unstable optimization (text-to-3D)	Difficulties in reproducibility of results

Development Prospects. Further development of neural network methods for generating and reconstructing 3D images is focused on improving their efficiency, versatility, and practical applicability. One key area is the development of hybrid representations that combine the advantages of implicit models, neural radiation fields, and accelerated storage structures, thereby achieving a balance between reconstruction quality and computational complexity.
An important perspective is to reduce the dependence on labeled 3D data through the use of self-learning, multimodal datasets and knowledge transfer from pre-trained 2D models, including diffusion architectures. This facilitates the wider application of these methods in data-constrained settings.
Special attention is being paid to accelerating training and rendering, including architectural optimization, efficient sampling schemes , and hardware-specific implementations. Improving

Web of Scientists and Scholars: Journal of Multidisciplinary Research

the geometric correctness and robustness of training remains a significant focus, particularly in text-to-3D approaches, which is critical for engineering and medical applications.
Overall, the combination of algorithmic improvements, new data sources, and computational optimizations forms the basis for the transition of neural network 3D technologies from experimental solutions to widespread industrial use.

Conclusion

Thus, neural network methods radically expand the possibilities of creating and reconstructing 3D content. Implicit representations (DeepSDF , Occupancy Networks offer compact and accurate shape representations, NeRF demonstrates outstanding photorealism for species synthesis, and new approaches based on 2D diffusion models (DreamFusion and later) open the way to convenient text- driven 3D generation.

References

1. Gao KY, Gao Y., He H., Lu D., Xu L., Li J. NeRF : Neural Radiance Field in 3D Vision: A Comprehensive Review // arXiv . 2022. Vol. 2210.00379v6. URL: https://arxiv.org/html/2210.00379v6
2. 3D Reconstruction Survey / Jülich Supercomputing Center. URL: https://juser.fz-juelich.de/record/1017988/files/3D_Reconstruction_Survey.pdf
3. Dellaert F. NeRF Explosion 2020: Survey and Resources on Neural Radiance Fields. URL: https://dellaert.github.io/NeRF/
4. Mildenhall B., Srinivasan PP, Tancik M. et al. NeRF : Representing Scenes as Neural Radiance Fields for View Synthesis // arXiv . 2020. Vol. 2003.08934. URL: https://arxiv.org/abs/2003.08934
5. Park JJ, Florence P, Straub J et al. DeepSDF : Learning Continuous Signed Distance Functions for Shape Representation // arXiv . 2019. Vol. 1901.05103. URL: https://arxiv.org/abs/1901.05103
6. Mescheder L., Oechsle M., Niemeyer M. et al. Occupancy Networks: Learning 3D Reconstruction in Function Space // arXiv . 2018. Vol. 1812.03828. URL: https://arxiv.org/abs/1812.03828
7. Saito S., Huang Z., Natsume R. et al. PIFu : Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization // arXiv . 2019. Vol. 1905.05172. URL: https://arxiv.org/abs/1905.05172
8. Wu Z., Song S., Khosla A. et al. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling // arXiv . 2016. Vol. 1610.07584. URL: https://arxiv.org/abs/1610.07584
9. Poole B., Ramesh A., Dhariwal P. et al. DreamFusion : Text-to-3D using 2D Diffusion // arXiv . 2022. Vol. 2209.14988. URL: https://arxiv.org/abs/2209.14988
10. Ravir S. et al. Proceedings of the International Conference on Bioscience and Systems (ICBS 2023). 2023. URL: https://cseweb.ucsd.edu/~ravir/icbs23.pdf