# Neural LerPlane Representations for Fast 4D Reconstruction of Deformable Tissues

Chen Yang[1], Kailing Wang[1], Yuehao Wang[2], Xiaokang Yang[1],
and Wei Shen[1(✉)]

[1] MoE Key Lab of Artificial Intelligence, AI Institute,
Shanghai Jiao Tong University, Shanghai, China
`wei.shen@sjtu.edu.cn`
[2] Department of Computer Science and Engineering,
The Chinese University of Hong Kong, Sha Tin, Hong Kong

**Abstract.** Reconstructing deformable tissues from endoscopic stereo videos in robotic surgery is crucial for various clinical applications. However, existing methods relying only on implicit representations are computationally expensive and require dozens of hours, which limits further practical applications. To address this challenge, we introduce LerPlane, a novel method for fast and accurate reconstruction of surgical scenes under a single-viewpoint setting. LerPlane treats surgical procedures as 4D volumes and factorizes them into explicit 2D planes of static and dynamic fields, leading to a compact memory footprint and significantly accelerated optimization. The efficient factorization is accomplished by fusing features obtained through linear interpolation of each plane and enables using lightweight neural networks to model surgical scenes. Besides, LerPlane shares static fields, significantly reducing the workload of dynamic tissue modeling. We also propose a novel sample scheme to boost optimization and improve performance in regions with tool occlusion and large motions. Experiments on DaVinci robotic surgery videos demonstrate that LerPlane accelerates optimization by over $100\times$ while maintaining high quality across various non-rigid deformations, showing significant promise for future intraoperative surgery applications.

**Keywords:** Fast 3D Reconstruction · Neural Rendering · Robotic Surgery

## 1 Introduction

Reconstructing deformable tissues in surgical scenes accurately and efficiently from endoscope stereo videos is a challenging and active research topic. Such techniques can facilitate constructing virtual surgery environments for surgery
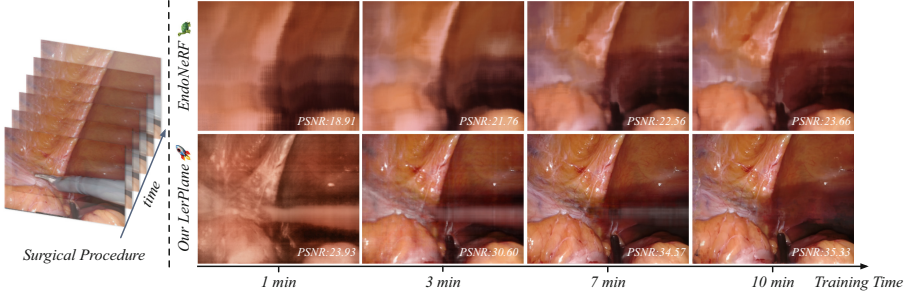
**Fig. 1.** Performance along with training time. We show the results of EndoNeRF (top) and LerPlane (bottom) with the same training time. LerPlane exhibits remarkable restoration of surgical scenes with just a few minutes of optimization.

robot learning and AR/VR surgery training and provide vivid and specific training for medics on human tissues. Moreover, real-time reconstruction further expands its applications to intraoperative use, allowing surgeons to navigate and precisely control surgical instruments while having a complete view of the surgical scene. This capability could reduce the need for invasive follow-up procedures and address the challenge of operating within a confined field of view.

Neural Radiance Fields (NeRFs) [16], a promising approach for 3D reconstruction, have demonstrated strong potential in accurately reconstructing deformable tissues in dynamic surgical scenes from endoscope stereo videos. EndoNeRF [26], a recent representative approach, represents deformable surgical scenes using a canonical neural radiance field and a time-dependent neural displacement field, achieving impressive reconstruction of deformable tissues. However, the optimization for dynamic NeRFs is computationally intensive, often taking dozens of hours, as each generated pixel requires hundreds of neural network calls. This computational bottleneck significantly constrains the widespread application of these methods in surgical procedures.

Recently, explicit and hybrid methods have been developed for modeling static scenes, achieving significant speedups over NeRF by employing explicit spatial data structures [8,9,27] or features decoded by small MLPs [5,17,25]. Nevertheless, these methods have only been applied to static scenes thus far. Adopting these methods to surgical scenes presents significant challenges for two primary reasons. Firstly, encoding temporal information is essential for modeling surgical scenes while naively adding a temporal dimension to the explicit data structure can significantly increase memory and computational requirements. Secondly, dynamic surgical scene reconstruction suffers from limited viewpoints, often providing only one view per timestep, as opposed to static scenes, which can fully use multi-view consistency for further regularization. This condition requires sharing information across disjoint timesteps for better reconstruction.

To address the aforementioned challenges, we propose a novel method for fast and accurate reconstruction of deformable tissues in surgical procedures, Neural LerPlane (Linear Interpolation Plane), by leveraging explicitly represented

multi-plane fields. Specifically, we treat surgical procedures as 4D volumes, where the time axis is orthogonal to 3D spatial coordinates. LerPlane factorizes 4D volumes into 2D planes and uses space planes to form static fields and space-time planes to form dynamic fields. This factorization results in a compact memory footprint and significantly accelerates optimization compared to previous methods [6,26], which rely on pure MLPs, as shown in Fig. 1. LerPlane enables information sharing across timesteps within the static field, thereby reducing the negative impact of limited viewpoints. Moreover, considering the surgical instrument occlusion, we develop a novel sample approach based on tool masks and contents, which assigns higher sampling probability to tissue pixels that have been occluded by tools or have a more extensive motion range. By targeting these regions, our approach allows for more efficient sampling during the training, leading to higher-quality results and faster optimization.

We summarize our contributions:

1. A fast deformable tissue reconstruction method, with rendering quality comparable to or better than the previous method in just 3 min, which is over 100× faster.
2. An efficient representation of surgical scenes, which includes static and dynamic fields, enabling fast optimization and high reconstruction quality.
3. A novel sampling method that boosts optimization and improves the rendering quality. Compared to previous methods, our LerPlane, achieves much faster optimization with superior quantitative and qualitative performance on 3D reconstruction and deformation tracking of surgical scenes, providing significant promise for further applications.

## 2   Method

### 2.1   Overview

LerPlane represents surgical procedures using static and dynamic fields, each of which is made up of three orthogonal planes (Sect. 2.3). It starts by using spatiotemporal importance sampling to identify high-priority tissue pixels and build corresponding rays (Sect. 2.4). Then we sample points along each ray and query features using linear interpolation to construct fused features. The fused features and encoded coordinate-time information are input to a lightweight MLP, which predicts color and density for each point (Sect. 2.5). To better optimize LerPlane, we introduce some training schemes, including sample-net, various regularizers, and a warm-up training strategy (Sect. 2.6). Finally, we apply volume rendering to produce predicted color and depth values for each chosen ray. The overall framework is illustrated in Fig. 2.

### 2.2   Preliminaries

Neural Radiance Field (NeRF) [16] is a coordinate-based neural scene representation optimized through a differentiable rendering loss. NeRF maps the 3D

coordinate and view direction of each point in the space into its color values $\boldsymbol{c}$ and volume density $\sigma$ via neural networks $\Phi_r$.

$$\boldsymbol{c}, \sigma = \Phi_r(x, y, z, \theta, \phi). \tag{1}$$

It calculates the expected color $\hat{C}(\mathbf{r})$ and the expected depth $\hat{D}(\mathbf{r})$ of a pixel in an image captured by a camera by tracing a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ from the camera center to the pixel. Here, $\mathbf{o}$ is the ray origin, $\mathbf{d}$ is the ray direction, and $t$ is the distance from a point to the $\mathbf{o}$, ranging from a pre-defined near bound $t_n$ to a far bound $t_f$. $w(t)$ represents a weight function that accounts for absorption and scattering during the propagation of light rays. The pixel color is obtained by classical volume rendering techniques [10], which involve sampling a series of points along the ray.

$$\hat{C}(\mathbf{r}) = \int_{t_n}^{t_f} w(t)\mathbf{c}(t)dt, \hat{D}(\mathbf{r}) = \int_{t_n}^{t_f} w(t)tdt, w(t) = \exp\left(-\int_{t_n}^{t} \sigma(s)ds\right)\sigma(t). \tag{2}$$
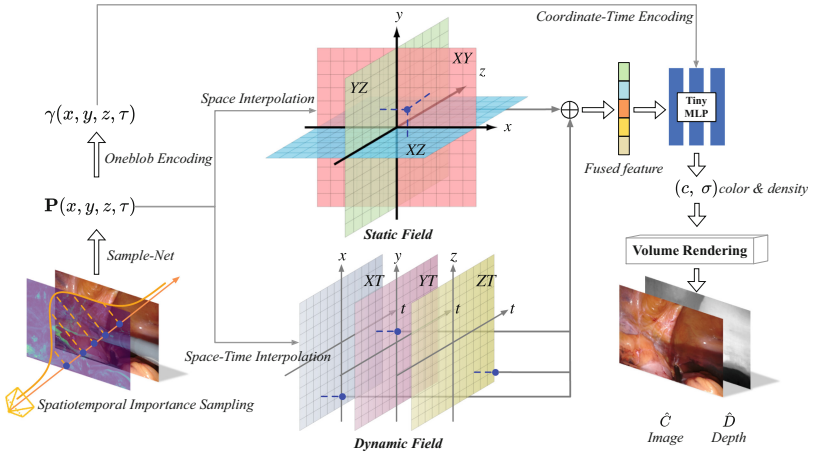


**Fig. 2.** Illustration of our fast 4D reconstruction method, LerPlane.

## 2.3 Neural LerPlane Representations for Deformable Tissues

A surgical procedure can be represented as a 4D volume, and we factorize the volume into 2D planes. Specifically, We represent a surgical scene using six orthogonal feature planes, consisting of three space planes (*i.e.*, *XY*, *YZ*, and *XZ*) for the static field and three space-time planes (*i.e.*, *XT*, *YT*, and *ZT*) for the dynamic field. Each space plane has a shape of $N \times N \times D$, and each space-time plane owns a shape of $N \times M \times D$, where $N$ and $M$ represent spatial and temporal resolution, respectively, and $D$ is the size of the feature.

To extract features from an image pixel $p_{ij}$ with color $C$ at a specific timestep $\tau$, we first cast a ray $\mathbf{r}(t)$ from $o$ to the pixel. We then sample spatial-temporal points along the ray, obtaining their 4D coordinates. We acquire a feature vector for a point $\mathbf{P}(x, y, z, \tau)$ by projecting it onto each plane and using bilinear interpolation $\mathcal{B}$ to query features from the six feature planes.

$$\mathbf{v}(x, y, z, \tau) = \mathcal{B}(F_{XY}, x, y) \odot \mathcal{B}(F_{YZ}, y, z) \ldots \mathcal{B}(F_{YT}, y, \tau) \odot \mathcal{B}(F_{ZT}, z, \tau), \quad (3)$$

where the $\odot$ represents element-wise multiplication, inspired by [4,22]. The fused feature vector $\mathbf{v}$ is then passed to a tiny MLP $\Theta$, which predicts the color $c$ and density $\sigma$ of the point. Finally, we leverage the Eq. 2 to get the predicted color $\hat{C}(\mathbf{r})$. Inspired by [14,17], we build the feature planes with multi-resolution planes, $e.g.$ $F_{XY}$ is represented by planes with $N = 128$ and $256$.

Existing methods [6,26] for reconstructing surgical procedures using pure implicit representations, requires traversing all possible positions in space-time, which is highly computationally and time-intensive. In contrast, LerPlane decomposes the surgical scene into six explicitly posed planes, resulting in a significant reduction in complexity and a much more efficient representation. This reduces the computational cost from $O(N^4)$ to $O(N^2)$ and enables the use of smaller neural networks, leading to a considerable acceleration in the training period. Besides, methods [7,12,19,20,23,26] using a single displacement field to supplement the static field struggle with handling variations in scene topology, such as non-linear deformations. In contrast, LerPlane can naturally model these situations using a dynamic field.

## 2.4   Spatiotemporal Importance Sampling

Tool occlusion in robotic surgery poses a challenge for reconstructing occluded tissues due to their infrequent occurrence in the training set, resulting in varied learning difficulties for different pixels. Besides, we observe that many tissues remain stationary over time, and therefore repeated training on these pixels contributes minor to the convergence, reducing efficiency. We design a novel spatiotemporal importance sampling strategy to address the issues above. In particular, we utilize binary masks $\{M_i\}_{i=1}^T$ and temporal differences among frames to generate sampling weight maps $\{\mathbf{W}\}_{i=1}^T$. These weight maps represent the sampling probabilities for each pixel/ray, drawing inspiration from [12,26]. One sampling weight map $\mathbf{W}_i$ can be determined by:

$$\mathbf{W}_i = \min(\max_{\substack{i-n<j \\ <i+n}}(\| \mathbf{I}_i M_i - \mathbf{I}_j M_j \|_1)/3, \alpha) \cdot \mathbf{\Omega}_i, \ \mathbf{\Omega}_i = \beta(M_i T / \sum_{i=1}^T M_i), \quad (4)$$

where $\alpha$ is a lower-bound to avoid zero weight among unchanged pixels, $\mathbf{\Omega}_i$ specifies higher importance scaling for those tissue areas with higher occlusion frequencies, and $\beta$ is a hyper-parameter for balancing augmentation among frequently occluded areas and time-variant areas. By unitizing spatiotemporal importance sampling, LerPlane concentrates on tissue areas and speeds up training, improving the rendering quality of occluded areas and prioritizing tissue areas with higher occlusion frequencies and temporal variability.

## 2.5 Coordinate-Time Encoding

Previous methods [6,26] apply positional encoded view direction $\gamma(\mathbf{d})$ to model view-dependent appearance. However, during endoscopic operations, camera movements are restricted. The view direction changes are typically minimal. Instead of view encoding, we propose using $\gamma(x, y, z, \tau)$ to enhance the spatiotemporal information. Specifically, the encoding along with the fused features $\mathbf{v}$ from feature planes is input to the MLP $\Theta$, which predicts $\sigma$ and $c$ of each point. Then we utilize Eq. 2 to render the expected color $\hat{C}$ and depth $\hat{D}$ of one specific ray.

## 2.6 Optimization

We adopt a joint supervision approach to optimize the tiny MLP $\Theta$ and feature planes using rendered color and depth. To further improve the optimization process, we propose several optimization schemes, including a sample-net for better-sampled points, a warm-up strategy to address outliers, and several regularizers.

*Sample-Net.* The sampling of spatiotemporal points is crucial for volume rendering, with a particular focus on sampling around tissue regions for optimal performance. We replaced the conventional two-stage time-consuming sampling strategy with a single sample-net and train it using histogram loss [3]. The sample-net is a lightweight single-resolution LerPlane model that provides more accurate sampling points for the full model.

*Regularizers.* We apply some regularization to address the limited information available in surgical scene reconstruction. We adopt 2D total variation (TV) loss for space planes in [5,8,24] and 1D TV loss on the space axis for space-time planes and a similar smooth loss on the time axis. Additionally, we introduce a minor time-invariant loss to separate the static and dynamic fields as much as possible, encouraging the features in space-time planes to remain unchanged.

*Warm-Up Training Strategy.* Since single-view captures cannot provide valid scale information, we leverage pseudo ground truth depth maps $D(\mathbf{r})$ generated by STTR-light [13] from stereo images to guide the optimization. Specifically, we apply a Huber loss for depth regularization:

$$\mathcal{L}_D = \begin{cases} 0.5 \Delta D(\mathbf{r})^2, & \text{if } |\Delta D(\mathbf{r})| < \delta \\ \delta \cdot (\Delta D(\mathbf{r}) - 0.5 \cdot \delta), & \text{otherwise} \end{cases} \tag{5}$$

where $\Delta D(\mathbf{r}) = |\hat{D}(\mathbf{r}) - D(\mathbf{r})|$ represents the absolute depth difference among valid depth values, $\delta$ is a threshold at which to change loss type. Considering that the predicted depth maps encounter a lot of unreliable depth values and missing areas [26], we design a simple by effective warm-up training strategy. Specifically, we apply the $\mathcal{L}_D$ to depths from both the sample-net and the full model during the first half of the training. In the remaining iterations, we disable the $\mathcal{L}_D$ and use other regularization to refine unreliable depths.

## 3   Experiments

### 3.1   Dataset and Evaluation Metrics

We evaluate our proposed method on the EndoNeRF dataset [26], a collection of typical robotic surgery stereo videos captured from stereo cameras at a single viewpoint during in-house DaVinci robotic prostatectomy procedures, which is designed to capture challenging surgical scenes with non-rigid deformation and tool occlusion. We evaluate our proposed method by comparing it to existing methods [15,26] using standard image quality metrics following [26], including PSNR, SSIM, and LPIPS. Additionally, to measure the consistency of the underlying 3D scene, we supplement these metrics using the FLIP metric [1,2]. For qualitative evaluation, we follow the exhibition method from [26].
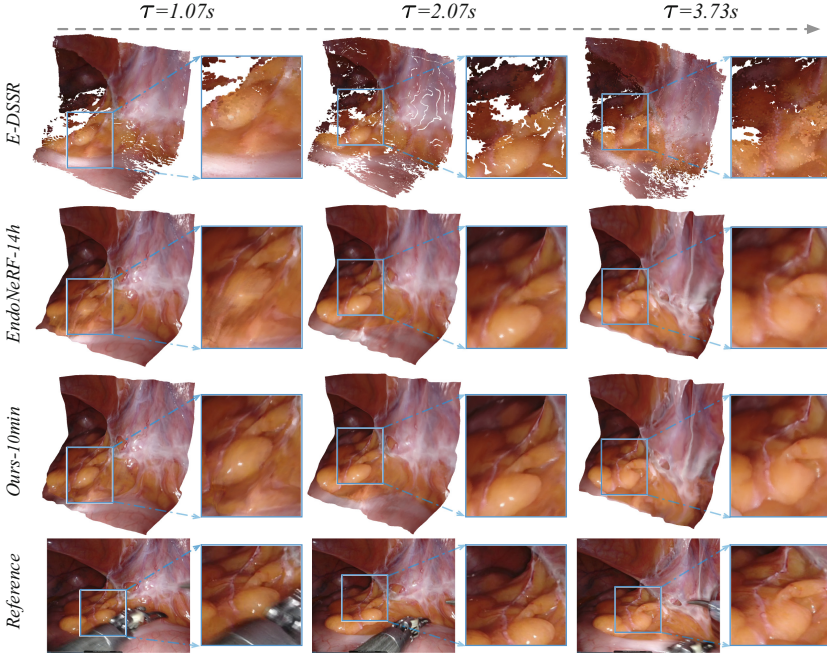


**Fig. 3.** Qualitative results on scene "traction" from different timesteps $\tau$.

### 3.2   Implementation Details

We normalize the scene into device coordinates (NDC) to handle single-view endoscopy videos and then project rays within the NDC space. The video duration is normalized to $[-1, 1]$. We use a two-stage sampling network with 128 and 256 dimensional plane features for the sample-net. Oneblob encoding [18] is applied to encode the spatiotemporal information. The full model consists of

four resolutions, 64, 128, 256, and 512 dimensions among space. Hyperparameters include $D = 32$ for feature planes, $j = 25$ for spatiotemporal importance sampling, $\xi = 16$ for Oneblob encoding dimensionality, and $\delta = 0.2$ for depth loss across all experiments. An Adam [11] optimizer is used with default values for optimization. In each iteration, 2048 rays are randomly sampled from the whole dataset to form a batch. The initial learning rate is set to 0.01. We apply a cosine schedule with a 512 iterations warming-up stage. We train all scenes with $9k$ and $32k$ iterations, which take around 3 and 10 minutes, respectively, on a single RTX 3090 GPU running the Ubuntu 20.04. Our LerPlane is implemented with pure Pytorch [21]. The code is available at https://github.com/Loping151/LerPlane.

### 3.3   Evaluation

We compare our proposed method, LerPlane, against two existing SOTA methods: the surfel warping-based method, E-DSSR [15] and NeRF-based method EndoNeRF [26]. We find that E-DSSR struggles to completely reconstruct surgical scenes, resulting in many holes and noisy points (see Fig. 3), which leads to poor numerical performance. In contrast, EndoNeRF achieves high-fidelity reconstruction of deformable tissues but requires around 14 h of optimization, which is computationally expensive and constrains intraoperative use. LerPlane, on the other hand, achieves comparable results to EndoNeRF with only 3 min of optimization, providing nearly 280-fold acceleration. Moreover, with a longer optimization time of 10 min, LerPlane outperforms both E-DSSR and EndoNeRF in terms of all metrics, as shown in Table 1. Our novel importance sampling and encoding strategies further enhance the ability of LerPlane to preserve details and produce accurate visualizations of deformable tissues, as demonstrated in Fig. 3. Our results demonstrate that LerPlane achieves significantly faster optimization without compromising reconstruction quality, showing great potential for future clinical applications in robotic surgery.

**Table 1.** Quantitative results on the EndoNeRF dataset. Please refer to Sect. 3.4 for explanations of the acronyms used.

| Methods | PSNR↑ | SSIM↑ | LPIPS↓ | FLIP↓ | Time↓ |
|---|---|---|---|---|---|
| E-DSSR [15] | 13.398 ± 1.387 | 0.630 ± 0.057 | 0.423 ± 0.047 | \ | \ |
| EndoNeRF [26] | 29.272 ± 2.836 | 0.921 ± 0.022 | 0.088 ± 0.020 | 0.085 ± 0.018 | 14 h |
| Ours-NS | 31.532 ± 1.665 | 0.886 ± 0.021 | 0.142 ± 0.020 | 0.112 ± 0.016 | 3 min |
| Ours-TS | 31.544 ± 1.669 | 0.886 ± 0.021 | 0.142 ± 0.020 | 0.111 ± 0.015 | 3 min |
| Ours-VE | 32.353 ± 1.742 | 0.897 ± 0.022 | 0.131 ± 0.024 | 0.103 ± 0.012 | 3 min |
| Ours-NE | 32.230 ± 1.655 | 0.895 ± 0.023 | 0.131 ± 0.020 | 0.102 ± 0.012 | 3 min |
| Ours-9k | 32.589 ± 1.451 | 0.901 ± 0.021 | 0.126 ± 0.028 | 0.103 ± 0.014 | 3 min |
| Ours-32k | 35.504 ± 3.076 | 0.935 ± 0.026 | 0.083 ± 0.022 | 0.075 ± 0.031 | 10 min |

### 3.4   Ablation Study

We conduct ablation studies on the EndoNeRF dataset to understand the key components and demonstrate their effectiveness. Table 1 shows the performance of all experiments.

1. *Sampling Strategy.* We compare with two different methods: naively avoiding tool masks, assigning equal weights to other pixels (Ours-NS), and assigning higher probabilities to highly occluded areas (Ours-TS), as in  [26]. Our method effectively prioritizes time-variant and highly occluded areas, significantly improving convergence speed.
2. *Encoding Strategy.* Experiments showed that coordinate-time encoding achieves better performance in all metrics compared to no encoding (Ours-NE) or direction encoding (Ours-VE), showing the effectiveness of the proposed encoding.

Further analysis of the optimization schemes is available in the Supplementary Materials.

## 4   Conclusion and Future Work

In this paper, we introduced LerPlane, a fast and accurate method for reconstructing deformable tissues from endoscopic videos. By utilizing multi-plane fields and spatiotemporal importance sampling, we can handle tool occlusion and large motion while significantly accelerating optimization. Our experiments show that LerPlane achieves rendering quality comparable to or better than EndoNeRF in just three minutes, which is over $100\times$ faster. We believe that LerPlane could improve robotic surgery scene understanding, benefiting various clinical-oriented tasks and intraoperative surgery applications.

Currently, the inference speed of our Lerplane is slow, In our future research endeavors, our main emphasis will revolve around improving the inference time of our approach, with the primary goal of efficiently supporting intraoperative operations. Moreover, we will dedicate our efforts to reducing the input data requirements, thereby aiming to broaden the applicability of LerPlane to a wider range of surgical scenarios.

# References

1. Andersson, P., Nilsson, J., Akenine-Möller, T., Oskarsson, M., Åström, K., Fairchild, M.D.: FLIP: a difference evaluator for alternating images. Proc. ACM Comput. Graph. Interact. Tech. **3**(2), 15-1 (2020)
2. Andersson, P., Nilsson, J., Shirley, P., Akenine-Möller, T.: Visualizing errors in rendered high dynamic range images. Eurographics (2021)
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-NeRF 360: unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5470–5479 (2022)
4. Chan, E.R., et al.: Efficient geometry-aware 3D generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16123–16133 (2022)
5. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: tensorial radiance fields. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022, Part XXXII. LNCS, vol. 13692, pp. 333–350. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19824-3_20
6. Corona-Figueroa, A., Frawley, J., Bond-Taylor, S., Bethapudi, S., Shum, H.P., Willcocks, C.G.: MedNeRF: medical neural radiance fields for reconstructing 3D-aware CT-projections from a single x-ray. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 3843–3848. IEEE (2022)
7. Fang, J., et al.: Fast dynamic radiance fields with time-aware neural voxels. In: SIGGRAPH Asia 2022 Conference Papers, pp. 1–9 (2022)
8. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5501–5510 (2022)
9. Hedman, P., Srinivasan, P.P., Mildenhall, B., Barron, J.T., Debevec, P.: Baking neural radiance fields for real-time view synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5875–5884 (2021)
10. Kajiya, J.T., Von Herzen, B.P.: Ray tracing volume densities. ACM SIGGRAPH Comput. Graph. **18**, 165–174 (1984)
11. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
12. Li, T., et al.: Neural 3d video synthesis from multi-view video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5521–5531 (2022)
13. Li, Z., et al.: Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6197–6206 (2021)
14. Liu, L., Gu, J., Zaw Lin, K., Chua, T.S., Theobalt, C.: Neural sparse voxel fields. Adv. Neural. Inf. Process. Syst. **33**, 15651–15663 (2020)
15. Long, Y., et al.: E-DSSR: efficient dynamic surgical scene reconstruction with transformer-based stereoscopic depth perception. In: de Bruijne, M., et al. (eds.) MICCAI 2021, Part IV. LNCS, vol. 12904, pp. 415–425. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87202-1_40
16. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. Commun. ACM **65**(1), 99–106 (2021)

17. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. (ToG) **41**(4), 1–15 (2022)
18. Müller, T., McWilliams, B., Rousselle, F., Gross, M., Novák, J.: Neural importance sampling. ACM Trans. Graph. (ToG) **38**(5), 1–19 (2019)
19. Park, K., et al.: NeRFies: deformable neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5865–5874 (2021)
20. Park, K., et al.: HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields. arXiv preprint arXiv:2106.13228 (2021)
21. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 32, pp. 8024–8035. Curran Associates, Inc. (2019)
22. Peng, S., Niemeyer, M., Mescheder, L., Pollefeys, M., Geiger, A.: Convolutional occupancy networks. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020, Part III. LNCS, vol. 12348, pp. 523–540. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58580-8_31
23. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-NeRF: neural radiance fields for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10318–10327 (2021)
24. Schwarz, K., Sauer, A., Niemeyer, M., Liao, Y., Geiger, A.: VoxGRAF: fast 3D-aware image synthesis with sparse voxel grids. arXiv preprint arXiv:2206.07695 (2022)
25. Sun, C., Sun, M., Chen, H.T.: 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. 5459–5469 (2022)
26. Wang, Y., Long, Y., Fan, S.H., Dou, Q.: Neural rendering for stereo 3D reconstruction of deformable tissues in robotic surgery. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022, Part VII. LNCS, vol. 13437, pp. 431–441. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16449-1_41
27. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: PlenOctrees for real-time rendering of neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5752–5761 (2021)